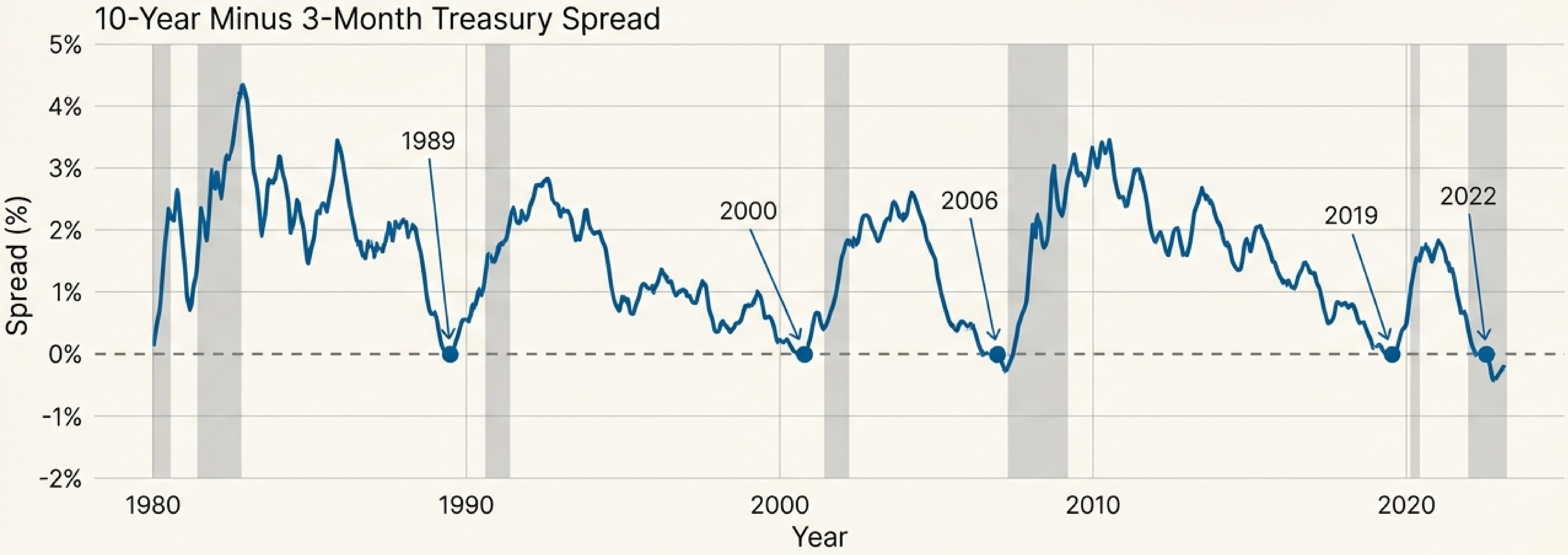
The Yield Curve Is Wall Street's Most Feared Recession Indicator



Historically, an inverted yield curve (10Y < 3M) has preceded every major U.S. recession.

Can We Build an Early-Warning System for Inversions?

The Prediction Task

Predict the probability that the 10Y-3M yield spread will be negative 12 weeks from today.

The Raw Material

Weekly constant maturity Treasury yields from FRED, spanning 1970-2025. We focus on the core 3M, 2Y, 5Y, and 10Y maturities.

The First Hurdle: Yields Behave Like a Random Walk

10Y Yield Level (Non-stationary)

ADF p-value: 0.015, KPSS p-value: 0.01

10Y-3M Spread (More Stationary)

ADF p-value: 0.034, KPSS p-value: 0.01

Forecasting the *level* of a yield is notoriously difficult. But forecasting the *relationship* between yields—the spread—is more tractable.

Our baseline 'Naïve' model, which assumes today's inversion state persists, achieves a test-set **AUC of 0.837**. This is the score to beat.

