National Impact Velocity (NIV)

Catching Preliminary Systematic Stress Before Recessions

Abstract

National Impact Velocity (NIV) is a parsimonious macroeconomic indicator designed to detect systematic stress accumulation before recessions by measuring the velocity of regenerative capital formation against cumulative economic friction. The framework synthesizes investment dynamics, monetary liquidity, capacity utilization, and credit circulation inhibitors into a single composite signal. Validated through six out-of-sample tests spanning 504 months (1970–2024), NIV achieves an AUC of 0.8538 at the 18-month horizon, filters 98.5% of false alarms via an L2-regularized ensemble, and contributes 41.71% orthogonal variance beyond the Fed Yield Curve. Results demonstrate that NIV functions as a structurally independent, leading indicator of macroeconomic regime shifts.

Project Directives Addressed

1. What did you do, why, and to what end?
2. How exactly do you evaluate it? (e.g., hyperparameter tuning, measures of success, measures others have used)
3. Carefully report a rich set of results coming from these evaluations.
4. Make a fair comparison to alternative approaches.

1 Overview

NIV (National Impact Velocity) is an attempt to address the significance of investment dynamics in catching preliminary systematic stress before recessions. This idea came from a real criticism of GDP as a single measure, which makes everyone focus on GDP instead of the more important underlying issues that cause stress in the economy. My thesis is that these stressors are the driving force.

This idea inspired my approach for a parsimonious equation designed to catch the relation between positive investment dynamics and massive cumulative drags and friction within the economy. Using the idea of regenerative capital—capital that has compounding exponential margins of value generation over time—as the liberator, you can see how efficient the economy is at mobilizing this at scale.

1.1 The NIV Equation

National Impact Velocity measures economic momentum versus friction:

$$\boxed{\mathrm{NIV}_t \;=\; \frac{u_t \cdot P_t^2}{(X_t + F_t)^{\eta}}} \tag{1}$$

where u_t is thrust (kinetic impulse), P_t is capital productivity, X_t is economic slack, F_t is friction drag, and η is an elasticity parameter.

1.2 Thrust (u) — Kinetic Impulse

$$u = \tanh\big(+1.0 \cdot \Delta G \;+\; 1.0 \cdot \Delta A \;-\; 0.7 \cdot \Delta r\big) \tag{2}$$

- $+\Delta G$: Investment YoY growth (positive = expansion).
- $+\Delta A$: M2 money YoY growth (positive = liquidity).
- $-\Delta r$: Fed Funds rate change (hikes subtract thrust).

Range: $[-1, +1]$ *via* tanh. *Positive = expansion impulse; negative = contraction impulse.*

Thrust is what I measured as the "invisible hand" within the economy, where the Federal Reserve and economic organizations have disproportionate power in fixing the state of levers and margins they possess. These have the highest impact in driving the variables within this equation.

The tanh function prevents the exponential explosion of values and creates a bounded spectrum for this impact, ensuring applicability across various eras represented in the entire FRED dataset. It also accounts for periods of significant exogenous shocks, such as the massive liquidity implosions and unchaining of the M2 money supply experienced during the 2008 and 2020 QE events.

Investment YoY Growth. I chose this variable because it serves as a proxy for economic confidence; the highest level of confidence is demonstrated by reinvesting and taking on higher risks. I acknowledge there is a sticky layer of investment that persists during events of systematic stress and concede the compensatory and narrow layers of such investment, such as the significance of foreign inflows and mandatory capital consumption. When you analyze the true feedback loop influenced by government decisions, you will find that much of it directly affects investment decisions, as the population experiences significant changes in their local economies due to various investment dynamics. This entire bottom-up force of year-over-year investment growth serves as a barometer for increasing economic activity; by identifying when certain thresholds are passed, one can determine a dynamic alpha at specific economic states that significantly influence the Federal Reserve's ability to implement additional measures, thereby greatly impacting investment dynamics.

M2 Money Growth. This variable is a proxy for liquidity dynamics, which are critical for the entire financial plumbing for the distribution of credit to private businesses. It represents, most importantly, the collective shared prosperity, as it can directly bolster the fractional reserves of merchant banks across the board, allowing broadstream liquidity to drive all economic factions. This indicator provides the base necessary for influencing investment dynamics at scale, allowing for risk premiums to fall and for credits to be distributed more freely.

