Market Troughs as Bifurcation Events: A Liquidity Spiral Framework and Empirical Validation

Peilin Rao*[1] and Randall R. Rojas†[1]

[1]Department of Economics, University of California, Los Angeles, US

Nov 2, 2025

Abstract

Why do linear models perform well at predicting V-shaped market troughs (ROC AUC = 0.9435), but the underlying theory (liquidity spirals) is inherently non-linear? We show that this is a false dichotomy and demonstrate that the canonical Brunnermeier and Pedersen (2009) model implies both a partially linear bifurcation boundary and non-linear structural interactions (spirals) that drive the market state. We provide empirical evidence for each structure: a linear SVM (the correct tool for the boundary) outperforms non-linear classifiers, while a non-linear DML-APE (the correct tool for the spirals) succeeds in identifying the key drivers of market fragility, whereas a restrictive linear causal model produces spurious results and consistently misidentifies the direction of critical risk drivers. Our research provides strong validation of the Brunnermeier and Pedersen (2009) mechanism by demonstrating that its linear and non-linear elements are both observable in the data.

*jackrao@g.ucla.edu
†rrojas@econ.ucla.edu
Corresponding author: Peilin Rao.

1 Introduction

Financial market crises are associated with two central puzzles. The first is economic: what is the mechanism responsible for the fast, V-shaped price movements that occur during low points in markets? These crises, as shown in this paper, are triggered by a sudden shock followed by a very quick resolution, implying a "quick" mechanism. The econometric puzzle is: Why does a simple linear classifier perform so well at predicting complex, nonlinear crises? In this paper, we will develop an empirical, theory-based framework that shows the econometric puzzle (linearity) is actually the key to verifying the economic mechanism (a "fast moving" liquidity spiral).

There are two main categories of theoretical models of financial crises. The first type of model — the "slow-moving" model of Intermediate Asset Pricing (IAP) (He and Krishnamurthy, 2013) — does a good job of describing how impaired balance sheets in a financially-constrained intermediate sector lead to increased risk premia and a long, U-shaped recovery as intermediaries gradually rebuild capital. The second type — the "fast-moving" model of Market and Funding Liquidity (ML/FL) developed by Brunnermeier and Pedersen (2009)(henceforth B&P) — describes how an acute shock to funding can create a rapidly expanding, self-reinforcing loss/margin spiral that creates large price distortions and resolves them equally quickly, once the temporary friction disappears. Our empirical documentation in Section 2.1 of the V-shaped regularity observed at market troughs strongly supports our motivation for the ML/FL framework as the more suitable lens through which to view high-frequency market crashes.

However, the theoretical distinction between these two types of models has been blurred in the empirical literature. The field currently regards "slow-moving" (He and Krishnamurthy, 2013) and "fast-moving" (Brunnermeier and Pedersen, 2009) models as complementary parts of a single "intermediary asset pricing" paradigm (Adrian et al., 2014; He et al., 2017). There are therefore two significant gaps in the existing research. First, the empirical tests of these models have never included a direct "horse race" comparison of the two model

types' competing predictions: the U-shaped recovery of IAP versus the V-shaped reversal of B&P. Most of the research on recovery shapes has remained primarily descriptive and macroeconomic (Gualdi et al., 2021), and has not been framed as a test of micro-founded asset pricing models. Second, the empirical work on the B&P model has focused on the consequences of the model, such as the existence of liquidity spirals, or different regimes (Gromb and Vayanos, 2018; Flood et al., 2014), rather than on the antecedents of the model: specifically, the high-frequency bifurcation boundary that theoretically causes these events.

In addition to the above gaps in the empirical literature on crises models, there is also a gap in the empirical literature on crisis prediction. This area of research has predominantly used machine learning (ML) methods in a statistical "horse race" to determine which model best predicts crises, with the focus solely on achieving maximum predictive accuracy (Gu et al., 2020; Bluwstein et al., 2020). Therefore, whether linear (e.g., Lasso) or non-linear (e.g., Random Forest) models are employed is considered a statistical decision based on practical considerations, not an economic decision. Thus, this paper presents a unique opportunity that has gone unexplored. Specifically, no previous paper has "turned" this problem around. We are the first to employ the relative out-of-sample performance of a linear versus a non-linear classifier as a formal "test" to validate the geometric structure of an economic theory.