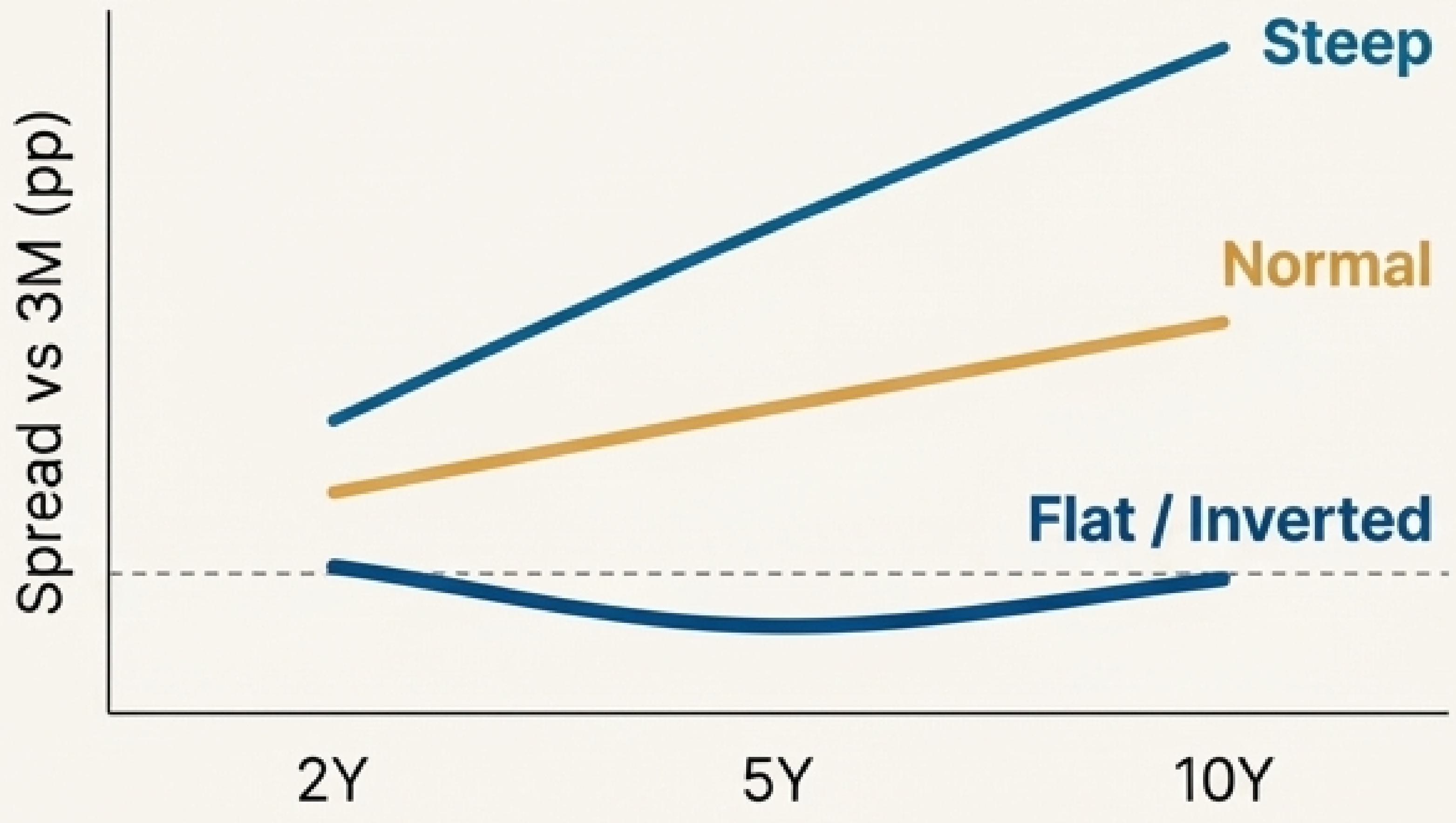
We Decomposed the Curve into Its Core Drivers

Principal Component Analysis (PCA) Loadings						
	y_3m	y_2y	y_5y	y_10y	Interpretation	Explained Variance
PC1	0.494	0.505	0.504	0.497	Level	97.5%
PC2	0.723	0.175	-0.290	-0.602	Slope	2.3%
PC3	-0.471	0.688	0.255	-0.489	Curvature	0.2%

Nearly all weekly yield variation is explained by just three intuitive factors: the overall level of rates, the slope of the curve, and its curvature. These factors become powerful inputs for our model.