Fed Funds Rate. This is a critical rate, as it sets the relational everyday basis for the distribution of credit to businesses at scale to drive investment. This system is designed to filter and establish the market rate of liquidity throughout the economy, where the largest amounts of liquidity are available at the lowest rates, creating a descending gradient based on the size of the liquidity transfer and the proportional burden shouldered by the debtor. The summation of all this lending at scale has an outsized influence on the microdecisions that guide the obligations and incentives that compose the market dynamics of the wider economy.

1.3 Efficiency (P) — Capital Productivity

$$P = \frac{\text{Investment} \times 1.15}{\text{GDP}} \tag{3}$$

- **Investment**: GPDIC1 (Real Private Domestic Investment).
- **1.15**: R&D / Education proxy multiplier.

Squared in Equation (1) (P^2) to reward productive capital allocation and punish hollow growth.

Efficiency represents regenerative capital, which is the fundamental principle behind NIV. Regenerative capital is a theory that money flowing into areas of the economy that have exponentially forming margins is able to cause disproportionate increases in growth. This metric measures the tangible value base of the economy; its value provides the basis for all forms of capital, and it allows the recycling of cash into investment with exponentially compounding margins through non-residential investment, residential investment, and change in inventories. This is observed through examples such as data centers, where their modularity not only powers the current state of AI but also facilitates future improvements, which are more likely to cause disproportionate ripple effects across the economy than can currently be projected.

1.15 Multiplier. I chose 15% as the standard percentage for this, as I felt it to be a balance that is able to address an average of regenerative capital. This weight can be modified, especially in times of high growth velocity, where tangible real-world value is harmonized with economic macro shifts, such as a national buildout to serve AI development at a national scale. It can also be modified downwards in times wherein the investment is directed towards capital that does not have compounding margins, such as times of high speculative capital during the financialization that dominates investment dynamics for the last few decades.

1.4 Slack (X) — Economic Headroom

$$X = 1 - \frac{\text{TCU}}{100} \tag{4}$$

- **TCU**: Total Capacity Utilization (typically 70–85%).

Range: $[0, 1]$. *High slack = room to grow. Low slack = economy near capacity limits.*

Slack. I chose idle capacity as one of the two primary detractor measures of regenerative capital, as idle capacity tends to have compounding negative margins on regenerative capital formation. Idle capacity chips away at the composite capital base of the economy and serves as a sticky detractor from investment within certain industries; idle capacity cascades and causes massive disproportionate influence on investment decisions in those industries. This also contributes to a decrease in value within investment dynamics that affect the entire capital base, as it can reveal any significant reversals in the current capital investment regime.

1.5 Drag (F) — Friction Forces

$$F = 0.4 \cdot s + 0.4 \cdot \max(0,\ r - \pi) + 0.2 \cdot \sigma \tag{5}$$

- s: Yield inversion penalty ($|T10Y3M|/100$ if inverted).
- $r - \pi$: Real interest rate (FedFunds/100− Inflation).
- σ: Rate volatility (12-month rolling StdDev/100).

All components add friction that slows capital circulation. Higher F = more economic drag.

Drag refers to the accumulation of credit circulation inhibitors and their potential to influence firm decision-making. Together, these factors pose a significant threat to regenerative capital when they dominate, as they prolong depressive cycles of capital formation and create inertia-driven investment dynamics that tend to prioritize cash conservation, thereby straining the capital flow towards higher-risk areas and compounding margins.

Yield Inversion Penalty. This measurement represents the macro headwinds that influence the health of the overall capital within the economy and the reflexive relationship between capital formators and the broader credit instruments that guide the market. Holding the shorter-term debt instruments will mean there is a shift towards more aggressive recoupling of credit risk, leading to strains on all business lending. This variable is weighted highly at 0.4, as it is a primary determinant for the short-term credit dynamics component of investment.

Real Interest Rate. Inflationary conditions are a key source of disharmony in the economy because they misalign investment dynamics with the market and create uncertainty for local capital expansion. The market interest rate has a massive impact with inflation on dictating the input dynamics of the business in forming the output to transfer to the wider economy; this leads to a disproportionate reduction in risk assumption, and compounding margins is not something that will be pursued. The max function ensures that negative real interest rates do not add any additional negative drag; it just stops being a detractor from the entire drag function. This factor is weighted highly at 0.4 because it significantly influences the final decision, which must be rational and capable of generating cash at its base value—an outcome that is uncertain in the current climate of disharmony caused by inflationary pressures and fluctuating market interest rates.