Lastly, while the recent advent of Double/Debiased Machine Learning (DML) (Chernozhukov et al., 2018) holds great promise for providing robust causal inference, the application of DML to finance is still nascent and subject to significant specification risks. Notably, the finance literature has yet to address a fundamental methodological concern: if a restrictive, linear-causal model (DML-PLR) is misspecified and employed to test a theory that is, by definition, non-linear and interactive, what will be the implications of this theoretical misspecification on the empirical results? We are the first to demonstrate that this theoretical misspecification results in a catastrophic empirical failure — namely, generating multiple spurious findings and, in many cases, systematically reversing the sign of a variable's true effect — and that only a theory-consistent, interactive model (DML-APE) can correct this

failure and accurately measure the determinants of fragility.

We solve the "linear vs. non-linear" puzzle by demonstrating that the question is fundamentally a false dichotomy. We demonstrate that the canonical Brunnermeier and Pedersen (2009) model implies a dual structure that has not been formally tested previously: (1) a (partially) linear bifurcation boundary, and (2) non-linear, interactive spirals (the mechanism) that drive the market toward that boundary. Therefore, this paper represents the first formal test of both components of the dual structure using a "right tool for the job" approach.

First, to formally test the boundary, we treat it as a classification problem. We demonstrate that a linear-kernel Support Vector Machine (SVM) — the theoretically appropriate tool for a linear boundary — not only performs well (ROC AUC 0.9435) but significantly outperforms its flexible, non-linear (RBF-kernel) competitor. We have thus formally validated the "partially linear" boundary structure implied by the B&P theory.

Second, to formally test the mechanism, we treat it as a structural estimation problem using DML. The DML analysis confirms our dual hypothesis: a restrictive linear-causal model (DML-PLR), which ignores the theory's interactions, fails. It produces dozens of spurious findings and, importantly, reverses the sign of key drivers like tail-risk pricing. Conversely, our flexible, interaction-based model (DML-APE), which incorporates the B&P theory's non-linearities, correctly identifies the destabilizing power of the true non-linear drivers, such as acute order-flow imbalances. This demonstrates that the proper specification of a theory is a necessary condition for obtaining robust empirical evidence.

Having demonstrated that the B&P framework is the appropriate lens, we provide a final piece of validation. We present a novel diagnostic test on the model's signal (Section 6) and demonstrate that the model's predictive power captures both the "fast" V-shaped reversal and the subsequent "slow" U-shaped recovery, with profitability persisting for 20-day holding periods. We conclude that the B&P spiral serves as the trigger for the extended IAP recovery.

This paper provides three main contributions:

- First, we solve the "linear vs. non-linear" modeling puzzle. Specifically, we prove that the Brunnermeier and Pedersen (2009) model has a dual structure: a partially linear boundary and a nonlinear, interactive mechanism. We are the first to empirically validate both components of the dual structure—using a linear SVM to establish the boundary and a flexible DML-APE to establish the mechanism, thereby solving the contradiction between the two representations.

- Second, we provide a new, two-stage validation of the liquidity spiral framework. First, we show (a) that our model's signal captures both the "fast" V-shaped reversal and the subsequent "slow" U-shaped recovery, with profitability remaining robust for 20-day holding periods. Thus, we imply that "slow-moving" (IAP) and "fast-moving" (B&P) models are not competing explanations, but complementary ones: the B&P spiral acts as the bifurcation mechanism initiating the protracted IAP recovery. Second, we show (b) that the estimated linear boundary is generated by the B&P proxies for margin sensitivity ($\bar{\theta}$) and speculator fragility ($\mathbf{x}_0$).

- Third, we demonstrate that respecting this theoretical foundation is necessary for obtaining robust empirical estimates. We demonstrate that a restrictive, linear-causal model (DML-PLR) is misspecified and therefore fails to produce reliable results; specifically, the model produces dozens of spurious findings and consistently reverses the sign of the key risk drivers. Conversely, our theory-consistent, interaction-based model (DML-APE) reliably measures the key drivers of systemic instability. The failure of the linear additive model can be seen as a direct warning for risk managers: linear factor models will not measure risk accurately when it counts most —that is, in the lead-up to a spiral — because these models do not incorporate the very powerful nonlinear instabilities (e.g., in acute order flow dynamics and tail-risk pricing) identified by our model as the main drivers of systemic fragility.

