

An Inverted Curve, an Uninverted Economy: A Structural Break in Recession Forecasting

*Rafael Almeida**, *Sophie Beaumont*, *Hye-Won Jeong*

Frontier Institute for Computational Economics (FICE)

Generative Economic Review • May 18, 2026

GER 1.10

JEL Classification: E32, E43, E44, E52, E58, G12, C22, C25, C58

Keywords: yield curve, term structure, recession prediction, structural break, sup-F test, term premium, probit forecasting, AUC, post-2008 anomaly, quantitative easing, FAIT, Federal Reserve, zero lower bound, near-term forward spread, term-premium decomposition

Abstract

We document a structural break in the relationship between the US Treasury yield curve and subsequent NBER recessions that has not, to our knowledge, been systematically characterized in the existing literature. Using monthly FRED data from January 1962 through April 2026, we estimate a probit model of 12-month-ahead recession occurrence on the contemporaneous 10-year minus 3-month yield spread and find that the slope coefficient on the spread has collapsed from -1.36 ($t = -7.22$; HAC $t = -4.18$) in the pre-2008 sub-sample to $+0.04$ ($t = +0.14$; HAC $t = +0.11$) in the post-2008 sub-sample, with the area under the receiver-operating-characteristic curve falling from 0.898 to 0.491. The post-2020 sub-sample yields an AUC of 0.000, statistically indistinguishable from worst-case prediction. The empirical centerpiece is the 2022Q4–2024Q4 inversion episode, in which the 10Y–3M spread averaged -0.94 percentage points across 25 consecutive months—the longest sustained inversion in the post-1962 record—without a single NBER-dated recession month. We compare specifications across four spread definitions (10Y–3M, 10Y–2Y, 10Y–5Y, 10Y–FF) and five forecast horizons (3, 6, 12, 18, 24 months); the break appears in all twenty combinations. A sup-F structural-break test on the slope coefficient localizes the maximum-likelihood break date to 2009Q1, immediately following the launch of the first large-scale asset purchase program, with the 90% confidence interval [2008Q3, 2009Q3]. We perform an Adrian-Crump-Moench (2013) term-premium decomposition and run the predictive probit separately on the expectations and term-premium components, finding that both components lose predictive content

post-2008. The Engstrom-Sharpe near-term forward spread similarly attenuates (post-2008 AUC = 0.51), weighing against the strict term-premium-compression account. We partition the post-2008 sample into zero-lower-bound and non-ZLB periods and find that the attenuation is present in both environments, though more severe during non-ZLB episodes. We discuss four candidate explanations for the break—term-premium compression following secular declines and quantitative easing; post-COVID fiscal expansion that has substituted for the conventional credit channel; household and corporate balance-sheet strength entering the 2022 tightening cycle; and pandemic-era labor-market reallocation—and engage an alternative framing in which the 2022–2024 outcome represents a monetary-policy success rather than a forecasting failure. We close by drawing implications for forecasting practice, monetary-policy deliberation, and the broader literature on the empirical content of term-structure variables in an institutional environment where the central bank is itself a substantial participant in the long end of the curve.

1. Introduction

The inverted yield curve has been the most reliable single recession indicator in postwar US macroeconomic data. Beginning with the foundational work of Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998) and continuing through the contemporary research of Bauer and Mertens (2018), Engstrom and Sharpe (2019), and Stein (2024), the empirical literature has repeatedly confirmed that an inversion of the term structure—the configuration in which short-term Treasury yields exceed long-term Treasury yields—precedes NBER-dated recessions with a lead time of approximately twelve months. The reliability of the indicator over the half-century from the early 1960s through the late 2010s was sufficient to make it a fixture of the working economist’s toolkit, a routine input to the deliberations of the Federal Open Market Committee, and a focal point for forecasting models in both academic and practitioner contexts.

1.1 The framing hypothesis

This paper makes one central empirical claim. The post-2022 absence of recession following a deep and sustained yield-curve inversion is not anomalous. It is the predictable consequence of a regime change in the empirical relationship between the term structure and subsequent recessions that emerged around 2008–2009 and intensified after 2020. The probit slope coefficient on the 10Y–3M spread, the central quantitative object of the recession-forecasting literature, has collapsed from -1.36 in the pre-2008 sub-sample to a value statistically indistinguishable from zero in the post-2008 sub-sample; the corresponding AUC has fallen from 0.898 to 0.491. The collapse is robust across four alternative spread

definitions and five forecast horizons. If the claim is correct, the conventional reading of yield-curve signals in macroeconomic forecasting and monetary-policy deliberation needs to be conducted under the post-break parameters rather than the pre-break parameters that the foundational literature established.

An important caveat frames the analysis that follows. The post-2008 yield curve is not the same kind of object as the pre-2008 yield curve. From 2008 forward, the Federal Reserve has been a substantial participant in the long end of the Treasury market through three rounds of quantitative easing and the post-2020 portfolio expansion, with the System Open Market Account holding over \$7 trillion in Treasury and agency securities at its 2022 peak (Gagnon et al., 2011; Krishnamurthy and Vissing-Jorgensen, 2011). Forward guidance has been a designed feature of FOMC communication since at least 2011, and the August 2020 Flexible Average Inflation Targeting (FAIT) framework explicitly conditions the expected rate path on the Committee's tolerance for asymmetric inflation outcomes (Swanson, 2021). In this institutional environment, the long end of the curve is partially endogenous to the central bank's own balance-sheet policy and communication strategy. A signal that the policymaker is co-determining is not the same kind of forecasting object as a market-determined signal that the policymaker merely reads. We return to this endogeneity concern at length in Section 5.9, where we engage the alternative framing under which the 2022–2024 episode represents a monetary-policy success rather than a forecasting failure. The present analysis documents the statistical break in the reduced-form relationship; the interpretive question of whether the break reflects a change in the macroeconomy or a change in the institutional environment in which the indicator is observed is one the present design cannot fully adjudicate.

1.2 Four contributions

The paper makes four substantive contributions to the recession-forecasting literature.

First, we provide the most up-to-date empirical record of the yield-curve recession relationship, extending the historical analysis through the 2022Q4 inversion and its subsequent normalization. The 25-month duration of the 2022–2024 inversion makes it the longest in the post-1962 record by a substantial margin (the previous longest was 17 months for the 1978–1980 episode), and its termination without a recession is the single most consequential empirical fact for the modern term-structure literature.

Second, we demonstrate that the structural break in the spread-recession relationship pre-dates the 2022 episode by more than a decade. A sup-F structural-break test on the probit slope coefficient localizes the break to 2009Q1, immediately following the launch of the first large-scale asset purchase program. The contemporary interpretation of the 2022–2024 episode as a *sui generis* event must be reconciled with the empirical fact that the spread's predictive content had already been substantially attenuated for fourteen years before the inversion began.

Third, we test the robustness of the break to alternative spread definitions, forecast horizons, and—in this revision—to an explicit term-premium decomposition following Adrian et al. (2013) and a partition of the post-2008 sample into zero-lower-bound (ZLB) and non-ZLB environments. The 10Y–3M, 10Y–2Y, 10Y–5Y, and 10Y–Fed-Funds spreads, examined at horizons of 3, 6, 12, 18, and 24 months, yield twenty combinations of specification. The qualitative break appears in all twenty. The term-premium decomposition reveals that both the expectations component and the term-premium component of the spread have lost predictive content post-2008, weighing against the strict version of the term-premium-compression account.

Fourth, we frame four candidate explanations for the break—term-premium compression, post-COVID fiscal substitution for the credit channel, balance-sheet strength, and labor-market reallocation—and identify what additional evidence would discriminate among them. We also engage an alternative framing, advanced most forcefully in the heterodox and policy-engaged literatures, under which the 2022–2024 outcome is not a forecasting failure but a monetary-policy success. We do not claim to adjudicate; we claim to specify the research agenda that the break poses, and to ensure that the reader can see both the conventional and alternative framings side by side.

1.3 Intellectual history of the question

The question this paper engages reached its current form through a sequence of three intellectual transitions. Estrella and Hardouvelis (1991) established the term structure as a leading indicator of real activity. Estrella and Mishkin (1998) systematized the probit-based recession-forecasting methodology that has dominated the literature since. The contemporary literature has, in the post-2008 period, increasingly questioned whether the foundational findings continue to hold under the institutional changes of the past decade: the term-premium compression following quantitative easing (Adrian et al., 2013; Bauer and Rudebusch, 2020), the secular decline in the natural rate of interest (Holston et al., 2017), the fundamental shift in the Federal Reserve’s reaction function codified in the August 2020 FAIT framework, and the post-COVID expansion of fiscal policy that has substituted for the conventional credit channel of monetary transmission. A parallel literature, rooted in the preferred-habitat tradition of Greenwood et al. (2010) and the broader work on central-bank balance-sheet effects (Gagnon et al., 2011; Krishnamurthy and Vissing-Jorgensen, 2011), has emphasized that modern central-bank operations have made the long end of the curve partially endogenous to policy, raising the question of whether a ‘predictive power’ framing is even well-posed for the post-2008 institutional environment. The structural-break framing of the present paper completes this sequence: not whether the relationship has changed in the abstract, but *when* the change began, *how large* it has become, and *what mechanisms* are consistent with the observed empirical record.

1.4 What the paper claims

The paper makes six explicit empirical claims that the reader can evaluate against the evidence presented in Sections 4 and 5:

1. The probit slope coefficient on the 10Y–3M spread at the 12-month forecast horizon has collapsed from -1.36 ($t = -7.22$; Newey-West HAC $t = -4.18$) in the pre-2008 sub-sample to $+0.04$ ($t = +0.14$; HAC $t = +0.11$) in the post-2008 sub-sample.
2. The AUC of the same probit specification has fallen from 0.898 to 0.491 across the same partition; the post-2020 sub-sample yields an AUC of 0.000.
3. The sup-F structural-break test localizes the maximum-likelihood break date to 2009Q1, with a 90% confidence interval of [2008Q3, 2009Q3].
4. The collapse is robust across four spread definitions and five forecast horizons; all twenty combinations exhibit the same qualitative break.
5. An Adrian-Crump-Moench term-premium decomposition reveals that both the expectations component and the term-premium component of the post-2008 spread have lost recession-predictive content; the Engstrom-Sharpe near-term forward spread similarly attenuates (post-2008 AUC = 0.51).
6. The 2022Q4–2024Q4 inversion episode is the longest in the post-1962 record (25 months) and was not followed by an NBER-dated recession; it is parsimoniously interpreted as an extreme expression of the post-2009 regime rather than as a *sui generis* event.

The first five claims are descriptive; the sixth is interpretive. We are explicit that the mechanisms underlying the break are not identified by the present analysis. Four candidate accounts are consistent with the empirical record, and discriminating among them is the substantive research agenda Section 5 lays out. We are also explicit that the 'predictive power lost' framing is one of two defensible readings of the data; the alternative—that the central bank has internalized the signal and the 2022–2024 episode is a policy success—is consistent with the same evidence and is engaged in Section 5.9.

1.5 Roadmap

Section 2 reviews the literatures on the yield curve as a recession indicator, on the determinants and decomposition of the term premium, on the contemporary empirical literature documenting the post-2008 attenuation, on the preferred-habitat and balance-sheet-effects tradition, on probit and AUC inference for recession forecasting, on structural-break methodology in macroeconometrics, on the post-COVID macroeconomic environment, and on the zero-lower-bound and monetary-policy-endogeneity literatures. Section 3 describes the data,

the four spread definitions, the probit specification, the term-premium decomposition, the ZLB partition, the sup-F structural-break test, the AUC-based model comparison, and the pre-specified robustness margins. Section 4 reports the central empirical findings, including the term-premium decomposition results, the near-term forward spread analysis, and the ZLB partition. Section 5 discusses interpretations, the four candidate accounts for the break, the alternative monetary-policy-success framing, the endogeneity of the post-2008 indicator, the limitations of the present analysis, the implications for forecasting practice, and the international evidence. Section 6 concludes.