Rate Volatility. This serves as a proxy for indecision and ineffective monetary policy, reflecting how rapidly changing interest rates can affect businesses' ability to make long-term decisions about investing in regenerative capital. It is weighted lower at 0.2, as it does not have the same overbearing impacts that the yield inversion magnitude and the real interest rate have on the stakeholders necessary for regenerative capital formation.

2 Out-of-Sample Testing Protocol

The core of the OOS testing was the walk-forward protocol; this is based on not showing all the data at once and only allows reliving the entire history at $t + h$, preventing look-ahead bias. This leads to predictions based on past data. The walk-forward started in 1970 and moved forward, allowing the model to see 5 months forward and be retrained at that increment.

This model ran on an *expanding window*, meaning that as it moved forward in time, it had a memory of all its previous history. There was a warm-up period until 1983 to allow the model to accumulate a considerable store of history. I also tested it on a *rolling window* that only keeps data for the last 15 years. However, in the successful OOS testing, the expanding window was the clear victor.

3 Six Validation Tests

In order to ensure NIV had maximal predictive power, I used six different stress tests.

3.1 Test 1: Calibrated Ensemble Performance

I used three different machine learning methods to "vote" on systematic stress risk, then combined all of them. This procedure ensured several diverse opinions more accurate than any single model, as there are many blind spots in one model alone.

L2 Logistic Regression. The first model was L2 Logistic Regression, which focuses on the larger trends dictating the data. It finds a mathematical link between all the variable inputs (thrust, drag, etc.) in the larger equation. It has a linear relationship between each input variable and uses a sigmoid function to bound the output between 0 and 1, representing a probability between 0% and 100%. This means that if thrust rises 1%, the systematic stress probability moves up a set amount within the bounds.

Hyperparameter: Ridge Regression Logic. The L2 regularization adds a penalty assigned to the size of the model coefficients. This means that if a variable input is very large, the score is penalized. This forces the model to distribute its entire attention among all the variables, forming the basis for the model being resistant to noise and other glitches within certain data. This avoids overfitting and hyperfixation with one part of the data.

Boosted Stumps (AdaBoost). The decision tree "stump" is a model with only one split, so it can only look at one variable and sets an internal cutoff point. Since these stumps are not complex enough to memorize data, this forces them to find a signal. The boosting process turns these weak stumps into a fully comprehensive measurement with an order: the first stump detects a systematic risk event, the second stump pays more attention to the months it got wrong by modifying the cutoff values. After 15 rounds, you capture many difficult, non-linear patterns that every previous stump missed. This makes it a great determinant of "if–then" truth within the data, such as "if real rates are high but investment efficiency is also high, the economy is fine."

Hyperparameters:

- `max_depth=1` — each model focuses on one problem at a time.
- `n_estimators=15` — how many stumps are in the chain, fixing the previous one's mistakes.
- `learning_rate=0.1` — how much each stump influences the final vote; the lower rate makes learning slow and conservative, allowing for more accurate depth in the final results.

Neural Network. The NIV components enter the input layer and pass through hidden layers that aim to uncover certain multidimensional relationships connected to times of high systematic risk, such as the combination of high drag and low efficiency happening once every 20 years. This used tanh to handle non-linear math of economic cycles. It had to run on a browser and a machine, which meant limited layers, neurons, and epochs—the data was underfit and there was not much value to be gained from it alone.

3.2 Test 2: Multi-Horizon Analysis

Most economic indicators are lagging, meaning they can identify states of recession only after the event. This test checks whether NIV can instead be a *leading* indicator in predicting systematic risk. This spanned 3-, 6-, 12-, and 18-month horizons.

Hyperparameters:

- **Horizon (h):** The time gap between input data and target prediction. These horizons took the data point at a certain time and predicted it at a fixed interval later.
- **Smoothing Window (12):** To prevent one month from changing the entire signal, a 12-month rolling mean was used. This normalizes jitters.
- **Lead Time Calibration (18 months):** The process of determining how far a signal appears before the recession begins. The model was trained to look for patterns 18 months before the crash. Probability thresholds: *yellow* at 12–35% and *red* at >35%.

3.3 Test 3: Expanding vs. Fixed Window

This was a test of whether using the entire historical record is better than using a shorter, more recent "snapshot." The economy was fundamentally different in the 1970s (fixed window), but one can also observe structural patterns that underpin several decades regardless (expanding window).

Hyperparameters:

- **Expanding Window:** Training data grows as the model moves forward, from 1970 to the most recent time.
- **Fixed Window:** Model memory is limited to the most recent 15 years. Data outside this timeframe does not factor into the analysis.
- **Walk-Forward Protocol:** Every 5 steps (months), the model is retrained; with the expanding window the training slice grows in every iteration. Only data until point t is used to predict the future.