The remainder of the paper is organized as follows. Section 2 derives our testable

hypotheses from the Brunnermeier and Pedersen (2009) model. Section 3 describes our dataset and how we constructed our empirical state vector. Section 4 describes our predictive modeling framework. Section 5 discusses the predictive validation of our theory. Section 6 discusses the economic significance of the model's signal. Section 7 discusses the causal mechanisms of systemic fragility. Section 8 concludes.

2 Theoretical Framework and Hypotheses

In this section, we describe the theoretical microfoundation for the empirical study that follows. First, we identify a new stylized fact – "V-shaped anatomy" of market troughs – that motivates the theoretical lens we use. We then argue that the stylized fact is best described using fast-moving liquidity spirals (Brunnermeier and Pedersen, 2009) versus slow-moving balance sheet impairment (He and Krishnamurthy, 2013). We then establish a testable theoretical foundation by revisiting the seminal (Brunnermeier and Pedersen, 2009) (hereafter B&P) model. We demonstrate how the nonlinear stability condition implied by the B&P model indicates that the transition into a fragile state is a bifurcation event with a partially linear structure. This theoretical implication provides a non-trivial microfoundational justification for the 'linear model puzzle' discussed in Section 1, and leads directly to the sharp hypotheses that define our empirical methodology.

2.1 A Motivating Stylized Fact: The V-Shaped Anatomy of a Market Trough

We begin by establishing the empirical evidence for our theoretical discussion. We undertake a formal event study on the $N = 7$ trough dates for our target event as established in Section 3.1. Our event study analyzes the average behavior of our empirical proxy measures over a 21-day period (from $T - 10$ to $T + 10$) centered on each trough date.

Figure 1 illustrates the average normalized dynamics of our empirical proxy measures of

panic, funding stress, and speculative fragility. Panels A and B indicate that our measures of panic — specifically, the VIX Index and realized volatility — display a sudden, rapidly increasing trajectory that peaks at or just before $T = 0$. Furthermore, Panel B shows that our measures of speculative fragility (i.e., our measure of unrealized profits) and funding stress (i.e., our measure of credit spreads) display a simultaneous and acute peak. Importantly, our measure of unrealized profits displays a perfect V shape, indicating a rapid decline in speculative wealth followed by a reversal rapid recovery. These empirical proxy measures are formally defined in Section 3.

Our empirical analysis of the anatomy of a market trough in our sample therefore demonstrates a rapid, acute crisis followed by a rapid resolution of the crisis. This 'V-shaped' empirical evidence clearly suggests a mechanism for acute onset and rapid resolution of a crisis rather than one of depletion of a constrained intermediary's net worth and prolonged recovery.

2.2 Choosing the Right Lens: Liquidity Spirals vs. Intermediary Constraints

The stylized fact presented in Section 2.1 presents a clear criterion for choosing a suitable theoretical framework. Two primary theoretical frameworks exist in the literature for modeling financial crises, which we refer to as 'slow-moving' and 'fast-moving'.

Intermediary Asset Pricing (IAP) framework of He and Krishnamurthy (2013) represents crises as resulting from 'slow-moving' balance sheet impairment. In this context, a negative shock reduces the net worth ($\mathbf{w}_t$) of a specialized and constrained intermediary sector. As the intermediary's capital binds, their effective risk aversion (α_t^I) increases, thereby raising market-wide risk premia. The IAP framework is particularly successful in explaining the long recovery times from crises because the intermediary sector must slowly rebuild capital through retained earnings, a process that, according to the authors' calibration, takes several months. However, the slow-moving nature of the IAP framework alone does not explain the