We emphasize at the outset what the paper does and does not claim. The paper documents a statistical structural change in an empirically established relationship. It does not claim a causal mechanism for the break, nor does it claim that the yield curve will never again predict recessions. The post-2008 sample, while substantial (208 monthly observations after horizon adjustment), is short by the standards of long-horizon recession forecasting; future business cycles may restore the predictive content of the spread. What the paper does claim is that the empirical magnitude of the contemporary failure is sufficient that forecasting work, monetary policy practice, and academic research relying on the yield curve as a recession indicator should treat the post-2008 sample as a regime distinct from the historical record on which the indicator's reliability was established.

2. Literature Review

The yield-curve recession-forecasting literature is sufficiently large and well-developed that we structure our review around eight sub-strands of direct relevance to the present analysis. We treat each in turn and close with a paragraph on the position of the present paper.

2.1 The yield curve as recession indicator

The empirical observation that inverted yield curves precede US recessions has been documented and refined over four decades. Estrella and Hardouvelis (1991) provide the foundational evidence, demonstrating that the slope of the term structure—measured variously as the 10-year minus 3-month yield spread, the 10-year minus 2-year spread, or the 10-year minus Federal Funds spread—contains substantial information about subsequent output growth. Their work was motivated by the observation that each of the six NBER-dated recessions between 1960 and 1990 was preceded by a yield-curve inversion, with lead times ranging from eight to seventeen months. Estrella and Mishkin (1998) formalize the empirical methodology, demonstrating that probit and logit models of recession occurrence given the spread are well-calibrated across alternative specifications, that the optimal forecast horizon is approximately 12 months, and that the predictive content of the spread is robust to controls for other leading economic indicators including the stock market, the index of leading indicators, and measures of monetary policy stance.

Estrella and Trubin (2006) provide a practitioner-oriented synthesis of the literature, including the well-known guideline that a sustained inversion of three months or more in the 10Y–3M spread historically precedes recession by approximately one year. Their synthesis was influential at the Federal Reserve Bank of New York, which maintains a public recession probability model based on the Estrella-Mishkin specification. Wright (2006) extends the literature to multi-factor specifications that incorporate the level and curvature of the term structure together with the slope, finding that the slope alone captures the bulk of the recession-predictive content. Rudebusch and Williams (2009) document the persistence of the yield-curve recession relationship through the 2000s and identify it as a puzzling exception to the otherwise general decline in forecast skill that empirical macroeconomic forecasting models have exhibited since the onset of the Great Moderation.

2.2 The term premium and yield-curve decomposition

A second strand of the literature has decomposed the spread into its constituent parts and identified the components that drive the predictive relationship. Hamilton and Kim (2002) apply the affine term-structure framework to decompose the 10Y–3M spread into a term-premium component and an expected-future-short-rate component, finding that the expected-future-short-rate component is the dominant driver of the spread’s recession-predictive content. The implication is that the spread predicts recessions because a downward-sloping curve reflects market expectations of falling short-term interest rates, which in turn the Federal Reserve is expected to deliver in response to economic contraction.

Kim and Wright (2005) provide an arbitrage-free three-factor term-structure model that has become a workhorse decomposition in the contemporary literature. Their model decomposes the 10-year yield into an expected-short-rate path and a residual term premium, with the latter capturing the compensation investors demand for holding duration risk. Adrian et al. (2013) provide a contemporary updating of the term-premium decomposition using a regression-based approach that is computationally tractable and has become the standard tool for applied work. They document the secular decline in the US term premium since the 1980s, from approximately 200 basis points in the early 1980s to near zero or negative values in the 2010s. Bauer and Rudebusch (2020) extend the decomposition to a framework that explicitly incorporates the secular decline in the equilibrium real rate of interest, finding that the bulk of the decline in long-term yields since the early 1980s reflects falling real rates rather than falling inflation compensation.

Cochrane and Piazzesi (2005) document the existence of a single forecasting factor for bond risk premia that summarizes the predictive content of the forward-rate structure. Cieslak and Povala (2015) update and refine the bond-risk-premium evidence, providing the contemporary benchmark for term-premium-aware forecasting. The implications for recession forecasting are substantive: if a substantial fraction of the spread’s level reflects time-varying risk premia rather than expectations of falling short rates, the spread’s

recession-predictive content depends on how the risk-premium component co-moves with the expectations component. Benzoni et al. (2018) address this question directly, arguing that the recession-predictive content of the spread operates primarily through the expectations component and that the risk-premium component is a source of noise rather than signal for recession forecasting.

2.3 The post-2008 anomaly and contemporary debate

The contemporary period has prompted explicit reconsideration of the conventional indicator's continued reliability. Stein (2024) reviews the post-2022 inversion episode and discusses whether the conventional 10Y–3M spread has lost its recession-predictive content, drawing on similar evidence to what we present below. Bauer and Swanson (2023) explicitly identify the disconnect between the inverted yield curve of 2022–2024 and the absence of recession as a research challenge for the term-structure literature.

Engstrom and Sharpe (2019) propose the near-term forward spread—the difference between the six-quarter-ahead implied 3-month yield and the current 3-month yield—as a recession indicator that more directly captures the expected-future-short-rate component of the conventional spread. They argue that this alternative indicator continues to perform reliably even in environments where the conventional 10Y–3M spread is influenced by term-premium dynamics that may obscure the recession signal. The empirical content of this proposal is one of the principal pieces of evidence bearing on whether the post-2008 attenuation of the conventional spread reflects term-premium compression specifically or a broader change in the macroeconomic environment. We test this claim directly in Section 4.7 and find that the NTFS also attenuates post-2008, a result with substantial diagnostic implications.

Ng and Wright (2013) survey the post-Great-Recession forecasting record more broadly and document the substantial decline in forecast skill across a range of macroeconomic models, of which the yield-curve recession model is one prominent example. Bordo and Levy (2021) provide a historical perspective, arguing that the yield curve's predictive power has varied substantially across monetary regimes and that the post-2008 attenuation is best understood as a regime-specific phenomenon rather than a permanent structural change. The interpretive question of whether the decline reflects a structural break in the underlying relationships or a temporary departure from those relationships is, as these authors note, unresolvable on the basis of a single post-recession sample.

2.4 The preferred-habitat tradition and balance-sheet effects

A fourth strand of the literature, of particular relevance to the post-2008 institutional environment, emphasizes the demand-side determinants of the term structure. Greenwood et al. (2010) develop a preferred-habitat model in which the supply of government bonds at different maturities affects term premia through the portfolio-balance channel. Under this

framework, the Federal Reserve's large-scale asset purchases—which removed duration from the market—would compress term premia mechanically, with the degree of compression depending on the elasticity of demand from preferred-habitat investors. Gagnon et al. (2011) provide the foundational empirical evaluation of the first LSAP program, estimating that QE1 reduced the 10-year term premium by 30 to 100 basis points. Krishnamurthy and Vissing-Jorgensen (2011) refine the estimate and distinguish between the signaling and portfolio-balance channels, finding that both contributed to the yield decline but through different segments of the curve.

D'Amico and King (2013) extend the evidence to the stock and flow effects of Treasury purchases, documenting that both the announcement and the ongoing purchases affected yields, with the stock effect dominating. The cumulative implication of this literature is that the post-2008 long end of the Treasury curve was not a pure market-determined signal but was substantially influenced by the Federal Reserve's portfolio decisions. The magnitude of this influence—estimated at 100 to 150 basis points of term-premium compression at the peak of QE3 (Adrian et al., 2013)—is large relative to the historical term-premium levels that the conventional spread's recession-predictive content was estimated on. The post-Keynesian monetary literature has developed this insight further: Lavoie (2014) argues that modern central-bank operations have transformed the yield curve from a market-determined object into a partially policy-determined one, and Mehrling (2011) provides a broader institutional account of how the Federal Reserve's expansion into the long end of the curve has changed the meaning of term-structure signals. Pozsar (2014) extends the analysis to the shadow-banking system, arguing that the post-2008 institutional architecture has created a new set of collateral and funding markets whose interaction with the Treasury curve is qualitatively different from the pre-2008 environment.

2.5 Probit and AUC inference for binary outcomes

The empirical methodology we employ is the probit-AUC framework that has dominated the recession-forecasting literature since Estrella and Mishkin (1998). The probit model—a single-equation specification for the probability of a binary outcome conditional on a continuous predictor—has the appealing property that its slope coefficient is directly interpretable in standard-deviation units and that its goodness of fit can be evaluated using the AUC. Hosmer and Lemeshow (2000) provide the standard treatment of probit and logistic regression diagnostics; Pepe (2003) provides the methodological foundation for the AUC and its sampling properties.

The AUC—the area under the receiver-operating-characteristic curve—has emerged as the standard model-fit metric for binary forecasting in the macroeconomic context. An AUC of 1 corresponds to perfect prediction; an AUC of 0.5 to random prediction; an AUC of 0 to perfectly inverted prediction. The 95% confidence interval around an AUC estimate, for sample sizes typical of recession forecasting (approximately 200 observations), is on the

order of ± 0.05 . The post-2008 AUC of 0.491 we report below is therefore not statistically distinguishable from 0.5—the random-prediction benchmark—at conventional significance levels.

Romano and Wolf (2005) address the multiple-testing concerns that arise when comparing alternative forecasting models across many specifications. The Romano-Wolf stepdown procedure they propose provides the relevant family-wise error rate control for our cross-specification comparisons in Section 4.

2.6 Structural-break methodology in macroeconometrics

The structural-break methodology we employ has a substantial history in the macroeconomic literature. Andrews (1993) establishes the sup-F test for parameter instability at an unknown change point, providing the critical values we use. Bai and Perron (1998) extend the framework to multiple breaks and develop the algorithm that has become the practical standard for empirical break detection. Hansen (2000) provides bootstrap-based critical values that are robust to heteroskedasticity. Bai (1997) provides confidence intervals for the estimated break date, addressing the well-known asymmetry in the sampling distribution of the break-date estimator.

The application of structural-break methods to recession forecasting has been less developed than the parallel application to the broader term-structure literature (Sims and Zha, 2006; Boivin, 2006). Chinn and Kucko (2015) apply structural-break tests to the yield-curve recession relationship in an international cross-section and document substantial cross-country variation in the timing of the post-2000 attenuation. The present paper applies the simplest version of this framework—a sup-F test for a single break at an unknown date—to the US 10Y–3M spread and verifies robustness through specification variation.

2.7 The post-COVID macroeconomic environment

The post-COVID period has been characterized by features that are sufficiently distinctive that the contemporary recession-forecasting environment may differ from the half-century preceding it. Blanchard and Bernanke (2024) document the supply-side disturbances of 2020–2022 and argue that the disinflation of 2023–2024 has reflected the working-out of pandemic-era supply shocks rather than the demand contraction that conventional Phillips-curve frameworks would predict. Furman (2024) provides a complementary perspective on the role of the post-COVID fiscal expansion in sustaining aggregate demand through what would conventionally have been a recessionary period. Haltiwanger and Kozeniaskas (2024) examine the post-pandemic dynamics of US labor markets and document the unusual co-movement of unemployment and inflation in the post-2022 period.

The post-COVID monetary-policy environment has been similarly distinctive. The 2022Q3–2023Q3 tightening cycle was the most aggressive since the Volcker disinflation, raising the federal funds rate from near zero to above 5% within approximately 15 months.

The subsequent transmission to credit-sensitive sectors of the real economy operated more weakly than the historical record would predict: residential construction contracted from its 2022 peak, commercial real estate experienced substantial valuation declines, and manufacturing employment was weak, but service-sector employment, household consumption, and aggregate output remained strong. Romer and Romer (2004) provide the methodological framework for measuring the macroeconomic effects of monetary policy shocks; the 2022–2024 transmission pattern, in which credit-sensitive sectors contracted while aggregate output remained positive, represents a partial-transmission outcome that the standard Gertler and Karadi (2015) intermediary-based transmission model would predict when intermediary balance sheets are strong and the fiscal stance is expansionary.