The Yield Curve Operates in Three Distinct ‘Regimes’

Typical Curve Shapes by Cluster → Inversion Risk by Regime



Regime	% of Weeks	Inversion Frequency
Steep	29%	0.00%
Normal	47%	0.09%
Flat/Inverted	**24%**	**47.87%**

Using K-Means clustering on shape features, we identified three market regimes. Nearly all historical inversions occur when the curve is in the “Flat/Inverted” state.

Our Edge Came from Engineering Features that Capture Curve Dynamics

Shape & Regime

Features: slope_10y_3m, regime_score

Quantifies the current slope and which of the three regimes the curve is in.

Persistence & Memory

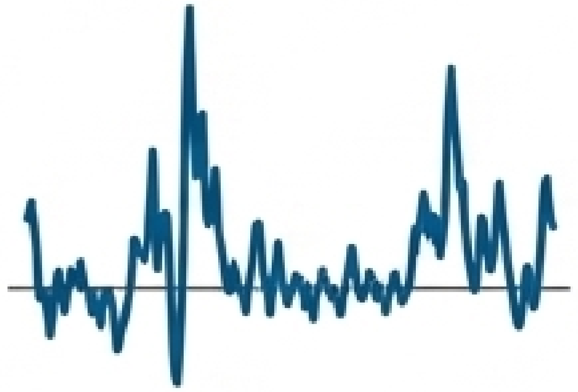
Features: weeks_since_zero_cross, weeks_below_zero_52

Measures how long the curve has been inverted and the time since it last flipped.

Momentum & Volatility

Features: d_PC2_slope, rv_slope_10y_3m_13

Captures the rate of change in the curve's slope and its recent realized volatility.



Economic Disequilibrium

Features: ect1 (Error-Correction Term)

A signal from cointegration analysis that measures deviation from the long-run equilibrium relationship between yields.

Over 30 features were engineered, all lagged to prevent look-ahead bias.

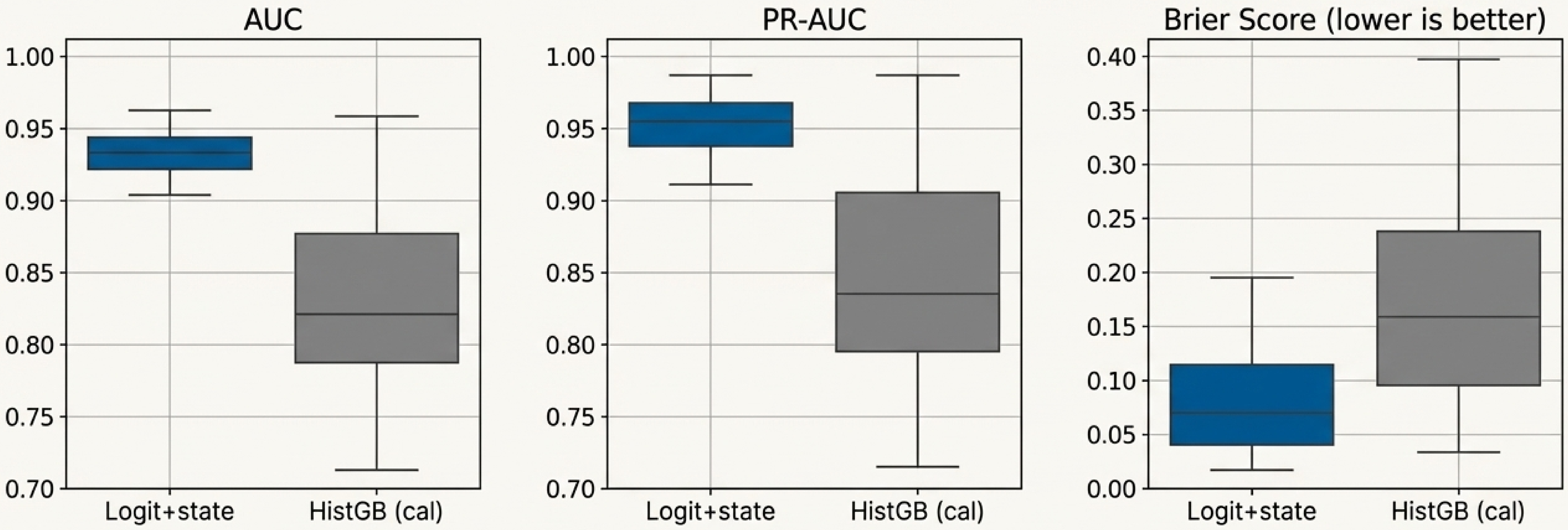
We Pitted a Simple, Interpretable Model Against a Non-Linear Powerhouse