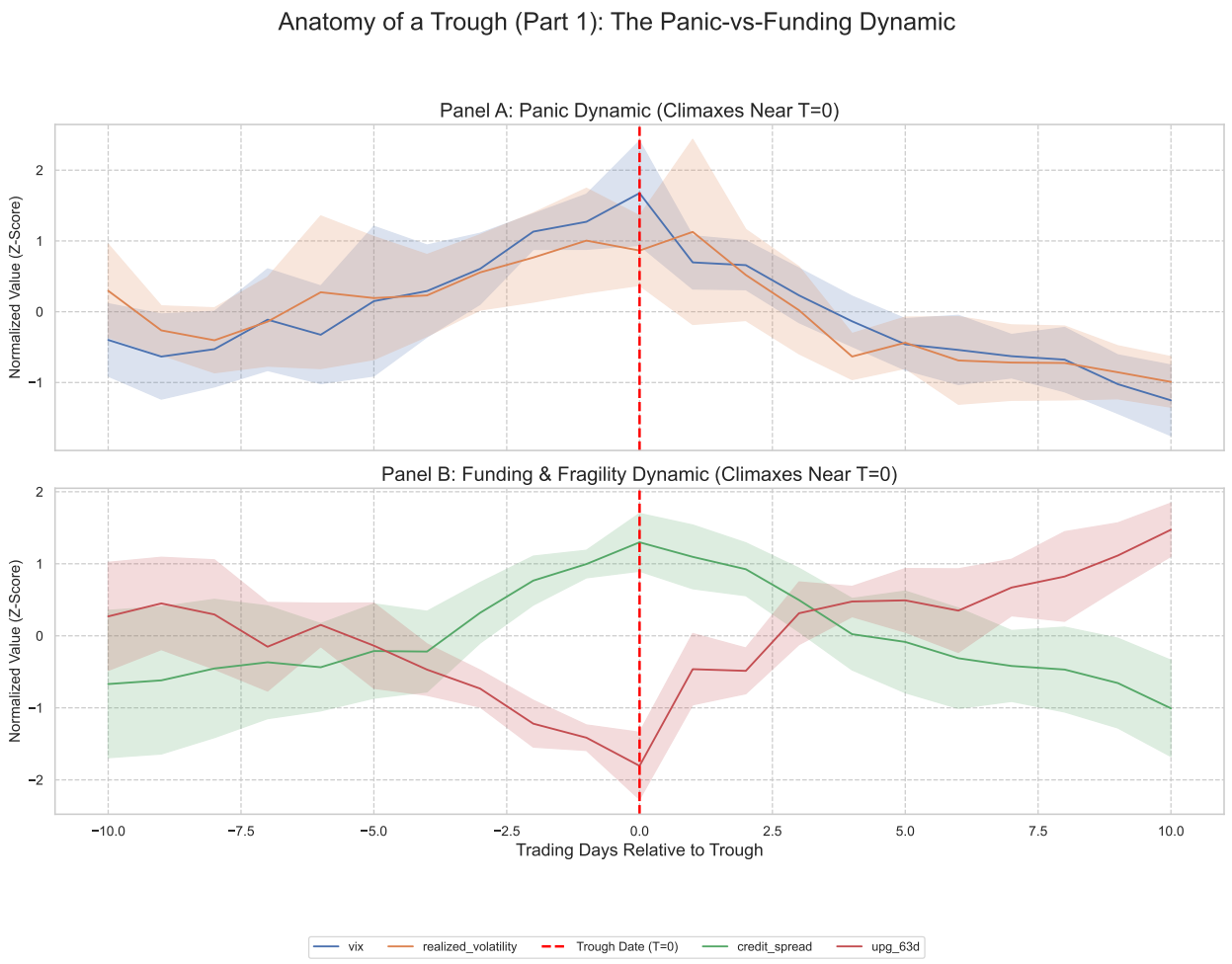


Figure 1: Anatomy of a Trough: Panic and Funding Dynamics

Notes: This figure plots the average normalized Z-score of key indicators in the 21-day window around the $N = 7$ identified market troughs ($T = 0$). Panel A shows proxies for market panic (vix, r ealized_volatility). Panel B shows proxies for funding stress (credit_spread) and speculator fragility (upg_63d). The shaded regions represent 95% confidence intervals. The plots demonstrate a sharp, simultaneous climax of all indicators at or near $T = 0$, consistent with a V-shaped, acute crisis event. The specific indicators plotted are our empirical proxies for these concepts, formally constructed in Section 3.

high-frequency, acute trigger of the trough itself (see Figure 1).