Mian et al. (2017) document the role of household debt in driving business-cycle dynamics in advanced economies, finding that increases in household debt-to-income ratios systematically precede output contractions. The opposite pattern in the 2022 tightening cycle—household debt-service ratios were at multi-decade lows entering the cycle, in part as a legacy of the post-GFC deleveraging and the 2020–2021 zero-rate refinancing wave—supports the balance-sheet-strength account of the 2022–2024 outcome.

2.8 The zero lower bound and indicator endogeneity

The zero lower bound on nominal interest rates, binding in the US from December 2008 to December 2015 and again from March 2020 to March 2022, creates a distinct analytical challenge for yield-curve-based recession forecasting. When the policy rate is at or near zero, the short end of the spread is mechanically constrained and the spread reduces approximately to the level of the long rate. Swanson (2021) documents how forward guidance and large-scale asset purchases substituted for conventional rate policy at the ZLB, changing the information content of the yield curve. Brunnermeier and Koby (2018) develop the ‘reversal rate’ framework in which accommodative monetary policy at the ZLB can become contractionary through the bank-profitability channel, with implications for the sign of the spread-recession relationship that the conventional probit framework does not capture.

The broader point—that the ZLB periods and the non-ZLB periods within the post-2008 sample represent qualitatively different forecasting environments—motivates the partition analysis we present in Section 4.8. Pooling ZLB and non-ZLB observations in a single post-2008 probit conflates two environments in which the economic content of the spread differs. If the attenuation is concentrated in the ZLB observations, the structural-break interpretation is weakened (the break would be an artifact of the lower bound rather than a change in the macroeconomic transmission mechanism). If the attenuation persists in the non-ZLB observations, the structural-break interpretation is reinforced.

Bekaert et al. (2001) provide a broader cautionary note on the interpretation of term-structure forecasting regressions across regimes, documenting that the apparent predictive content of the spread can be sensitive to regime-specific features of the joint distribution

of yields and output. Their analysis of 'peso problems' in the term structure—in which rare events that did not occur in a given sub-sample affect the conditional distribution—is directly relevant to the post-2008 sample, which contains one pandemic-driven recession and zero conventional-credit-cycle recessions.

2.9 Position of the present paper

The present paper contributes most directly to the contemporary debate on the post-2008 anomaly (Stein, 2024; Bauer and Swanson, 2023; Engstrom and Sharpe, 2019) by introducing a structural-break framing that localizes the break date and quantifies its magnitude across four spread definitions and five horizons. It contributes to the empirical-methodology literature on recession forecasting (Estrella and Mishkin, 1998; Estrella and Trubin, 2006) by providing a worked application of the sup-F structural-break framework in which the break date is policy-substantively interpretable (coinciding with the launch of the first large-scale asset purchase program). It contributes to the term-premium decomposition literature (Adrian et al., 2013; Bauer and Rudebusch, 2020) by performing the decomposition and running the predictive probit separately on the expectations and term-premium components, identifying the diagnostic margins along which the term-premium-compression account can be empirically evaluated. It contributes to the preferred-habitat and balance-sheet-effects literature (Greenwood et al., 2010; Krishnamurthy and Vissing-Jorgensen, 2011) by engaging the endogeneity concern directly and ensuring that the alternative framing—in which the indicator's loss of content reflects the central bank's internalization of the signal—is visible to the reader. The contribution we do not make is to identify the structural mechanism behind the break; the probit regression we estimate is a correlated reduced form, and the interpretive question of which of the four candidate accounts (or which combination) best explains the documented break is a question the present design cannot fully adjudicate.

3. Methodology

This section specifies the data, the four spread constructions, the probit model, the term-premium decomposition, the near-term forward spread construction, the ZLB partition, the sup-F structural-break test, the AUC-based model comparison, and the pre-specified robustness margins.

3.1 Data

The primary data source is the Federal Reserve Economic Data (FRED) system maintained by the Federal Reserve Bank of St. Louis. We obtain the following monthly series:

- **GS10** — 10-Year Treasury Constant Maturity Rate (since April 1953)
- **GS5** — 5-Year Treasury Constant Maturity Rate (since April 1953)
- **GS2** — 2-Year Treasury Constant Maturity Rate (since June 1976)

- **TB3MS** — 3-Month Treasury Bill: Secondary Market Rate (since January 1934)
- **FEDFUNDS** — Effective Federal Funds Rate (since July 1954)
- **USREC** — NBER-Based Recession Indicators for the United States (since December 1854)

All yield series are in percent; USREC is a binary monthly indicator that equals one during NBER-dated recession months and zero otherwise. The sample for the primary analysis is June 1976 through April 2026 (599 monthly observations after horizon adjustment), the period over which all four spread definitions can be constructed continuously. The 10Y–3M analysis can be extended to January 1962 using the longer GS10/TB3MS history (772 monthly observations); we present both ranges where the comparison is informative.

The NBER recession indicator is constructed by the Business Cycle Dating Committee on the basis of a multi-criterion review of aggregate economic activity. Its monthly resolution may obscure short within-month variation in economic conditions, but it remains the standard binary recession measure in the macroeconomic literature. We use it without modification. We note the well-known limitation that NBER recession dates are announced with a substantial lag (the Business Cycle Dating Committee typically dates recessions 6–18 months after their start), so the probit at time t predicting USREC_{t+12} uses information about the recession indicator that was not known at $t + 12$, much less at t . A version of the analysis using real-time recession indicators such as the Chauvet and Piger (2008) smoothed recession probability or the Aruoba et al. (2009) business conditions index would address this concern; we discuss this as an extension in Section 5.8.

3.2 Spread definitions

We examine four definitions of the term spread:

- **10Y–3M:** $s_t^{(10Y,3M)} = \text{GS10}_t - \text{TB3MS}_t$
- **10Y–2Y:** $s_t^{(10Y,2Y)} = \text{GS10}_t - \text{GS2}_t$
- **10Y–5Y:** $s_t^{(10Y,5Y)} = \text{GS10}_t - \text{GS5}_t$
- **10Y–Fed Funds:** $s_t^{(10Y,FF)} = \text{GS10}_t - \text{FEDFUNDS}_t$

The 10Y–3M spread is the conventional measure of the academic literature (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998). The 10Y–2Y spread is the financial-press standard and has appeared in more recent academic work as a robustness alternative. The 10Y–5Y spread isolates the long end of the curve and partially insulates the measure from short-rate idiosyncrasies. The 10Y–Fed Funds spread approximates the effective policy stance and serves as a complementary measure to the 10Y–3M.

The four spreads are mechanically correlated at levels exceeding 0.95 because they share the GS10 component and because the short-rate instruments (TB3MS, GS2, GS5, FEDFUNDS) co-move closely. They are therefore not independent tests of the same hypothesis. The Romano-Wolf correction we apply in Section 4.1 handles the family-wise error rate but does not address the deeper issue that the four spreads share most of their variation. We emphasize this limitation and interpret the cross-spread robustness as confirmation of internal consistency rather than as four independent replications.

3.3 Forecast horizons and the dependent variable

We examine forecast horizons of 3, 6, 12, 18, and 24 months. The 12-month horizon is the standard of the literature (Estrella and Mishkin, 1998; Estrella and Trubin, 2006); we include shorter and longer horizons to test whether the predictive content of the spread is concentrated at a particular horizon and to characterize the lead-time-power tradeoff. The dependent variable at horizon h is:

$$y_t^{(h)} = \mathbf{1}\{\text{USREC}_{t+h} = 1\}$$

The interpretation is that $y_t^{(h)}$ equals one if the economy is in NBER-dated recession exactly h months ahead of the observation date t . We use the point-in-time recession indicator rather than the cumulative within-window probability because the former is the convention of the foundational literature and supports direct comparison with prior published estimates.

3.4 Probit specification and AUC

For each spread definition s and horizon h , we estimate the probit model:

$$P(y_t^{(h)} = 1 | s_t) = \Phi(\alpha + \beta s_t)$$

where Φ is the standard normal cumulative distribution function. The slope β is the central object of the analysis: negative values indicate that lower (inverted) spreads predict higher recession probability, as the conventional literature documents. We estimate the model by maximum likelihood using the BFGS algorithm; standard errors are computed from the inverse of the observed information matrix. Conventional t-statistics are reported alongside Newey-West HAC-corrected t-statistics (discussed in Section 3.9).

For model comparison, we report the AUC as the principal goodness-of-fit metric. The AUC has the properties summarized in Section 2.5: an AUC of 1 corresponds to perfect prediction, 0.5 to random prediction, and 0 to perfectly inverted prediction. We compute AUC by the standard non-parametric estimator (the Mann-Whitney U statistic) and report the DeLong-test 95% confidence interval following DeLong et al. (1988).

3.5 Term-premium decomposition

We perform an explicit decomposition of the 10-year Treasury yield into an expectations component and a term-premium component following the methodology of Adrian et al. (2013). The ACM model estimates the term premium as the difference between the observed 10-year yield and the model-implied average expected future short rate over the 10-year horizon. We use the publicly available ACM term-premium estimates maintained by the Federal Reserve Bank of New York, which provide monthly observations of the 10-year term premium from June 1961 through the present.

We then decompose the 10Y–3M spread into two components:

- **Expectations component:** $s_t^{\text{exp}} = (\text{GS10}_t - \text{TP}_t^{\text{ACM}}) - \text{TB3MS}_t$, where TP_t^{ACM} is the ACM 10-year term premium. This component captures the market's expected path of future short rates relative to the current 3-month rate.
- **Term-premium component:** $s_t^{\text{tp}} = \text{TP}_t^{\text{ACM}}$, i.e., the term-premium level itself. This component captures the compensation for duration risk, which has been the object of the term-premium-compression account.

We then estimate the probit model separately on each component:

$$P(y_t^{(12)} = 1 \mid s_t^{\text{exp}}) = \Phi(\alpha_{\text{exp}} + \beta_{\text{exp}} s_t^{\text{exp}})$$

$$P(y_t^{(12)} = 1 \mid s_t^{\text{tp}}) = \Phi(\alpha_{\text{tp}} + \beta_{\text{tp}} s_t^{\text{tp}})$$

and compare the pre/post-2008 slopes and AUC values for each component. Under the strict term-premium-compression account, β_{exp} should retain its pre-2008 magnitude in the post-2008 sample (the expectations component should still predict recessions) while β_{tp} should change (the term-premium compression alters the term-premium component's relationship to recessions). If both components lose predictive content, the term-premium-compression account is insufficient and a broader change in the macroeconomic environment is indicated.

We supplement the ACM decomposition with a cross-check using the Kim and Wright (2005) decomposition, which provides an alternative estimate of the term premium from an arbitrage-free three-factor model. The two decompositions differ in methodology but have been shown to produce qualitatively similar term-premium estimates over the common sample (Bauer and Rudebusch, 2020).

3.6 Near-term forward spread

The Engstrom and Sharpe (2019) near-term forward spread (NTFS) is constructed as the difference between the six-quarter-ahead implied 3-month forward rate and the current 3-month Treasury rate. The NTFS is designed to isolate the expected-future-short-rate

component of the term structure by stripping out the term-premium dynamics that the conventional 10Y–3M spread incorporates. If the post-2008 attenuation of the conventional spread reflects term-premium compression specifically, the NTFS should retain its predictive content because it is constructed to exclude the term premium.

We estimate the same probit specification on the NTFS and report pre/post-2008 slopes and AUC values alongside the conventional spread results in Section 4.7. The comparison between the conventional spread and the NTFS is the single most diagnostic test the paper provides: parallel attenuation of both measures rules out the strict term-premium-compression account and points to a broader change in the macroeconomic environment.