3.4 Test 4: Tactical Translation Benchmark

This OOS test pits the NIV model against the Fed Yield Spread (10Y-3M) to see if NIV can improve upon GDP forecasting. The main measure of success was the RMSE (Root Mean Square Error): the lower the score, the better the forecast. The purpose was to determine whether NIV has a deeper systematic intuitive process that serves as a reinforced predictor of economic health compared to the industry standard.

Grid Search. This tested all combinations of smoothing windows and lags to find the optimal filter. Twenty different configurations were run against the Fed model.

Hyperparameters:

- **Window Size:** Number of months used to smooth out noise.
- **Lag Time:** Delay between data recording and model prediction.
- **RMSE:** Average distance between model prediction and actual outcome. Errors are squared (penalizing large mistakes more than small ones), averaged, and then square-rooted.

3.5 Test 5: Component Analysis

The goal of this test was to identify which components of the NIV model (thrust, efficiency, slack, or drag) were driving the predictions. This was a determination of variable rank by contribution to deflect away from "black box signals" and the dangers of a lack of transparency.

Coefficient Analysis. Every variable is assigned a coefficient that serves as a source of mathematical weight relative to the final score. Regime detection builds on this: you can track the month-to-month contribution of each component to the total change in the NIV score and identify the driver of each drop.

Hyperparameters:

- **Coefficient Magnitude:** Derived from L2 Logistic Regression, calculating linear relationships after variable changes. The share of a certain component (e.g., `thrust_share`) is then compared against the sum of absolute changes in all other components.
- **Thrust Regime Threshold:** A conditional hyperparameter for regime detection. If the thrust share exceeds 75% of total NIV movement, it flags a thrust regime. When crossed, the model widens confidence intervals by 20% to account for the higher volatility in aggressive monetary shifts.
- **Efficiency Proxy Multiplier:** Set at 1.15 by default in the core NIV calculation, scaling the investment-to-GDP ratio as a proxy for capital productivity.

3.6 Test 6: Forensic Analysis

The motivation behind this test is to prove that NIV can serve on its own and not be a replica of bond market signals. This was done to show that NIV has orthogonal information that the Fed Yield Spread cannot capture.

Correlation and Hybrid Testing Hyperparameters:

- **Correlation Analysis:** Monthly time series of the smoothed NIV composite compared against the monthly Fed yield spread average. This led to a 76% correlation, proving that NIV commits the other 24% of variance as orthogonal information.
- **RMSE:** Head-to-head forecasting simulation.
- **Hybrid Model:** Both models combined to maximize predictive power. Weights were tuned to find the best balance, ending on a 60/40 split with a weighted linear contribution.

4 Results for All Six Validation Tests

4.1 Test 1: Ensemble

The main purpose of this test was to fulfill the commonly accepted principle that a macro model's performance is dominantly derived from its ability to ignore false alarms. This ensemble ML voting mechanism using a regression probability threshold of >35% was able to suppress the noise generated from the ML calculations across all layers.

Table 1: False-alarm suppression via ensemble voting (>35% threshold).

Model	Months >35%	Std. Dev.
Neural Network	474	0.1001
Logistic Regression	230	0.2418
Boosted Stumps	50	0.1675
Final Ensemble	**7**	**0.0883**

Frequency of >35% Risk Vote. Forensic Proof. Due to a lack of memory, the neural network was heavily underfit, becoming a regurgitation of the baseline. The logistic regression tracked linear trends that had elevated results almost half the time, directly attributed to times when the economy was accelerating healthily. The calibration of all three models together led to cross-verifications against the boosted stumps, which were tuned to be highly sensitive. This filtered out 98.5% of all underlying ML model noise and allowed critical warnings to trigger 7 times in 42 years.

Table 2: Ensemble probability distribution across 504 months.

Percentile	Ensemble Probability
10th–50th	0.0000
60th	0.0500
70th	0.0668
80th	0.1304
90th	0.1833
100th (Max)	0.6364

Probability Deciles (All 504 Months). For at least 50% of the entire existence, the NIV ensemble outputs a mathematically flat 0.00000. This reinforces the fact that when the NIV ensemble probability moves above 0, there is physical macroeconomic friction.

Confidence Bound Mechanics. All three models do not output just a probability; they also calculate the disagreement between the logistic, boosted, and neural models to generate a lower and upper bound.