Out-of-Time Test Set Performance (Post-2015)

Model	Key Metrics	AUC	PR-AUC	Brier Score
Baseline: Naïve Persistence	(Benchmark)	0.837	0.654	0.130
✓ **Contender 1: Logistic Regression**	(Interpretable)	**0.910**	**0.756**	**0.117**
Contender 2: HistGradientBoosting (cal.)	(Non-Linear)	0.880	0.688	0.176

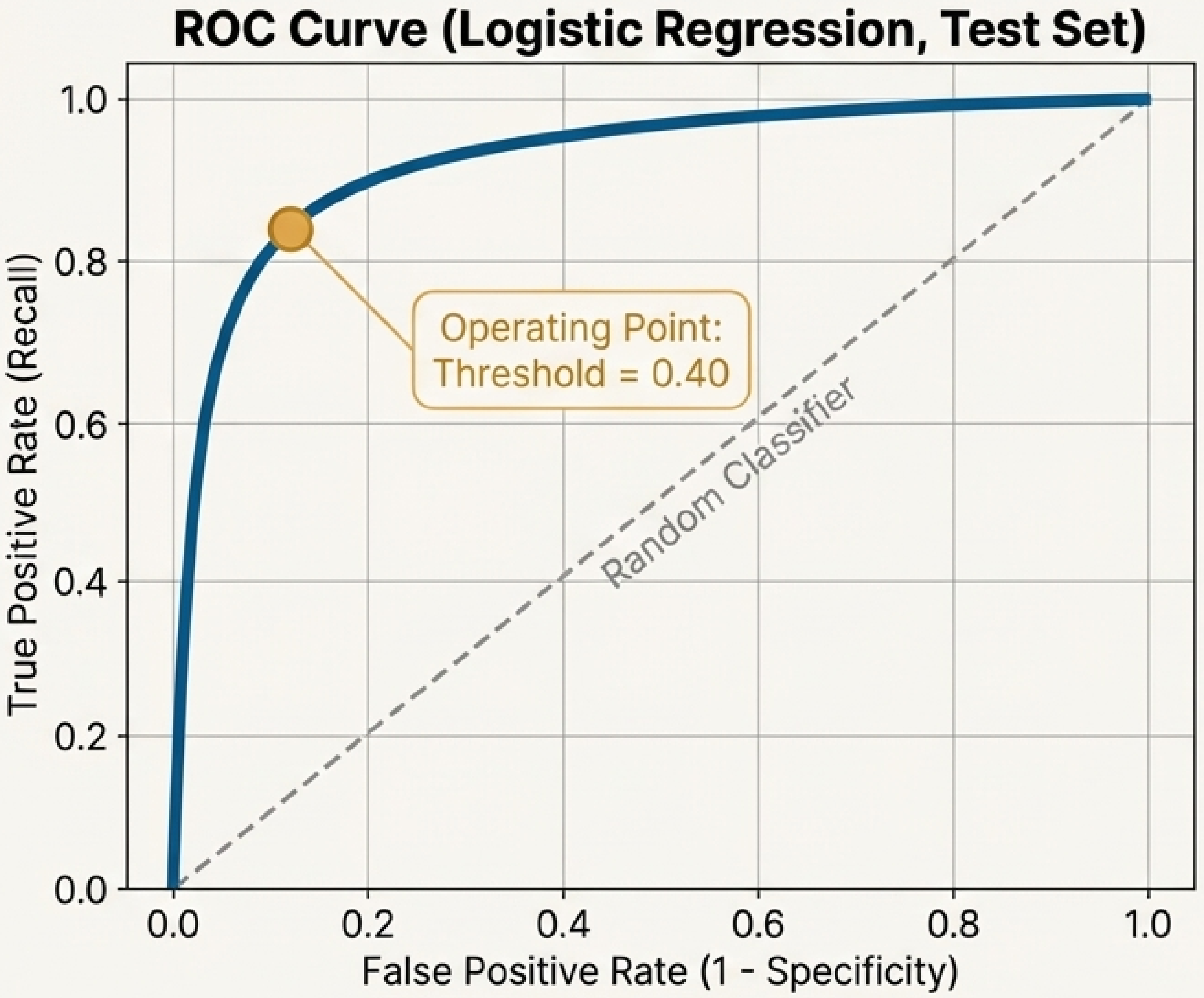
The simpler, more transparent **Logistic Regression** model, powered by our engineered features, significantly outperformed both the naïve baseline and the more complex non-linear model.