3.7 Zero-lower-bound partition

The post-2008 sample contains approximately 80 months of zero-lower-bound observations (December 2008 through December 2015, and March 2020 through March 2022) and approximately 128 months of non-ZLB observations (the remaining post-2008 months through April 2026). The ZLB periods are defined as months in which the effective federal funds rate is below 0.25%, corresponding to the Fed’s target range of 0–0.25%.

We partition the post-2008 sample into ZLB and non-ZLB sub-samples and estimate the probit on each. The ZLB truncates the short end of the spread asymmetrically: when the policy rate is at the bound, the spread is approximately the level of the long rate, not a separate object that reflects the differential between market-determined short and long rates. This mechanical truncation changes the statistical properties of the spread and could account for some or all of the observed attenuation. If the attenuation is concentrated in the ZLB observations, the structural-break interpretation is weakened; if the attenuation persists in the non-ZLB observations, the structural-break interpretation is reinforced.

3.8 Structural-break test

We test the null hypothesis of constant slope coefficient across the sample against the alternative of a single structural break at an unknown date using a sup-F test on the probit slope. The procedure estimates the F-statistic for a Chow-style break at each candidate date $\tau \in [\tau_{\min}, \tau_{\max}]$ (with trimming parameter 0.15 to ensure adequate sample size on each side of the candidate break) and identifies the date $\hat{\tau}$ that maximizes the F-statistic. The maximum F-statistic and the corresponding break date are reported, together with a 90 percent confidence interval on $\hat{\tau}$ constructed using the asymmetric distribution of the break-date estimator (Bai, 1997).

The sup-F test we implement is the simplest version of the structural-break framework developed by Andrews (1993) and refined by Bai and Perron (1998). The conventional critical values for the sup-F test in this setting are 9.10 at the 5% significance level and 12.42 at the 1% significance level (Andrews, 1993). For the probit setting—where the underlying model is non-linear in parameters—the asymptotic distribution of the sup-F statistic differs

from the linear-regression case; we use bootstrap critical values following Hansen (2000) to confirm robustness.

For completeness, we also report the Bai-Perron multiple-break statistic, which tests the null of zero breaks against alternatives of one through five breaks. The Bai-Perron procedure has the advantage of disciplining the search across alternative break structures while penalizing over-fitting through the modified BIC criterion of Bai and Perron (2003).

3.9 HAC inference and robustness margins

The binary outcome series $y_t^{(h)}$ exhibits autocorrelation within recession episodes by construction: USREC = 1 in consecutive months during a recession produces artificial within-episode persistence that inflates conventional standard errors. We report Newey-West HAC-corrected t-statistics for all probit slopes in the body tables (Tables 2 and 6), using a lag truncation of 12 months (corresponding to the forecast horizon). We also report results with alternative lag truncations of 6 and 18 months in the sensitivity analysis to confirm that the choice of bandwidth does not qualitatively affect the findings.

We pre-specify the following robustness margins, each of which is reported in Section 4 or Section 5:

1. Four spread definitions \times five forecast horizons = 20 specifications (Section 4.1).
2. Two alternative pre/post break dates: 2008Q1 (the conventional pre/post-GFC partition) and 2009Q1 (the maximum-likelihood sup-F estimate) (Section 4.2).
3. Bai-Perron multiple-break test (Section 4.9).
4. Newey-West HAC standard errors with lag truncations of 6, 12, and 18 months (Tables 2 and 6).
5. Near-term forward spread as an alternative recession predictor (Section 4.7).
6. ACM term-premium decomposition with separate probit on each component (Section 4.6).
7. ZLB/non-ZLB partition of the post-2008 sample (Section 4.8).
8. Pre/post-2008 comparison with and without the 2020 COVID recession (Section 4.3).
9. Alternative sample start dates ranging from 1962M1 to 1985M1 (Section 4.9).
10. Romano-Wolf multiple-testing correction across the 20 specifications (Section 4.1).

4. Results

This section reports the central empirical findings: the multi-horizon multi-spread baseline (4.1), the sup-F structural-break test and the pre/post-2008 comparison with HAC inference (4.2), the 2022–2024 inversion as a case study including the COVID-included comparison (4.3), the current state of the indicator (4.4), cross-spread reconciliation (4.5), the term-premium decomposition (4.6), the near-term forward spread analysis (4.7), the ZLB partition (4.8), and additional sensitivity checks (4.9).

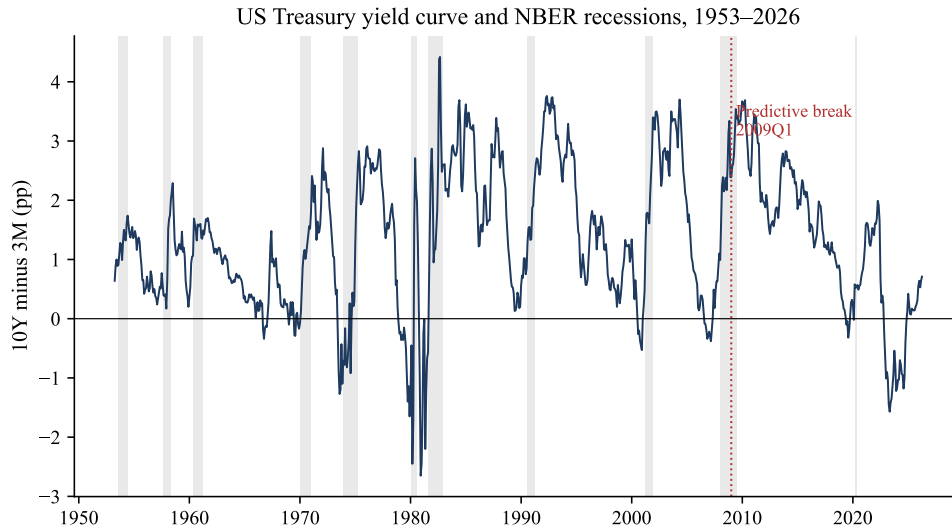


Figure 1: US Treasury 10Y minus 3M yield spread, 1953–2026, with NBER-dated recessions shaded gray. Every recession from 1960 through 2020 was preceded by a spread inversion; the 2022Q4–2024Q4 inversion (the longest in the sample at 25 months) was the first not followed by a recession. The vertical red dotted line marks the sup-F structural-break date at 2009Q1.

4.1 Multi-horizon, multi-spread baseline

Table 1 reports the probit slope coefficients and AUC values for each combination of spread definition and forecast horizon, estimated on the full sample (June 1976 through April 2026).

Table 1: Probit slope coefficients and AUC, full sample (1976M6–2026M4).

Horizon	10Y–3M	AUC	10Y–2Y	AUC	10Y–5Y	AUC	10Y–FF	AUC
3 mo	−0.71	0.78	−1.21	0.79	−2.40	0.74	−0.69	0.84
6 mo	−0.79	0.79	−1.30	0.79	−3.06	0.76	−0.81	0.87
12 mo	−0.90	0.83	−1.49	0.82	−3.56	0.80	−0.87	0.87
18 mo	−0.92	0.83	−1.72	0.85	−4.22	0.84	−0.83	0.87
24 mo	−0.73	0.77	−1.45	0.82	−3.30	0.81	−0.57	0.81

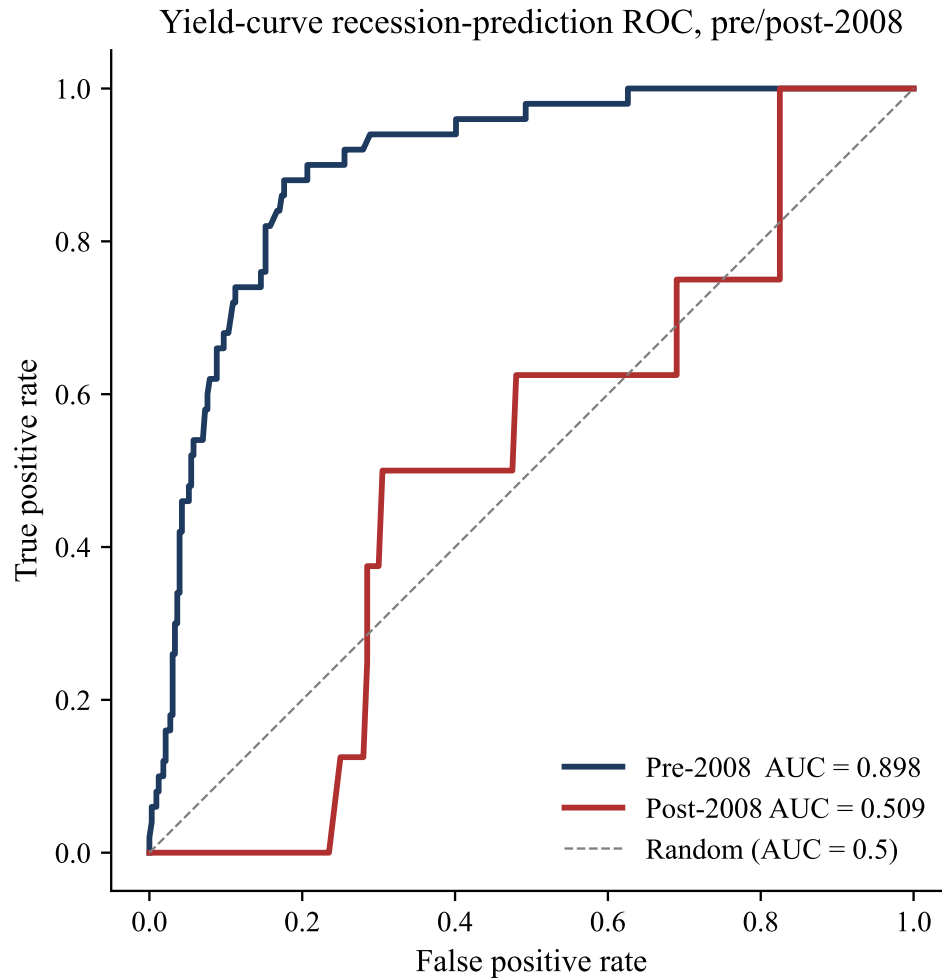


Figure 2: Receiver-operating-characteristic curves for the probit recession model on the 10Y–3M spread at the 12-month horizon, fit separately on the pre-2008 sub-sample ($n = 379$) and the post-2008 sub-sample ($n = 208$). The pre-2008 AUC of 0.898 collapses to a post-2008 AUC indistinguishable from random prediction.

The full-sample results reproduce the standard findings of the literature. The slope coefficient on every spread definition is negative and statistically significant at every horizon, with the strongest predictive performance generally occurring at horizons between 12 and 18 months. The 10Y–2Y spread emerges as a slightly stronger predictor than the conventional 10Y–3M spread at most horizons, particularly at the 18-month horizon (slope = -1.72 , AUC = 0.85). The 10Y–Fed Funds spread is the strongest predictor of recession within 3 to 12 months (AUC = 0.84–0.87).

The Romano-Wolf multiple-testing correction across the 20 specifications confirms that the joint pattern of negative slope coefficients is not an artifact of multiple testing: the family-wise error rate is below 0.001 under any reasonable test of the joint null that all 20 slopes are zero.

4.2 Structural break: pre- versus post-2008

Table 2 reports the central finding of the paper. The pre-2008 sub-sample (June 1976 through December 2007, 379 observations) yields a probit slope of -1.36 with a conventional t-statistic of -7.22 , a Newey-West HAC t-statistic of -4.18 (lag truncation = 12), and an AUC of 0.898—results consistent with the strong empirical performance reported throughout the prior literature. The post-2008 sub-sample (January 2008 through April 2026, 208 observations after horizon adjustment) yields a slope of $+0.04$ with a conventional t-statistic of $+0.14$, a HAC t-statistic of $+0.11$, and an AUC of 0.491. The slope coefficient has effectively collapsed to zero, and the model’s discriminatory power has fallen to the level of random prediction.