Table 3: Mean confidence-bound width by regime.

Regime	**Mean Bound Width**
Green (Healthy)	0.5405
Yellow (Transition)	0.9120

When the economy enters a regime shift, the underlying ML models disagree fully. At the peak of the GFC, the logistic model voted 0.5052, the boosted stumps voted 0.5915, and the neural net voted 0.6642. Due to this structural volatility, the ensemble widened the confidence bounds to the absolute maximum (0.0000 to 1.0000). The model became self-aware of the extreme liquidity issues represented through thrust injections, and correctly flagged the entire environment as highly unstable.

Seven Highest-Risk Months. These were the months where the ensemble systematic risk probability crossed the >35% threshold.

Table 4: The seven months exceeding the 35% ensemble risk threshold.

Date	**Ensemble Prob.**	**Votes (Log / Boost / NN)**	**Macro Context**
Jan 1984	0.6364	0.507 / 0.422 / 0.420	Post-Volcker Inflation
Oct 1994	0.5000	0.616 / 0.515 / 0.539	1994 Bond Massacre
Nov 1994	0.4286	0.616 / 0.314 / 0.580	Fed Hikes 75 bps
Nov 2007	0.3939	0.505 / 0.592 / 0.664	GFC Core Onset
Dec 1994	0.3750	0.655 / 0.354 / 0.582	Soft Landing Friction
Jan 1995	0.3750	0.672 / 0.462 / 0.575	Soft Landing Peak
Feb 1995	0.3750	0.678 / 0.300 / 0.634	Soft Landing Resolution

Final Evaluation. Test 1 proves that NIV generates alpha in the dimensions of detecting systematic risk and leading structural contractions. The single linear model generated 230 false alarms; boosted stumps alone would have generated 50. With the L2-regularized, tanh-bounded ensemble, the noise reduces to 7 critical regime alerts in 42 years.

4.2 Test 2: Multi-Horizon

Model accuracy conventionally decays when a time horizon extends; predicting the economy 3 months out is easier than predicting it 18 months out. The data from Test 2 proves NIV does the opposite. By extracting AUC, Brier Score, Optimal F1, and Optimal Threshold, NIV reveals itself to be a structurally optimized 1.5-year early warning system.

Multi-Horizon Master Matrix. Evaluated on a sliding baseline of 500–512 months with 40 strict systematic contractions.

Table 5: Multi-horizon performance metrics. The model *improves* with longer horizons.

Horizon	Ens. AUC	Log. AUC	Ens. α	Brier	Opt. F1
3 mo	0.7702	0.7434	+0.0268	0.0949	0.3471
6 mo	0.7444	0.7283	+0.0161	0.1160	0.2875
12 mo	0.8243	0.7835	+0.0408	0.0972	0.3590
18 mo	**0.8538**	0.8229	+0.0309	**0.0891**	**0.4545**

Anti-Lag Phenomenon. The mathematical delta between the model's weakest point (6 months) and its absolute peak (18 months) uncovers the full gestation period of systematic risk:

Table 6: Performance surge from 6-month to 18-month horizon.

Metric	Improvement
AUC Discrimination	+14.70%
F1 Score	+58.09%
Brier Score (Calibration Error)	−23.19%

Forensic Proof. The model performs its worst at 6 months (AUC: 0.7444) but dominates at 18 months (AUC: 0.8538). At 6 months prior to a structural breakdown, the thrust of an economy has already been altered to a consistently negative state and headline metrics are artificially elevated by lag time. Then 18 months prior, the liquidity and capital efficiency that compose NIV start to get negatively impacted through M2 constraints or idle capacity growth. It takes 1.5 years for the destruction of regenerative capital to fully metastasize into headline GDP.

Ensemble Alpha. Test 2 allows one to definitively measure whether adding Boosted Stumps and Neural Networks helped the logistic regression or just added noise:

Table 7: Ensemble alpha (AUC improvement) over standalone logistic regression.

Horizon	Alpha
3 mo	+0.0268
6 mo	+0.0161
12 mo	**+0.0408**
18 mo	+0.0309

The highest alpha was +0.0408, occurring exactly at the 12-month horizon. One year out from a regime shift, linear trends begin to break as the economy undergoes high volatility and non-linear regime shifts. The non-linear "if–then" of the boosted stumps was required to bridge the 12-month gap.

Probabilistic Precision. The critical warning thresholds between 31% and 33% were computed for long-horizon forecasts without being explicitly programmed:

Table 8: Naturally converging optimal thresholds at long horizons.