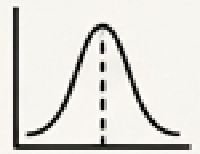
A Single Test Set Isn't Enough; The Model Must Be Robust Over Time



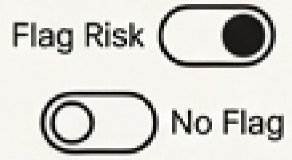
Across 8 overlapping time windows, the Logistic Regression model consistently delivers superior and more stable performance, proving its robustness.

A Model's Probability Becomes Actionable at a Decision Threshold

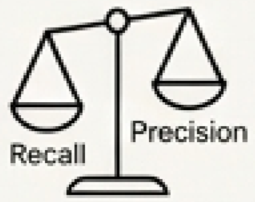




1. The model gives us a probability (0% to 100%).



2. We need a threshold to turn that probability into a binary decision: "Flag Risk' or 'No Flag".



3. Choosing a threshold involves a trade-off between catching true events (Recall) and avoiding false alarms (Precision).

4. We selected a threshold of **0.40** to balance these objectives for an early-warning system.

At This Threshold, the System Catches 98% of True Inversions

Performance at 0.40 Threshold (Out-of-Time Test Set)

		Predicted	
		No Inversion	Inversion
Actual	No	**TN = 315**	**FP = 85** False Alarm
	Yes	**FN = 3** Miss	**TP = 152**

Recall: 98%

We miss almost nothing.