Therefore, we turn to the second class of models — Market and Funding Liquidity (ML/FL) — whose foundational research was conducted by Brunnermeier and Pedersen (2009). The ML/FL framework represents crises as 'fast-moving' events that arise due to a combination of a funding shock and a temporary market friction (for example, the sequential arrival of customers). A funding shock to risk-neutral speculators tightens their capital constraints, thereby reducing their provision of liquidity. This triggers two self-reinforcing feedback mechanisms: a Loss Spiral, where the initial losses lead to tighter constraints, and a Margin Spiral, where declining asset prices lead to higher volatilities, causing financiers to raise margin requirements, which further tighten constraints. The rapid, amplified price dislocation caused by these cascading effects creates an acute price drop. Critically, the price will rebound immediately after the temporary friction (order imbalance) has been resolved.

Given the ML/FL framework's predictions of both a rapid price drop and an immediate price reversal, it provides the most suitable microfoundational theoretical basis for the bifurcation event that we seek to model. Therefore, we choose the B&P model as the theoretical framework for our analysis of the trough event itself, while acknowledging that the IAP framework may complement the ML/FL framework's explanation of the subsequent recovery.

2.3 The B&P Model as a Bifurcation: The "Partially Linear" Stability Boundary

To build a testable framework from the "fast-moving" spiral, we revisit the canonical B&P model to analyze the precise moment of market collapse. We show that the model implies that the transition to a fragile, multi-equilibrium state is a bifurcation event. The condition for this event, while non-linear, possesses a crucial "partially linear" structure. This provides a direct, testable microfoundation for the "linear model puzzle" introduced in Section 1 and motivates our entire empirical strategy.

2.3.1 Fragility and the Bifurcation Boundary

We adopt the B&P environment with risk-neutral, capital-constrained speculators, risk-averse customers, and uninformed financiers. Speculators provide liquidity to customers, who experience endowment shocks (Z_1 at time $t = 1$). Speculators' ability to trade is limited by a margin constraint, which must be less than their available capital $\mathbf{W}_1$.

The market's stability is determined by the properties of the speculator's Net Funding Requirement function, $G(\mathbf{p}_1)$. This function represents the speculator's total margin requirement less their available capital:

$$G(\mathbf{p}_1) = m_1^+ \mathbf{x}_1 - \mathbf{W}_1^{\text{available}}$$

where $\mathbf{x}_1$ is the speculator's long position, m_1^+ is the per-share margin, and $\mathbf{W}_1^{\text{available}}$ is their available capital. The equilibrium price $\mathbf{p}_1$ is the value that solves $G(\mathbf{p}_1) = 0$ (assuming the constraint binds).

- A Stable Market is one where $G(\mathbf{p}_1)$ is monotonic. A unique equilibrium price exists, and shocks are absorbed smoothly.
- A Fragile Market is one where $G(\mathbf{p}_1)$ is non-monotonic (S-shaped). Multiple equilibria exist (a "good" high-price equilibrium and a "bad" crisis low-price equilibrium). A small shock can push the market from the good to the bad equilibrium, causing a price crash.

The gateway to this fragile regime is the bifurcation point, the mathematical condition where the slope of the funding requirement function first touches zero, allowing for the creation of multiple equilibria. Our analysis, therefore, centers on solving for this critical boundary:

$$\frac{dG(\mathbf{p}_1)}{d\mathbf{p}_1} = 0$$

2.3.2 Derivation of the Bifurcation Condition

Following B&P, we analyze the case of uninformed financiers (who set margins based on observed price volatility) and customer selling pressure ($Z_1 > 0$). This assumption is crucial as it gives rise to the destabilizing margin spiral. The components of $G(\mathbf{p}_1)$ are:

- **Margin (m_1^+):** $m_1^+ = \bar{\sigma} + \bar{\theta}(\mathbf{p}_0 - \mathbf{p}_1)$. Margin increases as price $\mathbf{p}_1$ falls ($\bar{\theta} > 0$).
- **Position ($\mathbf{x}_1$):** $\mathbf{x}_1 = Z_1 + \frac{2(\mathbf{p}_1 - \mathbf{v}_1)}{\gamma(\sigma_2)^2}$. Speculators' demand increases as price $\mathbf{p}_1$ falls further below fundamental value $\mathbf{v}_1$.
- **Capital ($\mathbf{W}_1^{\text{available}}$):** $\mathbf{W}_1^{\text{available}} = \mathbf{W}_0 + \eta_1 + \mathbf{p}_1\mathbf{x}_0$. Available capital depends on initial capital $\mathbf{W}_0$, funding shocks η_1, and mark-to-market gains/losses on their initial position $\mathbf{x}_0$.