Horizon	Optimal Threshold
12 mo	31.00%
18 mo	33.00%

This empirical validation is the precursor to the wider probability thresholds: yellow at 12–35%, red at >35%. The multi-horizon math confirms that when NIV crosses 31–33%, the economy is set to undergo severe systematic risk in 12–18 months.

Final Assessment of Test 2. By achieving the 58% surge in F1 score and driving the Brier error down to 0.0891 precisely at the 18-month mark, the math proves that the NIV framework is structurally the *opposite* of lagging headline indicators. Its purpose is to isolate the velocity of regenerative capital motion and the exact moment the economy begins to systematically weaken.

4.3 Test 3: Expanding vs. Fixed Window

The purpose of this test was to solve one of the main points of contention in quantitative macroeconomics: Does the model need to remember the entire dataset or only the last 10–15 years?

Flatline Defect (Zero-Probability Analysis). A core issue during macroeconomic modeling is that a model becomes useless if it defaults to 0.00% because it lacks historical context.

Table 9: Zero-probability output frequency by window type.

Window Type	Zero-Output Months	Fraction
Expanding	263	52.2%
Fixed (15 yr)	394	78.2%

The 15-year fixed window is entirely blind for 80% of its existence since it forgets any data older than 15 years, viewing structural macro friction like prolonged zero-interest-rate regimes as normal.

Table 10: Number of months crossing specific risk thresholds by window type.

Threshold	Expanding	Fixed
>10% Risk	131	93
>20% Risk	42	58
>30% Risk	8	40

Risk Threshold Crossings. The expanding window spends 131 months accurately warning of low-to-moderate friction (>10%), but strictly limits high-panic (>30%) warnings to 8 months in 42 years. In contrast, the fixed window sits at 0% most of the time and then violently spikes to >30% fifty times. The fixed window does not ramp up; it only panics.

Table 11: Top positive divergences: where expanding window caught friction but fixed was blind.

Date / Event	Expanding	Fixed	Delta
Oct 1994 (Rate Shock)	50.00%	0.00%	+50.00%
Jan 1995 (Rate Shock)	37.50%	0.00%	+37.50%
Apr 2023 (Rate Hikes)	37.50%	0.00%	+37.50%

Maximum Divergence. The rate dynamics derive from the fact that the expanding window remembers the Volcker shock of the early 1980s, so it could correctly flag an environment as highly stressed. The fixed window had no precedent for 5% interest rates and output a flat 0.00%.

Final Assessment of Test 3. The forensic data rejects the use of a 15-year rolling window. By retaining cumulative economic memory from the 1970s, the expanding window protocol prevents the flatline and captures the full magnitude of systematic stress during every major structural liquidity shock of the last forty years.

4.4 Test 4: GDP Forecast Grid Search (NIV vs. Fed Yield Spread)

The purpose of this test was to benchmark NIV against the Fed Yield Spread (10Y-3M) for GDP forecasting via RMSE.

Volatility and Consistency Matrix. The grid search revealed a stark contrast in stability across all 20 configurations:

Table 12: RMSE volatility comparison across 20 grid-search configurations.

Model	RMSE Std. Dev.
NIV	0.0023
Fed Spread	0.0061

The Fed yield spread accuracy is nearly 3× more volatile based on tuning. The Fed's worst configuration resulted in the largest 0.1662 error, while NIV's worst was 0.1582. This proves the physical data underlying NIV has deeper systematic stability than the bond market.

Table 13: Mean RMSE by lag horizon. NIV dominates at zero lag; Fed at longer lags.

Lag	NIV RMSE	Fed RMSE	Winner
0 mo	0.1511	0.1594	NIV (+0.0083)
3 mo	0.1528	0.1538	NIV (+0.0010)
6 mo	0.1525	0.1493	Fed (−0.0032)
12 mo	0.1513	0.1462	Fed (−0.0050)

Performance by Lag Time. NIV dominates the bond market at no lag, meaning it is optimal for generating current economic snapshots of implied trajectory. The Fed Yield Spread does not achieve its optimal predictive state until a 12-month lag—it takes a year for the bond market to accurately translate the systematic risk that NIV detects immediately.

Table 14: Mean RMSE by smoothing window. NIV needs less smoothing to achieve clarity.