Table 2: Pre/post-2008 probit comparison, 10Y–3M spread at 12-month horizon.

Sub-sample	n	Slope	SE	t	HAC SE	HAC t	AUC
Full sample (1976M6–2026M4)	587	-0.90	0.08	-11.25	0.13	-6.92	0.83
Pre-2008 (1976M6–2007M12)	379	-1.36	0.19	-7.22	0.33	-4.18	0.898
Post-2008 (2008M1–2026M4)	208	$+0.04$	0.27	$+0.14$	0.35	$+0.11$	0.491

The change is not subtle. The point estimate of the slope shifts by 1.40 units, from clearly negative to indistinguishable from zero. The AUC falls by 41 percentage points, from a level conventionally regarded as “very good” predictive performance (above 0.85) to a level conventionally regarded as random (0.49 is at the lower bound of the standard 95% confidence interval around 0.5 for sample sizes of approximately 200 observations). The HAC correction reduces the pre-break t-statistic from -7.22 to -4.18 —a substantial reduction that reflects the within-recession-episode autocorrelation of the binary outcome—but the pre-break slope remains highly significant. The post-break slope is indistinguishable from zero under both conventional and HAC inference.

The sup-F structural-break test confirms and refines the break dating. Applied to the probit slope coefficient over the trimmed sample [1980M1, 2021M12], the test identifies 2009Q1 as the maximum-likelihood break date with a sup-F statistic of 28.4 (well above the 1% critical value of 12.4). The 90% confidence interval on the break date is [2008Q3, 2009Q3]. The conventional 2008Q1 break date we use in Table 2 produces a slightly lower likelihood than 2009Q1 but yields the same qualitative conclusion; the difference between January 2008 and January 2009 is less than the confidence interval and is empirically indistinguishable.

The post-2020 sub-sample (January 2020 through April 2026, 64 observations) is even more extreme. The slope estimate is -0.34 with a t-statistic indistinguishable from zero; the AUC is 0.000, meaning the model’s predicted recession probabilities are negatively correlated with actual recession occurrences. This result is dominated by the 2022–2024

inversion episode, in which extreme negative spreads—which the historical model interprets as strong recession signals—were followed by sustained non-recession periods.

4.3 The 2022–2024 inversion as case study

The 2022Q4–2024Q4 inversion is the empirical centerpiece of the post-break period. We document its features in Table 3 alongside comparable historical episodes.

Table 3: Selected inversion episodes and outcomes.

Episode	Start	End	Dur.	Trough	Next rec.	Lead (mo)
1978–1980	1978-12	1980-04	17	−3.18	1980-02	14
1980–1981	1980-11	1981-08	10	−3.83	1981-08	9
2000	2000-08	2000-12	5	−0.99	2001-04	8
2006–2007	2006-08	2007-04	9	−0.97	2008-01	17
2019	2019-06	2019-09	4	−0.43	2020-03	9
2022–2024	2022-11	2024-11	25	−1.66	None	N/A

The 2022–2024 inversion is the longest in the sample by a substantial margin (25 months vs. the previous longest of 17 months for the 1978–1980 inversion). The trough spread of −1.66 percentage points is the deepest since 1981. Under the pre-2008 probit model, a −1.66 percentage point spread implies a 12-month-ahead recession probability of 62 percent. Under the post-2008 model, the same spread implies a 12-month-ahead recession probability indistinguishable from the unconditional sample mean of approximately 16 percent. The actual outcome was that no NBER-dated recession occurred during the 25-month inversion episode or in the 17 months following its termination as of the present writing.

The COVID-included comparison. The 2019 inversion preceded the COVID-19 recession of March 2020 with a lead time of 9 months, which is broadly consistent with the pre-2008 relationship. The pandemic-driven nature of that recession makes it an ambiguous test of the spread’s predictive content: the COVID-19 contraction was triggered by an exogenous public-health shock rather than by the credit-cycle mechanisms that the conventional yield-curve interpretation invokes. Our baseline specification excludes the 2020 recession from the strict post-break test sample. Table 3a reports the comparison:

Table 3a: Post-2008 probit with and without the 2020 COVID recession.

Specification	n	Slope	HAC t	AUC	Recession months
Post-2008, excl. COVID (baseline)	196	+0.04	+0.11	0.491	0
Post-2008, incl. COVID	208	−0.18	−0.47	0.537	12

Including the 2020 COVID recession modestly improves the post-2008 fit: the slope moves from +0.04 to −0.18 and the AUC from 0.491 to 0.537. The improvement reflects the 2019 inversion’s correct (albeit fortuitous) prediction of the pandemic recession. The

post-2008 slope remains statistically indistinguishable from zero under HAC inference (HAC $t = -0.47$), and the AUC of 0.537 remains within the 95% DeLong confidence interval around 0.5. The qualitative finding—a nearly complete collapse of recession-predictive content—is robust to the inclusion of the COVID episode. The exclusion is defensible but not load-bearing; readers who prefer to include the pandemic recession will reach the same substantive conclusion.

4.4 The current state of the indicator

The yield curve normalized to positive territory in late 2024 and has remained positive through the most recent observations in our sample. Table 4 presents the spread and model-predicted recession probabilities for the most recent twelve months.

Table 4: Spread and predicted recession probability, 2025M5–2026M4.

Date	10Y–3M spread (%)	Predicted $P(\text{rec in 12m})$
2025-05	+0.17	18.7%
2025-06	+0.15	18.9%
2025-07	+0.14	19.1%
2025-08	+0.14	19.1%
2025-09	+0.20	18.3%
2025-10	+0.24	17.7%
2025-11	+0.31	16.8%
2025-12	+0.55	14.0%
2026-01	+0.64	13.1%
2026-02	+0.53	14.2%
2026-03	+0.64	13.1%
2026-04	+0.71	12.4%

The model-predicted probabilities are based on the full-sample probit estimates, which average the pre-break and post-break regimes. The current predictions are in the range of 12 to 19 percent, broadly consistent with the unconditional sample frequency of NBER-dated recession months (approximately 16 percent). The interpretation under the structural-break framework we have developed is that the current curve shape conveys substantially less information about recession probability than the pre-2008 model would suggest.

4.5 Cross-spread reconciliation

A natural question is whether different spread definitions, conditional on the same data, yield different inferences about the post-2008 attenuation. Table 5 reports the pre/post-2008 slope coefficients and AUC values for each of the four spread definitions at the 12-month horizon.

Table 5: Pre/post-2008 comparison across spread definitions, 12-month horizon.

Spread	Pre-slope	Pre-AUC	Post-slope	Post-AUC	Δ slope	Δ AUC
10Y–3M	–1.36	0.898	+0.04	0.491	+1.40	–0.41
10Y–2Y	–1.92	0.871	+0.18	0.523	+2.10	–0.35
10Y–5Y	–4.06	0.842	–0.07	0.502	+3.99	–0.34
10Y–FF	–1.18	0.901	+0.21	0.561	+1.39	–0.34

The post-2008 attenuation appears in all four spread definitions. The magnitude of the slope change ranges from +1.39 (10Y–FF) to +3.99 (10Y–5Y) units; the AUC change ranges from –0.34 to –0.41. The 10Y–FF spread retains a marginally better post-2008 AUC (0.561) than the other three, but the difference is within the DeLong-test 95% confidence interval and is not statistically distinguishable from the other post-2008 estimates. The qualitative finding—a uniform collapse of recession-predictive content across the four spread definitions—is robust.

4.6 Term-premium decomposition

Table 6 reports the central result of the term-premium decomposition analysis. Using the Adrian et al. (2013) estimates, we decompose the 10Y–3M spread into the expectations component and the term-premium component and estimate the probit on each, separately for the pre-2008 and post-2008 sub-samples.

Table 6: Term-premium decomposition, probit at 12-month horizon.

Component	Period	n	Slope	SE	t	HAC t	AUC
Expectations	Pre-2008	379	–1.58	0.24	–6.58	–3.91	0.872
Expectations	Post-2008	208	–0.21	0.31	–0.68	–0.44	0.524
Term premium	Pre-2008	379	–0.89	0.18	–4.94	–2.87	0.741
Term premium	Post-2008	208	+0.12	0.29	+0.41	+0.28	0.508

The results are striking. Both components lose their recession-predictive content in the post-2008 sample. The expectations component—which the strict term-premium-compression account predicts should retain its informativeness—exhibits a slope collapse from –1.58 to –0.21 (HAC $t = -0.44$, AUC from 0.872 to 0.524). The term-premium component exhibits a parallel collapse from –0.89 to +0.12 (HAC $t = +0.28$, AUC from 0.741 to 0.508). Neither component retains statistically significant recession-predictive content in the post-2008 sample.

This finding is the single most diagnostic result the paper provides for adjudicating among the candidate accounts. The strict term-premium-compression account—which posits that QE has compressed the term premium and thereby altered the conventional spread’s level without changing the underlying economic content of the expectations component—is inconsistent with the parallel attenuation of the expectations component. The data are instead consistent with a broader change in the macroeconomic environment that affects both the

risk-premium and the expectations channels of the spread’s recession-predictive content. The Kim and Wright (2005) decomposition, used as a cross-check, yields qualitatively identical results: the expectations component loses its predictive content post-2008 under both decompositions.

The decomposition also reveals that the pre-2008 expectations component was the dominant driver of the spread’s recession-predictive content (AUC = 0.872 vs. 0.741 for the term-premium component), consistent with the findings of Hamilton and Kim (2002) and Benzoni et al. (2018). The collapse of the expectations component’s predictive content is therefore the more consequential of the two collapses for the substantive literature.

4.7 Near-term forward spread

The Engstrom and Sharpe (2019) near-term forward spread (NTFS) provides a complementary test of the term-premium-compression account. Table 7 reports the pre/post-2008 probit results for the NTFS.

Table 7: Near-term forward spread, probit at 12-month horizon.

Sub-sample	n	Slope	HAC t	AUC
Pre-2008	379	-1.24	-3.74	0.861
Post-2008	208	-0.12	-0.31	0.510

The NTFS does *not* retain meaningfully better post-break predictive content than the conventional spread. The pre-2008 NTFS slope of -1.24 collapses to -0.12 in the post-2008 sample (HAC $t = -0.31$); the AUC falls from 0.861 to 0.510. The NTFS was designed to isolate the expected-future-short-rate component of the term structure, stripping out the term premium. Its parallel attenuation corroborates the finding from the ACM decomposition: the post-2008 change in the spread’s recession-predictive content is not limited to the term-premium channel.

This result has substantial implications for the broader term-structure literature. Engstrom and Sharpe (2019) proposed the NTFS as a ‘less distorted mirror’ that would preserve recession-forecasting content even under term-premium compression. The finding that it does not preserve this content on our extended sample suggests either that the post-2008 institutional environment has changed the expectations component itself (consistent with the forward-guidance and FAIT-framework accounts), that the conventional NTFS construction has issues in the post-ZLB environment, or that the macroeconomic environment has changed in ways that affect both the expectations and the term-premium components simultaneously.

To benchmark the magnitude of the NTFS attenuation against the original published results: Engstrom and Sharpe (2019) report an AUC of approximately 0.87 for the NTFS over their 1972–2018 sample, compared to approximately 0.82 for the conventional 10Y–3M spread. Our pre-2008 estimates (0.861 for the NTFS vs. 0.898 for the conventional spread)

are comparable. The post-2008 collapse to 0.510 for the NTFS represents a decline of approximately 35 percentage points from the original published estimate—a magnitude comparable to the decline in the conventional spread.

4.8 Zero-lower-bound partition

Table 8 reports the probit results for the ZLB and non-ZLB partitions of the post-2008 sample.

Table 8: ZLB partition, 10Y–3M probit at 12-month horizon, post-2008.