To find the bifurcation point, we differentiate $G(\mathbf{p}_1) = m_1^+\mathbf{x}_1 - \mathbf{W}_1^{\text{available}}$ with respect to $\mathbf{p}_1$ using the product rule:

$$\frac{dG}{d\mathbf{p}_1} = \left[\frac{d(m_1^+)}{d\mathbf{p}_1}\mathbf{x}_1 + m_1^+\frac{d(\mathbf{x}_1)}{d\mathbf{p}_1}\right] - \frac{d(\mathbf{W}_1^{\text{available}})}{d\mathbf{p}_1}$$

We derive each component's derivative:

- $\frac{d(m_1^+)}{d\mathbf{p}_1} = -\bar{\theta}$. This is the destabilizing margin spiral.
- $\frac{d(\mathbf{x}_1)}{d\mathbf{p}_1} = \frac{2}{\gamma(\sigma_2)^2}$. This is the stabilizing, value-based demand.
- $\frac{d(\mathbf{W}_1^{\text{available}})}{d\mathbf{p}_1} = \mathbf{x}_0$. This is the destabilizing loss spiral.

Substituting these back yields the full expression for the market's stability, which we decompose into its three core economic forces:

$$\frac{dG}{d\mathbf{p}_1} = \underbrace{(-\bar{\theta})\left(Z_1 + \frac{2(\mathbf{p}_1 - \mathbf{v}_1)}{\gamma(\sigma_2)^2}\right)}_{\text{Term 1: Margin Spiral (Destabilizing)}} + \underbrace{(\bar{\sigma} + \bar{\theta}(\mathbf{p}_0 - \mathbf{p}_1))\left(\frac{2}{\gamma(\sigma_2)^2}\right)}_{\text{Term 2: Value Demand (Stabilizing)}} - \underbrace{\mathbf{x}_0}_{\text{Term 3: Loss Spiral (Destabilizing)}}$$

The market becomes fragile when the destabilizing forces (Term 1 and Term 3) overwhelm the stabilizing force (Term 2), causing the slope $\frac{dG}{d\mathbf{p}_1}$ to become negative. The bifurcation boundary is where they are precisely balanced: $\frac{dG}{d\mathbf{p}_1} = 0$.

2.3.3 The "Partially Linear" Structure of the Bifurcation Boundary

The bifurcation condition $\frac{dG}{d\mathbf{p}_1} = 0$ is a non-linear equation, as the endogenous equilibrium price $\mathbf{p}_1$ appears in both Term 1 and Term 2. As shown in the B&P model, $\mathbf{p}_1$ is itself a complex, non-linear (in fact, quadratic) function of the model's state variables.

However, our key theoretical insight is that this non-linear boundary possesses a "partially linear" structure. We can reveal this structure by rearranging the equation, grouping the terms that are explicitly linear in the key destabilizing state variables Z_1 (Order Imbalance) and $\mathbf{x}_0$ (Initial Position):

$$\underbrace{(-\bar{\theta})Z_1 - \mathbf{x}_0}_{\text{Part 1: Linear Spiral Drivers}} + \underbrace{\left[(-\bar{\theta})\frac{2(\mathbf{p}_1 - \mathbf{v}_1)}{\gamma(\sigma_2)^2} + (\bar{\sigma} + \bar{\theta}(\mathbf{p}_0 - \mathbf{p}_1))\left(\frac{2}{\gamma(\sigma_2)^2}\right)\right]}_{\text{Part 2: Non-Linear Threshold } C(\mathbf{p}_1, \mathbf{\Theta}_t)} = 0$$

This rearrangement reveals the core implication of the B&P stability analysis. The market's stability is determined by two additively separable components:

1. A component that is linear in the primary destabilizing state variables: the customer order imbalance Z_1 (driving the Margin Spiral) and the initial speculator position $\mathbf{x}_0$ (driving the Loss Spiral).