Window	NIV RMSE	Fed RMSE	Winner
3 mo	0.1539	0.1560	NIV (+0.0021)
6 mo	0.1526	0.1535	NIV (+0.0009)
9 mo	0.1519	0.1522	NIV (+0.0003)
12 mo	0.1506	0.1499	Fed (−0.0007)
18 mo	0.1506	0.1493	Fed (−0.0013)

Performance by Smoothing Window. NIV operates best at lower smoothing configurations (3–9 months) compared to the Fed yield curve, which requires 12–18 months of averaging. NIV's mechanics make it inherently less noisy.

Table 15: Best single configurations from the 20-point grid search.

Configuration	NIV	Fed	Note
Best tactical (smooth=12, lag=0)	0.1489	0.1530	Elite real-time
Max delta (smooth=3, lag=0)	0.1520	0.1663	+8.6% rel. improvement

Optimal Configurations.

Final Assessment. NIV is not a secondary delayed indicator to the bond market but rather a complement superior in some regards. The Fed Yield Curve is better for long-range, delayed forecasting (12 months). NIV is mathematically dominant with minimal lag, showing deeper systematic stability.

4.5 Test 5: Component Analysis

The purpose of this test was to address the "black box" problem by isolating the relative weight of each variable's contribution.

Feature Importance Matrix. Using Gini impurity metrics from the ensemble, the algorithm ranked each variable's predictive power:

Table 16: Feature importance scores ranked by Gini impurity reduction.

Feature	Importance
`efficiency_sq`	**0.9328**
`niv_smoothed`	0.5560
`rate_vol`	0.4901
`slack`	0.4260
`niv_acceleration`	0.3759
`niv_percentile`	0.3679
`drag`	0.1719
`thrust`	0.1341
`niv_raw`	0.0408
`real_rate`	0.0352
`spread` (Fed 10Y-3M)	**0.0298**
`niv_momentum`	0.0090

Forensic Proof.

Absolute Dominance of Efficiency. The entire thesis of NIV centered around regenerative capital as the ultimate liberator of economic growth. The algorithm agrees completely: `efficiency_sq` at 0.9328 is the most powerful predictor of structural health in the entire framework.

Irrelevance of Bond Sentiment. Within a multivariate environment, the algorithm determined that the bond market (`spread`) scored only 0.0298. Physical capital efficiency is 31.2× more important to the model's systematic risk detection than the yield curve. The model front-runs the bond market by tracking where the money goes rather than the sentiment guiding it.

The "Sticky Detractors." `rate_vol` (0.4901) and `slack` (0.4260) are the main friction components, meaning the model relies heavily on interest rate instability and idle capacity to confirm when systematic drag overwhelms economic thrust.

Table 17: Forensic breakdown of the last six major systematic contraction cycles.

Target Era	Pre-Contraction Window	NIV at Onset	Fed Spread at Onset	Dominant Trigger
∼1979 Shock	7 Months	2.60	0.57	Thrust
∼1980 Shock	17 Months	−3.08	0.52	Thrust
∼1989 Cycle	9 Months	−1.01	1.89	Thrust
∼2000 Dot-Com	9 Months	6.75	0.91	Thrust
∼2006 GFC	19 Months	−0.31	1.28	Thrust
∼2019 COVID-19	3 Months	2.04	0.41	None (Exogenous)

Structural Onset Diagnostics. Forensic Proof.

Thrust Regimes. In 5 out of 6 modern structural crises, the regime shift was triggered by thrust (proxy of liquidity and investment dynamics). Long before lagging headline metrics register a drop, central banking levers and the plumbing fuelling economic risk-taking fracture, creating a noticeable liquidity issue.

2019/2020 Exogenous Anomaly. The pre-contraction block leading into 2020 identifies the dominant trigger as *none*. The COVID-19 crash was exogenous and not a structural failure of regenerative capital formation. The model correctly identified that macro-thrust and efficiency were not the root causes; the plumbing was relatively stable before this external shock.

Final Assessment of Test 5. This test converted the abstract theory of NIV into hard mathematical laws. It proves that systematic risk prediction is dominated by the squared efficiency of capital (0.9328 importance). Furthermore, modern structural breakdowns derive dominantly from measurable failures in liquidity and investment thrust, reducing the need to rely solely on bond market sentiment.

4.6 Test 6: Forensic Analysis (The "Orthogonality" Proof)

The main criticism of any macroeconomic model is that it may be reverse-engineering the Federal Reserve Yield Curve without adding new information. The purpose of this test is to quantify the overlap and the value of divergence.

Table 18: Correlation and orthogonality between NIV and Fed Yield Spread.