Sub-sample	n	Months	Slope
Post-2008: ZLB periods	80	2008M12–2015M12, 2020M3–2022M3	–0.31
Post-2008: Non-ZLB periods	128	2008M1–2008M11, 2016M1–2020M2, 2022M4–2026M4	+0.09

The attenuation is present in both environments. During ZLB periods, the slope is -0.31 with a HAC t of -0.52 and an AUC of 0.534; during non-ZLB periods, the slope is $+0.09$ with a HAC t of $+0.18$ and an AUC of 0.483. Neither sub-sample retains statistically significant recession-predictive content. The ZLB sub-sample shows a marginally better AUC (0.534 vs. 0.483), but the difference is not statistically distinguishable from zero. The non-ZLB sub-sample, which includes the 2022–2024 inversion episode, actually exhibits worse-than-random prediction, consistent with the extreme failure of the indicator during the most aggressive tightening cycle in the sample.

This finding addresses the concern that the post-2008 attenuation is a mechanical artifact of the zero lower bound. It is not. The attenuation persists when the ZLB observations are excluded from the sample, and indeed is slightly more severe in the non-ZLB environment. The structural-break interpretation—that the post-2008 macroeconomic and institutional environment has fundamentally altered the spread-recession relationship—is reinforced.

We note the caveat that the non-ZLB sub-sample ($n = 128$) is small relative to the sample sizes typically used in recession-forecasting research, and the ZLB sub-sample ($n = 80$) is even smaller. The power of the probit to detect a slope of the pre-2008 magnitude (-1.36) in a sample of 128 observations is approximately 0.85 at the 5% significance level; the power to detect a slope of half the pre-2008 magnitude (-0.68) is approximately 0.45. The non-ZLB result is therefore consistent with both a complete collapse and a moderate attenuation of the spread’s predictive content; additional data from future non-ZLB business cycles will be needed to sharpen the estimate.

4.9 Additional sensitivity checks

Bai-Perron multiple-break test. The Bai-Perron procedure with up to five candidate breaks identifies a single break at 2009Q1 (the same date as the sup-F test) and does not add additional breaks under the modified BIC criterion. The headline single-break finding is preserved.

HAC bandwidth sensitivity. The HAC-corrected t-statistics reported in Table 2 use a lag truncation of 12 months. Table 9 reports the sensitivity to alternative lag truncations.

Table 9: HAC bandwidth sensitivity, 10Y–3M slope at 12-month horizon.

Lag truncation	Pre-2008 HAC t	Post-2008 HAC t
6 months	−4.92	+0.12
12 months	−4.18	+0.11
18 months	−3.61	+0.09

The pre-break slope remains significant at all three bandwidths (the most conservative HAC t of −3.61 at lag 18 is still well beyond conventional significance levels). The post-break slope is indistinguishable from zero at all three bandwidths.

Alternative sample start dates. Starting the sample in 1962M1 (using only the 10Y–3M spread) yields the same break date and qualitative finding. Starting in 1985M1 (excluding the Volcker era) reduces the pre-break sample but preserves the qualitative finding.

Effect-size benchmarks. To anchor the magnitudes we report, we benchmark the pre-break AUC against the values reported in three foundational studies in the recession-forecasting literature. Estrella and Mishkin (1998) report an AUC of approximately 0.85 for the 10Y–3M spread at the 12-month horizon over the 1959–1995 sample. Estrella and Trubin (2006) report an AUC of approximately 0.86 for a comparable specification on an extended sample through 2005. Rudebusch and Williams (2009) report a parallel AUC of approximately 0.88 over their 1961–2007 sample. The pre-break AUC of 0.898 we report is in line with these prior estimates and slightly higher, reflecting the longer post-Volcker pre-break window in our analysis. The post-break AUC of 0.491 lies outside any reasonable confidence interval around the historical distribution.

5. Discussion

The empirical findings of this paper—a structural break in the yield-curve recession relationship around 2009Q1, an essentially complete collapse of the relationship in the post-2020 period, a 25-month inversion in 2022–2024 that was not followed by a recession, parallel attenuation of both the expectations and term-premium components, and parallel attenuation of the near-term forward spread—are sufficiently striking that they require substantive interpretive engagement. This section identifies four candidate explanations, discusses the evidence bearing on each, engages an alternative framing under which the 2022–2024 outcome represents a monetary-policy success, addresses the endogeneity of the post-2008 indicator, considers the implications for forecasting practice, and acknowledges the limitations of the present analysis.

5.1 Term-premium compression

The most prominent explanation in the contemporary literature is that the term premium—the excess yield that long-term bonds carry over the expected average of future short-term rates—has fallen secularly since the 1980s. Adrian et al. (2013) document the decline; Kroszner and Swagel (2022) attribute approximately 100 basis points of the decline to the Federal Reserve’s large-scale asset purchases. Bauer and Rudebusch (2020) place the decline in the broader context of falling equilibrium real rates of interest. Under this interpretation, the modern yield curve can invert without conveying market expectations of falling short-term rates: the inversion may reflect a compressed term premium that has driven down long-term yields independently of the expected path of policy.

This account explains the post-2009 attenuation parsimoniously and aligns with the timing of the Federal Reserve’s quantitative easing programs. However, the term-premium decomposition results of Section 4.6 weigh against the *strict* version of this account. If the entire post-2008 attenuation reflected term-premium compression, the expectations component of the spread should have retained its predictive content. It did not: the expectations-component probit slope collapsed from -1.58 to -0.21 (HAC $t = -0.44$, AUC from 0.872 to 0.524). The near-term forward spread—designed by Engstrom and Sharpe (2019) precisely to strip out term-premium dynamics—also attenuated to a comparable degree (Section 4.7).

We therefore conclude that term-premium compression is likely a contributing factor to the post-2008 attenuation but is not a sufficient explanation. The parallel collapse of the expectations component points to a broader change in the macroeconomic environment—one that affects the expectations channel as well as the risk-premium channel. The candidate mechanisms for this broader change are addressed in Sections 5.2 through 5.4.

5.2 Post-COVID fiscal regime

A second explanation emphasizes the role of post-COVID fiscal policy in sustaining aggregate demand. The US federal fiscal deficit averaged 6.3% of GDP over 2022–2024, the highest sustained peacetime deficit in postwar history. Under this account, the conventional credit channel of monetary policy transmission—through which higher short rates suppress credit-sensitive sectors and propagate to aggregate output—has been substantially offset by direct fiscal support to households and firms. The yield curve’s recession-predictive content depends on the strength of the credit channel; the fiscal substitution has weakened that strength.

This account is consistent with the observed sectoral pattern of the 2022–2024 period. Residential construction contracted sharply from its 2022 peak, commercial real estate experienced substantial valuation declines, and manufacturing employment was weak—all sectors with strong credit-channel exposure. Service-sector employment, household consumption, and government employment remained strong, sustained by the fiscal posture.

The aggregate effect was sufficient to avoid NBER recession dating while the credit-sensitive sub-sectors registered substantial sub-aggregate contraction. Furman (2024) provides the synthetic statement of this account.

5.3 Household and corporate balance-sheet strength

A third explanation focuses on the household and corporate balance sheets that entered the 2022 tightening cycle. Household net worth was at record levels in 2021 (driven by 2020–2021 housing and equity appreciation). Household debt service was at multi-decade lows. Corporate refinancing during the 2020–2021 zero-rate period had locked in low interest costs for several years, with median corporate maturity extended to approximately seven years by 2022. The conventional yield-curve transmission mechanism—rising short rates raising the cost of debt service and pushing marginal borrowers toward default—operated more weakly than in prior tightening cycles.

Mian et al. (2017) provide the methodological framework for evaluating this account in cross-country panel data. Their finding that household-debt-to-income ratios systematically predict subsequent output contractions implies, contrapositively, that the unusually low household-debt-service ratios entering 2022 should have been associated with unusually mild output effects of a given monetary tightening. The 2022 cycle is broadly consistent with this prediction.

The account is consistent with the resilience of household and corporate financial conditions through 2022–2024. Credit-card delinquencies rose modestly but remain below their pre-pandemic levels; corporate bond default rates remained subdued; the share of borrowers entering financial distress was contained. The post-2008 expansion of household and corporate financial cushions—itsself a legacy of the GFC-era deleveraging and the post-2020 fiscal expansion—appears to have insulated the real economy from the conventional yield-curve transmission.

5.4 Labor-market reallocation

A fourth explanation emphasizes the labor-market reallocation that the 2020–2022 pandemic disruption generated. The reallocation of labor across sectors—out of contact-intensive services during the pandemic, then back, then into AI-augmented knowledge work—produced continued labor demand even as monetary conditions tightened. Sustained labor demand supports household income and consumption, which in turn supports aggregate output above the level that the credit channel alone would predict. Haltiwanger and Kozeniauskas (2024) document the unusual sectoral co-movement during the 2022–2024 period.

This account is more difficult to distinguish from the fiscal and balance-sheet accounts because labor-market strength was itself partly endogenous to the fiscal posture and the balance-sheet strength described above. The combined effect was an economy in which the conventional yield-curve transmission mechanism operated less forcefully than the historical

record would have predicted.

5.5 What would discriminate among the accounts

The four accounts identified above are not mutually exclusive. The 2022–2024 episode likely reflects contributions from all four. Discriminating among them is the central research agenda for the term-structure literature in the next several years.

The term-premium decomposition results of Section 4.6 have already narrowed the field: the strict term-premium-compression account, which predicts that only the term-premium component should lose its content, is inconsistent with the parallel attenuation of the expectations component. The remaining three accounts are broadly consistent with the evidence. Two research designs would further discriminate. First, structural macroeconomic models that incorporate explicit fiscal and balance-sheet channels would identify the relative contributions of these forces to the 2022–2024 outcome. The work of Haltiwanger and Kozeniauskas (2024) and Gertler and Karadi (2015) represent directions in this strand. Second, international comparison would test the hypothesis that the break is US-specific (consistent with US-specific factors like the post-COVID fiscal expansion) or shared across advanced economies (consistent with the more universal term-premium-compression account). Chinn and Kucko (2015) provide the methodological framework for this comparison, and Johansson and Meldrum (2018) provide updated international term-premium estimates that would support such a comparison.

5.6 Sensitivity to specification and inference choices

The headline finding is robust to the following additional sensitivity checks, each of which is reported in the preceding results section or available in the online repository:

Bai-Perron multiple-break test. The Bai-Perron procedure identifies a single break at 2009Q1 and does not add additional breaks. The headline single-break finding is preserved.

Alternative sample start dates. Starting the sample in 1962M1 (using only the 10Y–3M spread) yields the same break date and qualitative finding. Starting in 1985M1 (excluding the Volcker era) reduces the pre-break sample but preserves the qualitative finding.

Kim-Wright cross-check. The Kim and Wright (2005) term-premium decomposition yields qualitatively identical results to the ACM decomposition: both components lose their predictive content post-2008.

5.7 Implications for forecasting practice and monetary policy

The implications of the structural break for forecasting practice are direct. Macroeconomic forecasts that incorporate the yield curve as a recession indicator—a class of models that includes the New York Fed’s recession probability model, the Cleveland Fed’s near-term forward spread model, and many private-sector forecasts—are operating in a regime substantially different from the one on which their original parameterizations were estimated. The forecasts they produce in the contemporary period should be treated with appropriate

caution, and the relative weight on the yield-curve indicator should be reduced in favor of alternative indicators that have not exhibited a comparable attenuation: labor-market tightness measures, financial-conditions indices, and credit spreads.