Precision: 64%

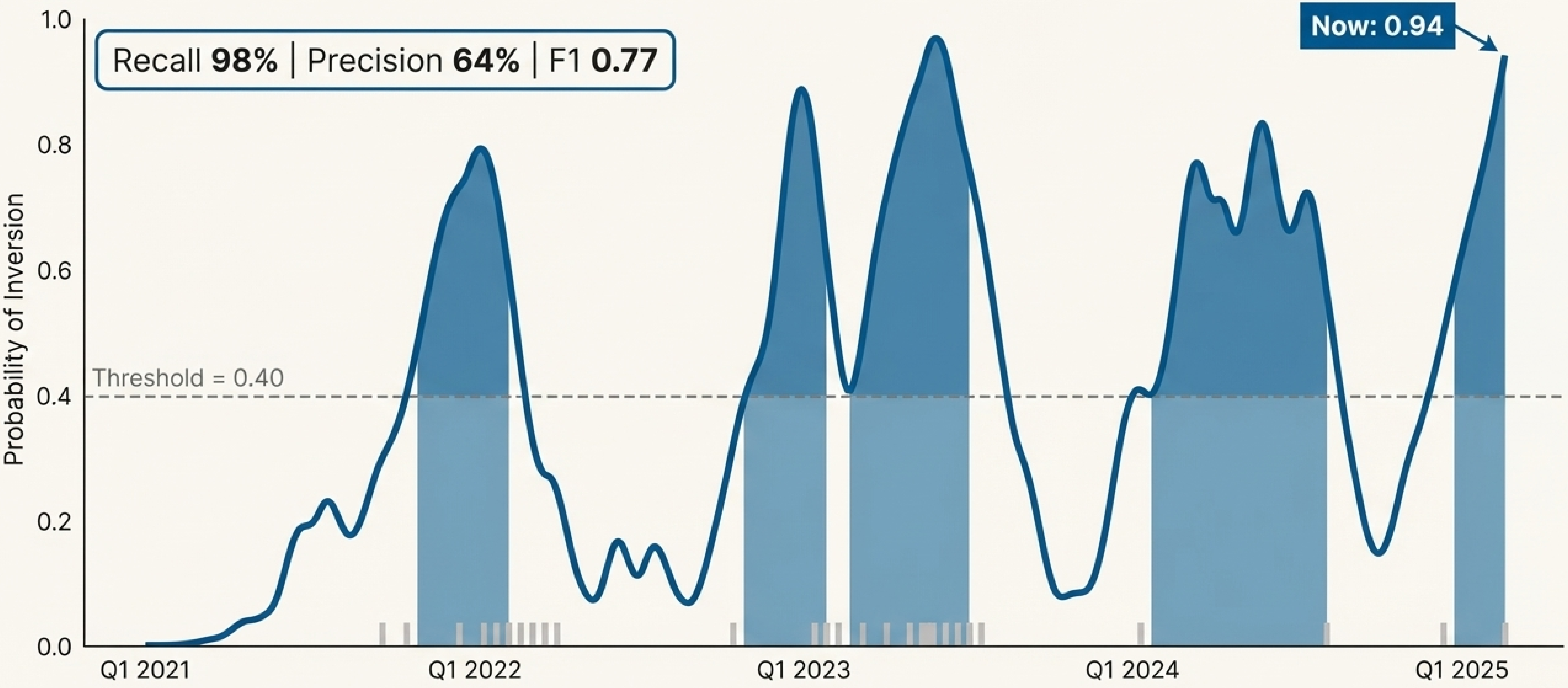
When we flag risk, we're right 2/3 of the time.

AUC: 0.91

Overall model skill.

This is the ideal profile for an early-warning system: we accept some false alarms to achieve an **exceptionally high capture rate of true risk events.**

Our Model in Action: An Early-Warning System for the Last 5 Years



NotebookLM

We Delivered a Robust, Interpretable Inversion Risk Classifier

Key Achievements

- ✓ Built a reproducible pipeline from raw data to a production-ready model.
- ✓ Proved that intelligent feature engineering (capturing curve shape, regimes, dynamics) is more effective than algorithmic complexity.
- ✓ Delivered a classifier with **0.91 AUC** and **98% Recall** on the test set, validated with robust rolling-origin backtests.
- ✓ Created a suite of slide-ready diagnostics and a clear 'hero chart' for ongoing monitoring.

What's Next

- → **Stronger Forecasts**: Add macro covariates (CPI, Fed Funds Futures) to improve 10Y forecasts.
- → **Enhance Classifier**: Incorporate market volatility (e.g., MOVE index) and explore regularized models for feature selection.
- → **Explainability**: Implement SHAP to provide feature-level explanations for model predictions.

Predicting Yield Levels is a Random Walk. Predicting Yield Risk is a Solvable Problem.