Metric	**Value**
Smoothed Correlation	0.7635
Coefficient of Determination (R^2)	58.29%
Orthogonal Variance	41.71%

Variance Audit. The 0.76 correlation confirms NIV is structurally sound—it correctly shares the baseline macroeconomic reality priced in by the bond market. The 41.71% orthogonal variance is independent of the Yield Curve, mathematically isolating the physical mechanics of capital thrust and idle capacity that cannot be derived from interest rate spreads alone.

Hybrid Allocation Matrix. An optimization algorithm found the best blend of Fed Yield Spread and NIV to maximize structural forecast stability:

Table 19: Optimized hybrid model allocation.

Component	**Weight**
Fed Model	60%
NIV Model	40%

The algorithm assigned a massive 40% structural weighting, proving that relying fully on the bond market leaves structural forecasts informationally incomplete.

Table 20: RMSE comparison: pure Fed model vs. 60/40 hybrid.

Model	**RMSE**
Pure Fed	0.146425
60/40 Hybrid	0.148831
Delta	+0.002407

RMSE Baseline. The hybrid blend incurs a marginal RMSE penalty of +0.0024. However, this sacrifices marginal aggregate accuracy in exchange for massive structural diversification: resistance to false signals when the Federal Reserve manipulates the yield curve, as the 40% NIV allocation enforces real-world physical investment dynamics as a hedge against bond market volatility.

Final Assessment of Test 6. With 41.71% orthogonal variance and a mathematically mandated 40% weighting in an optimally robust hybrid allocation, NIV is proven as an independent and structurally valuable macroeconomic diagnostic tool—not a bond market derivative.

5 Final Assessment of the NIV Framework

Across all six validation tests, the math holds up through the components of thrust, drag, and velocity of regenerative capital. NIV successfully catches the preliminary stressors that precede a systematic collapse long before lagging headline indicators register a shift.

6 Fair Comparison to Alternative Approaches

To justify the creation of NIV, it must be fairly evaluated against current industry-standard methodologies. These approaches have distinct strengths, but NIV adds orthogonal structural value from the physical realities of regenerative capital formation.

The Fed Yield Curve. The Federal Reserve's standard recession probability model relies on the 10Y-3M Treasury yield spread, hinging on the bond market being an effective aggregator of future economic expectations. Its strength lies in simplicity, high liquidity, and real-time aggregation of collective intelligence from massive economic stakeholders.

However, the yield curve measures sentiment and not physical realities. In the modern era of quantitative easing, massive central bank intervention, and forward guidance, the yield curve can become frequently artificially manipulated. It has a reflexive relationship with bond traders' expectations of Fed actions rather than the wider economy's capital base. NIV incorporates the yield curve (yield inversion penalty) but restricts its influence. As shown in Test 6, NIV extracts 41.71% orthogonal variance by tracking the physical plumbing of M2 liquidity and investment dynamics, front-running the bond market's sentiment-based reactions.

Conference Board LEI. The LEI aggregates ten diverse economic components (jobless claims, manufacturing orders, stocks, and credit indices) into a single linear index to predict turning points. Its strength is providing a broad, diverse snapshot across manufacturing, employment, and financial markets.

The structural flaw is that LEI relies on linear arithmetic—it averages all components together. This means it cannot detect non-linear regime changes or complex multi-faceted traps (e.g., manufacturing looking fine while a massive credit drag is building). It also includes coincident or lagging data that convolutes forecasting. By utilizing an L2-regularized ensemble of boosted stumps and neural networks, NIV detects the exact non-linear threshold where systematic risk breaks the economy. The "if–then" triggers from the Boosted Stumps find when components outperform each other (e.g., when Drag overwhelms Thrust), filtering out 98.5% of false alarms as demonstrated in Test 1.

Dynamic Stochastic General Equilibrium (DSGE) Models. DSGE models are the primary academic and central banking standard: massive, complex theoretical models that simulate the entire economy by modeling household and firm behavior responding to shocks over time. Their strength is mathematical rigor and a highly structured theoretical framework for how policy changes should impact human behavior.

The structural flaw is the assumption of equilibrium return. In the lead-up to the 2008 GFC, standard DSGE models completely failed because they did not model the friction of the financial sector and credit plumbing—they assumed money flowed without resistance. NIV entirely removes the assumption of economic equilibrium. Designed from a systems architecture perspective, it assumes the economy is a fluid, dynamic network that can accumulate massive friction. By measuring Slack (idle capacity) and Drag (credit circulation inhibitors), NIV accurately models the plumbing failures that DSGE models historically ignore.