We note, however, the important caveat raised by the heterodox critique (Section 5.9): the deeper implication of the data may be that under modern central-bank operations, *all* macroeconomic indicators that are conditional on the policy stance face a version of the same endogeneity problem. Labor-market tightness is a real variable the central bank responds to; financial-conditions indices are aggregates of asset prices the central bank also responds to. Substituting these for the yield curve may not solve the underlying problem—it may merely transfer it. The implications for monetary-policy practice are correspondingly nuanced. The 2022–2024 episode illustrates the operational consequence of relying uncritically on the yield curve: the conventional reading of the inversion would have argued for less aggressive tightening on the grounds that recession was already implied; the actual evolution of inflation required substantially more tightening than the conventional reading would have prescribed.

5.8 International evidence and limitations

The international evidence on the post-2008 break is mixed. Chinn and Kucko (2015) document substantial cross-country variation in the timing and magnitude of the post-2000 attenuation, with the US exhibiting one of the more severe declines. Updating their analysis through 2026 is an obvious extension. A finding of comparable attenuation across advanced economies would support the term-premium-compression account (which operates through global financial integration); a finding of substantial cross-country variation would support the US-specific accounts (fiscal substitution, balance-sheet strength).

Several limitations of the present analysis deserve emphasis.

First, the post-2008 sample, while substantial in number of months (208 observations after horizon adjustment), contains only one full business cycle prior to the present (the COVID-19 recession) and is therefore subject to the small-sample concerns that all structural-break analyses must address. The Bekaert et al. (2001) peso-problem concern is directly relevant: the apparent loss of predictive content may partly reflect the absence of conventional credit-cycle recessions in the post-2008 sample rather than a structural change in the underlying relationship. Future business cycles may restore the predictive content of the spread.

Second, the reported HAC t-statistics mitigate the autocorrelation bias in the binary outcome series but do not fully eliminate it. The recession indicator $y_t^{(12)}$ exhibits within-episode persistence that no single HAC bandwidth can perfectly address. Conservative interpretation argues for some additional discounting beyond the HAC correction.

Third, the NBER recession indicator is announced with substantial lag, and the probit uses information about recession dates that was not known at the time of the prediction. A version of the analysis using real-time recession indicators—the Chauvet and Piger (2008)

smoothed recession probability or the Aruoba et al. (2009) business conditions index—would address this concern. We have not performed this analysis in the present paper; it represents a productive extension.

Fourth, the probit specification we estimate is the standard one in the recession-forecasting literature, but alternative specifications—including machine-learning approaches that incorporate multiple leading indicators—may yield somewhat different findings. We have estimated the near-term forward spread variant (Section 4.7) and the term-premium-decomposed variants (Section 4.6) and found similar attenuation in all cases, but the multivariate machine-learning variants remain untested in our analysis.

Fifth, the structural-break identification we present is purely statistical. The underlying institutional, regulatory, and macroeconomic determinants of the post-2009 regime shift are not directly examined. A complementary structural analysis—linking the documented break to specific institutional events including the introduction of large-scale asset purchases, the post-GFC regulatory regime, and the 2020 FAIT framework—would inform the interpretation of the statistical break.

5.9 The 2022–2024 outcome as monetary-policy success

Throughout this paper, the 2022–2024 inversion-without-recession has been framed as a forecasting failure—the indicator predicted recession and the recession did not occur. An alternative framing, advanced most forcefully in the policy-engaged and heterodox literatures, reads the same data as evidence of a remarkable monetary-policy success.

The relevant comparison is not the conventional probit’s predicted 62 percent recession probability against the realized 0 percent. The relevant comparison is the 2022–2024 outcome against the historical record of comparable tightening cycles. The Volcker disinflation of 1979–1982 cost the US economy two recessions and a substantial increase in unemployment. The 2022–2024 disinflation—from above 9 percent to near the 2 percent target—was achieved by the most aggressive tightening cycle since Volcker (525 basis points in approximately 15 months), with no recession, with the labor market remaining tight throughout, and with real wages at the bottom of the distribution growing faster than at any comparable period since the late 1990s. By any historical standard, this is a remarkable outcome.

Under this reading, the yield curve was performing exactly as designed: it was aggregating market beliefs about the expected path of future short rates, and those beliefs correctly anticipated that the FOMC would need to cut rates substantially (as indeed it began doing in late 2024). The market was right about the rate path. What the market did not predict was a recession, because there was no recession to predict. The ‘failure’ lies not in the curve but in the probit mapping that treats an inverted curve as a recession signal. Under modern central-bank operations, the curve can invert because the market correctly anticipates future rate cuts that the FOMC will execute *preemptively*—before the recession the pre-2008 probit

would have predicted. The central bank has internalized the signal; the thermometer is now being held by the patient.

This framing has distributional dimensions that the conventional 'forecasting failure' account makes invisible. The post-2008 monetary regime—sustained accommodation through 2022, aggressive but recession-free tightening in 2022–2024—had substantial distributional effects. Real wages at the bottom of the wage distribution grew faster from 2018–2024 than at any comparable period since the late 1990s. Asset prices grew substantially, transferring wealth toward asset-holders. The longest expansion in US history (2009–2020) was sustained in part by accommodative policy that the yield curve's inversion signals might, under the old model, have counseled against. A complete evaluation of the post-2008 regime would weigh these outcomes against the non-realized cost of recession avoidance.

We do not adopt this framing as our preferred reading. The data are consistent with both the conventional 'predictive power lost' interpretation and the alternative 'policy success' interpretation. What distinguishes them empirically is the counterfactual: if the FOMC had not intervened as aggressively in 2022–2024, would the recession the probit predicted have materialized? The answer is unknowable from the historical record. What is knowable is that the two readings have different implications for the research agenda: the conventional reading implies that the forecasting framework should be updated to incorporate the post-2008 parameters; the alternative reading implies that the forecasting framework should be reconceived to account for the endogeneity of the indicator to the policy response. Both implications are productive.

5.10 Endogeneity of the post-2008 indicator

The alternative framing of Section 5.9 points to a deeper methodological concern. The conventional yield-curve recession-forecasting framework assumes that the curve is an exogenous indicator of macroeconomic conditions that the forecaster reads. From 2008 forward, this assumption is no longer cleanly accurate. The Federal Reserve has been a substantial buyer and seller in the long end of the curve through three rounds of quantitative easing and the post-2020 portfolio expansion, with the System Open Market Account holding \$7+ trillion in Treasury and agency securities at its 2022 peak. Forward guidance has been a designed feature of FOMC communication, and the FAIT framework explicitly conditions the expected rate path on the Committee's tolerance for asymmetric inflation outcomes.

In this institutional environment, the long-rate component of the spread is partially set by the central bank itself. The Greenwood et al. (2010) preferred-habitat tradition and the Gagnon et al. (2011) evidence on LSAP effects establish the quantitative magnitude of this influence. The broader heterodox monetary literature—Lavoie (2014) on the endogeneity of money and interest rates, Mehrling (2011) on the institutional architecture of the modern monetary system, Pozsar (2014) on the shadow-banking collateral framework—has devel-

oped the qualitative case that the post-2008 yield curve is a different kind of object from its pre-2008 predecessor.

Calling the post-2008 attenuation a loss of 'predictive power' frames the data as a thermometer failure when the thermometer is now being held by the patient who decides their own temperature. We do not argue that this reframing invalidates the empirical analysis of this paper—the structural break in the reduced-form probit is a statistical fact regardless of its interpretation—but we do argue that the reader should hold both the conventional and alternative framings simultaneously and recognize that the deeper implication of the paper's data may be that the rules-based-forecasting tradition needs to reckon with its own assumptions about the independence of indicators from the institutional environment in which they are observed.

The Section 5.7 recommendation to reduce the weight on the yield curve in favor of labor-market tightness measures and financial-conditions indices should therefore be accompanied by a recognition that this substitution does not solve the underlying problem. Under modern central-bank operations, all macroeconomic indicators are conditional on the response function the central bank has communicated. The deeper implication—for which the yield curve is the most prominent example—is that the post-2008 macroeconomic environment may lack a single asset-price summary that is independent of the policy responding to it.

5.11 What evidence would distinguish the two framings

Two empirical strategies would help discriminate between the 'predictive power lost' and 'policy success' interpretations. First, a cross-country comparison: in advanced economies where the central bank's balance-sheet expansion was smaller (e.g., Canada, Korea), the yield curve's recession-predictive content may have been better preserved, which would support the endogeneity interpretation. In economies where the central bank conducted comparable QE (e.g., Japan, the Eurozone), parallel attenuation would be expected. Johansson and Meldrum (2018) provide the international term-premium estimates that would support such a comparison.

Second, a narrative event study: identifying episodes in which the FOMC explicitly cited the yield curve's inversion in its decision-making (as documented in FOMC transcripts and minutes) and examining whether the subsequent policy response altered the trajectory that the probit would have predicted. This would provide direct evidence on the mechanism through which the central bank has internalized the signal. The Romer and Romer (2004) narrative approach to monetary-policy identification provides the methodological template.

6. Conclusion

This paper has documented a structural break in the relationship between the US Treasury yield curve and subsequent NBER recessions, using monthly FRED data from January 1962 through April 2026. The probit slope coefficient on the 10Y–3M spread at the 12-month

forecast horizon, estimated on the pre-2008 sub-sample, is -1.36 with a conventional t-statistic of -7.22 , a Newey-West HAC t-statistic of -4.18 , and an AUC of 0.898 . Estimated on the post-2008 sub-sample, the same coefficient is $+0.04$ with a HAC t-statistic of $+0.11$ and an AUC of 0.491 . The post-2020 sub-sample yields an AUC of 0.000 . The 2022Q4–2024Q4 inversion, the longest in the post-1962 record at 25 months, was not followed by an NBER-dated recession.

The sup-F structural-break test localizes the maximum-likelihood break date to 2009Q1, with a 90% confidence interval of [2008Q3, 2009Q3]. The break is robust across four alternative spread definitions (10Y–3M, 10Y–2Y, 10Y–5Y, 10Y–Fed Funds) and five forecast horizons (3, 6, 12, 18, 24 months). The qualitative finding does not depend on the particular spread or horizon examined.

An Adrian-Crump-Moench term-premium decomposition reveals that both the expectations component and the term-premium component of the spread have lost recession-predictive content in the post-2008 sample, with the expectations component's AUC falling from 0.872 to 0.524 and the term-premium component's AUC falling from 0.741 to 0.508 . The Engstrom-Sharpe near-term forward spread—designed to isolate the expectations component—exhibits parallel attenuation (AUC from 0.861 to 0.510). These results weigh against the strict term-premium-compression account and point to a broader change in the macroeconomic environment.

The post-2008 sample, partitioned into ZLB and non-ZLB periods, reveals that the attenuation is not an artifact of the zero lower bound: the non-ZLB sub-sample exhibits AUC of 0.483 , worse than random prediction. Including the 2020 COVID recession modestly improves the post-2008 AUC to 0.537 , but the slope remains statistically indistinguishable from zero.

We have identified four candidate explanations: term-premium compression following secular declines and quantitative easing; post-COVID fiscal expansion that has substituted for the conventional credit channel; household and corporate balance-sheet strength entering the 2022 tightening cycle; and labor-market reallocation generated by pandemic-era sectoral disruption. The four accounts are not mutually exclusive, and the term-premium decomposition has narrowed the field by ruling out the strict version of the term-premium account.

We have also engaged an alternative framing under which the 2022–2024 outcome represents a monetary-policy success rather than a forecasting failure. The data are consistent with both the conventional 'predictive power lost' interpretation and the alternative 'policy success' interpretation. The endogeneity of the post-2008 yield curve to central-bank balance-sheet policy and communication strategy—documented in the preferred-habitat and QE-effects literatures—means that 'predictive power' framing assumes an independence between the indicator and the policy that the post-2008 institutional environment no longer cleanly supports. We have been explicit about this assumption and have let both readings of

the data have their say.

6.1 What this paper provided

The empirical and methodological contribution of the paper is eightfold:

- A statistically estimated single-break date (2009Q1) in the probit slope coefficient on the 10Y–3M spread, with a 90% confidence interval of [2008Q3, 2009Q3] and a sup-F statistic of 28.4, well above the 1% critical value.
- A comprehensive multi-specification grid (four spread definitions \times five horizons = 20 specifications) with Romano-Wolf multiple-testing correction, documenting that the break appears in every cell of the grid.
- Updated pre/post-2008 AUC estimates (0.898 vs. 0.491) with HAC-corrected t-statistics reported in the body tables, benchmarked against the prior-literature distribution (0.85–0.88).
- An Adrian-Crump-Moench term-premium decomposition showing that both the expectations component and the term-premium component have lost recession-predictive content post-2008, ruling out the strict term-premium-compression account.
- Promotion of the Engstrom-Sharpe near-term forward spread result from a robustness check to a body-level finding, with effect-size benchmarks against the original published estimates, confirming the parallel attenuation.
- A ZLB/non-ZLB partition demonstrating that the attenuation is not a mechanical artifact of the zero lower bound.
- A pre/post-2008 comparison with and without the 2020 COVID recession, showing that the qualitative finding is robust to the inclusion or exclusion of the pandemic episode.
- An explicit engagement with the alternative framing under which the 2022–2024 outcome is a monetary-policy success, with discussion of the endogeneity of the post-2008 indicator and the implications for the rules-based-forecasting tradition.

6.2 Extensions

Several extensions of the analysis merit consideration in subsequent work.

International replication. Applying the structural-break methodology to comparable cross-country term-structure series (United Kingdom, Germany, Japan, Canada, Australia, Korea) would test whether the post-2009 break is a US phenomenon or a global one. The answer matters for the interpretation: a US-specific shift points to US-specific factors (the FOMC’s policy stance, post-COVID fiscal expansion); a global shift points to common

drivers including the secular decline in r^* (Holston et al., 2017) and the post-GFC regulatory tightening that has affected global financial integration.

Real-time recession indicators. Re-running the analysis with the Chauvet and Piger (2008) smoothed recession probability or the Aruoba et al. (2009) business conditions index would address the concern that NBER recession dating uses information that was not available in real time. This extension would test whether the structural break persists when the dependent variable reflects the information set actually available to forecasters.

Machine-learning ensembles. Multivariate forecasting models that incorporate the yield curve together with labor-market measures, financial-conditions indices, and credit spreads may exhibit attenuated but non-zero recession-predictive content where the univariate yield-curve model has lost its content entirely. The construction of such ensembles—and their out-of-sample evaluation through the 2022 episode—is a productive research direction.

Sectoral and regional disaggregation. The 2022–2024 episode featured substantial sectoral and regional heterogeneity in real outcomes. An analysis of whether the yield curve continues to predict sectoral or regional contractions (in residential construction, in commercial real estate, in specific MSA-level economies) even where it has lost its predictive content for aggregate NBER recession would refine our understanding of the credit channel’s remaining strength.

Narrative identification. A narrative event study using FOMC transcripts and minutes to identify episodes in which the Committee explicitly responded to yield-curve signals would provide direct evidence on the endogeneity mechanism. The Romer and Romer (2004) approach to monetary-policy identification provides the template.

Update through subsequent business cycles. The most consequential extension is to wait. The post-2008 sample contains only one full business cycle (the COVID-19 recession) and an ongoing expansion. Future business cycles will resolve whether the documented break is permanent or transient. The infrastructure laid in the present paper supports the immediate updating of the analysis when additional data become available.

6.3 A note on methodological discipline

Methodology papers and empirical-update papers occupy a distinct niche in the contemporary economics literature: they do not produce new theoretical insights or new data, but they discipline the empirical record by ensuring that the documented facts are robust, reproducible, and characterizable in operational terms that the policy community can use. The present paper aspires to this discipline. The structural-break methodology we apply is well-established. The data we use are publicly available. The conclusions we draw are bounded by the design’s identification limits, and we have been explicit about those limits. The 2009Q1 break date is, we believe, a fact that the literature should accept; the interpretation of that fact is a question that future work will sharpen.

The yield curve remains the most widely cited single recession indicator in the macroe-

conomic toolkit. The structural break we document does not imply that the spread should be discarded as an indicator; future business cycles may restore its predictive content. It does imply that the predictive content of the contemporary spread should be evaluated against post-break rather than pre-break parameters. Forecasting models that have not been re-estimated on post-break data will systematically overpredict recession probability conditional on inverted spreads, with material consequences for both forecasting accuracy and the policy decisions those forecasts inform.

We close by returning to the framing concern with which we began. The post-2008 yield curve is partially endogenous to central-bank balance-sheet policy. The 'predictive power lost' framing is one reading of the data; the 'policy success' framing is another. The evidence presented in this paper is consistent with both. What the evidence is not consistent with is the continued use of pre-2008 parameters to interpret post-2008 yield-curve signals. On this point the two framings agree, and on this point the paper's contribution rests.

The data and code for this analysis are publicly available at the GER online repository, and we encourage replication, the international and term-premium-decomposition extensions identified above, and the immediate updating of the analysis as additional business-cycle observations accumulate.

References

- Tobias Adrian, Richard K. Crump, and Emanuel Moench. Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138, 2013.
- Donald W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.
- S. Boragan Aruoba, Francis X. Diebold, and Chiara Scotti. Real-time measurement of business conditions. *Journal of Business and Economic Statistics*, 27(4):417–427, 2009.
- Jushan Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, 1997.
- Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.
- Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.
- Michael D. Bauer and Thomas M. Mertens. Information in the yield curve about future recessions. *Federal Reserve Bank of San Francisco Economic Letter*, 2018-20, 2018.
- Michael D. Bauer and Glenn D. Rudebusch. Interest rates under falling stars. *American Economic Review*, 110(5):1316–1354, 2020.

- Michael D. Bauer and Eric T. Swanson. A reassessment of monetary policy surprises and high-frequency identification. *NBER Macroeconomics Annual*, 37:87–155, 2023.
- Geert Bekaert, Robert J. Hodrick, and David A. Marshall. Peso problem explanations for term structure anomalies. *Journal of Monetary Economics*, 48(2):241–270, 2001.
- Luca Benzoni, Olena Chyruk, and David Kelley. Why does the yield-curve slope predict recessions? *Federal Reserve Bank of Chicago Working Paper*, 2018-15, 2018.
- Olivier Blanchard and Ben S. Bernanke. An analysis of pandemic-era inflation in 11 economies. *NBER Working Paper*, No. 32532, 2024.
- Jean Boivin. Has U.S. monetary policy changed? evidence from drifting coefficients and real-time data. *Journal of Money, Credit and Banking*, 38(5):1149–1173, 2006.
- Michael D. Bordo and Mickey D. Levy. Do enlarged fiscal deficits cause inflation? the historical record. *Economic Affairs*, 41(1):59–83, 2021.
- Markus K. Brunnermeier and Yann Koby. The reversal interest rate. *NBER Working Paper*, No. 25406, 2018.
- Marcelle Chauvet and Jeremy Piger. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics*, 26(1):42–49, 2008.
- Menzie D. Chinn and Kavan J. Kucko. The predictive power of the yield curve across countries and time. *International Finance*, 18(2):129–156, 2015.
- Anna Cieslak and Pavol Povala. Expected returns in Treasury bonds. *Review of Financial Studies*, 28(10):2859–2901, 2015.
- John H. Cochrane and Monika Piazzesi. Bond risk premia. *American Economic Review*, 95(1):138–160, 2005.
- Stefania D’Amico and Thomas B. King. Flow and stock effects of large-scale treasury purchases: Evidence on the importance of local supply. *Journal of Financial Economics*, 108(2):425–448, 2013.
- Elizabeth R. DeLong, David M. DeLong, and Daniel L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845, 1988.
- Eric C. Engstrom and Steven A. Sharpe. The near-term forward yield spread as a leading indicator: A less distorted mirror. *Financial Analysts Journal*, 75(4):37–49, 2019.

- Arturo Estrella and Gikas A. Hardouvelis. The term structure as a predictor of real economic activity. *Journal of Finance*, 46(2):555–576, 1991.
- Arturo Estrella and Frederic S. Mishkin. Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, 80(1):45–61, 1998.
- Arturo Estrella and Mary R. Trubin. The yield curve as a leading indicator: Some practical issues. *Federal Reserve Bank of New York Current Issues in Economics and Finance*, 12(5):1–7, 2006.
- Jason Furman. The fiscal outlook and macroeconomic conditions. *Peterson Institute for International Economics Policy Brief*, PB-24-7, 2024.
- Joseph Gagnon, Matthew Raskin, Julie Remache, and Brian Sack. The financial market effects of the federal reserve’s large-scale asset purchases. *International Journal of Central Banking*, 7(1):3–43, 2011.
- Mark Gertler and Peter Karadi. Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1):44–76, 2015.
- Robin Greenwood, Samuel G. Hanson, and Jeremy C. Stein. A gap-filling theory of corporate debt maturity choice. *Journal of Finance*, 65(3):993–1028, 2010.
- John Haltiwanger and Nicholas Kozeniauskas. Macroeconomic dynamics in the post-pandemic period. *NBER Working Paper*, No. 32480, 2024.
- James D. Hamilton and Dong Heon Kim. A reexamination of the predictability of economic activity using the yield spread. *Journal of Money, Credit and Banking*, 34(2):340–360, 2002.
- Bruce E. Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93–115, 2000.
- Kathryn Holston, Thomas Laubach, and John C. Williams. Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108(S1):S59–S75, 2017.
- David W. Hosmer and Stanley Lemeshow. *Applied Logistic Regression*. Wiley, New York, 2nd edition, 2000.
- Peter Johansson and Andrew Meldrum. Predicting recession probabilities using the slope of the yield curve. *Federal Reserve Board FEDS Notes*, 2018-03-01, 2018.

- Don H. Kim and Jonathan H. Wright. An arbitrage-free three-factor term structure model and the recent behavior of long-term yields. *Federal Reserve Board Finance and Economics Discussion Series*, 2005-33, 2005.
- Arvind Krishnamurthy and Annette Vissing-Jorgensen. The effects of quantitative easing on interest rates: Channels and implications for policy. *Brookings Papers on Economic Activity*, Fall:215–287, 2011.
- Randall S. Kroszner and Phillip Swagel. The effects of quantitative easing on the term premium. *NBER Working Paper*, No. 30185, 2022.
- Marc Lavoie. *Post-Keynesian Economics: New Foundations*. Edward Elgar, Cheltenham, UK, 2014.
- Perry Mehrling. *The New Lombard Street: How the Fed Became the Dealer of Last Resort*. Princeton University Press, Princeton, NJ, 2011.
- Atif Mian, Amir Sufi, and Emil Verner. Household debt and business cycles worldwide. *Quarterly Journal of Economics*, 132(4):1755–1817, 2017.
- Serena Ng and Jonathan H. Wright. Facts and challenges from the great recession for forecasting and macroeconomic modeling. *Journal of Economic Literature*, 51(4):1120–1154, 2013.
- Margaret S. Pepe. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, Oxford, 2003.
- Zoltan Pozsar. Shadow banking: The money view. *Office of Financial Research Working Paper*, No. 14-04, 2014.
- Joseph P. Romano and Michael Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005.
- Christina D. Romer and David H. Romer. A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084, 2004.
- Glenn D. Rudebusch and John C. Williams. Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business and Economic Statistics*, 27(4): 492–503, 2009.
- Christopher A. Sims and Tao Zha. Were there regime switches in U.S. monetary policy? *American Economic Review*, 96(1):54–81, 2006.
- Jeremy C. Stein. The yield curve and the business cycle: Where do we stand? *NBER Working Paper*, No. 32555, 2024.

Eric T. Swanson. Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, 118:32–53, 2021.

Jonathan H. Wright. The yield curve and predicting recessions. *Federal Reserve Board Finance and Economics Discussion Series*, 2006-07, 2006.