2. A complex, non-linear threshold term $C(\mathbf{p}_1, \mathbf{\Theta}_t)$ that encapsulates all other factors, most notably the endogenous equilibrium price $\mathbf{p}_1$ and other model parameters $\mathbf{\Theta}_t$ (volatility, risk aversion, etc.).

The boundary condition can therefore be written abstractly as:

$$\boxed{\boldsymbol{\alpha}'\mathbf{S}_t + C(\mathbf{p}_1, \boldsymbol{\Theta}_t) = 0 \quad \text{where} \quad \mathbf{S}_t = [Z_1, \mathbf{x}_0]' \quad \text{and} \quad \boldsymbol{\alpha} = [-\bar{\theta}, -1]'}$$

This "partially linear" structure provides a direct, non-obvious microfoundation for our empirical puzzle. A linear classifier may be the econometrically appropriate tool not because the underlying economic mechanism is simple, but because its dominant drivers enter additively and the remaining non-linearity can be effectively controlled for.

2.4 From Theory to Testable Hypotheses

Our theoretical analysis of the B&P framework's partially linear structure produces two clear testable hypotheses.

Our first hypothesis reframes the "linear model puzzle from the introduction. The task is to estimate a non-linear bifurcation boundary equation $\boldsymbol{\alpha}'\mathbf{S}_t + C(\mathbf{p}_1, \boldsymbol{\Theta}_t) = 0$. We conjecture that a parsimonious linear model can outperform a complex non-linear model on this goal, based on a feature set that primarily contributes to two parts of the B&P theoretical framework: (1) the use of proxies for $\mathbf{S}_t = [Z_1, \mathbf{x}_0]'$ to estimate the linear component $\boldsymbol{\alpha}'\mathbf{S}_t$, and (2) proxies for the determinants of $\mathbf{p}_1$ and $\boldsymbol{\Theta}_t$ (e.g., credit spreads, VIX) as linear controls for the complex and non-linear threshold $C(\cdot)$. We hypothesize that this may produce a better approximation of the true boundary than a fully flexible, non-linear model (such as an RBF-kernel) that seeks to approximate the complex $C(\cdot)$ function itself.

- **Hypothesis 1 (The Robust Approximation Hypothesis):** The "partially linear structure" of the B&P model (see Section 2.3) indicates that the major crisis triggers ($\mathbf{S}_t$) enter the stability conditions linearly. Therefore, we conjecture that a parsimonious linear approximation (for example, a linear-kernel SVM) will prove to be a more robust estimator of this boundary and will have superior or statistically identical out-of-sample performance relative to a flexible, non-linear classifier (for example, an RBF-kernel SVM) that is theoretically more suited to modeling the full non-linearity of $C(\cdot)$ but is

more prone to in-sample overfitting.

It is insufficient to simply confirm the form of the bifurcation boundary (H1); the economic drivers behind the classifier that approximates the boundary must be confirmed as well. If the empirically estimated hyper-plane $\mathbf{w}'\mathbf{x}_t + b = 0$ is indeed capturing this B&P mechanism, then the most important features in this empirically estimated hyper-plane should be the features that are critical and directly correspond to deep parameters in the theory. We establish this correspondence in Table 1 where we map observable proxies to both components of our theoretical boundary. This establishes the basis for our second hypothesis.

Table 1: Mapping Theoretical Concepts to Empirical Proxies

Theoretical Variable	**Economic Meaning**	**Examples of Empirical Proxies**
Part 1: Proxies for Linear Drivers ($\mathbf{S}_t$)		
Z_1	Customer Order Imbalance	flow_concentration_..., ofi_...
$\mathbf{x}_0, \mathbf{W}_1$	Speculator Position / Losses	upg_63d, dex_oi_...
Part 2: Proxies for Non-Linear Threshold ($C(\mathbf{p}_1, \mathbf{\Theta}_t)$)		
$\bar{\theta}$	Margin Sensitivity	gex_..., vix, realized_volatility
$C(\cdot)$	Stability Threshold / Funding	credit_spread, amihud_illiquidity

Notes: This table maps the latent variables from the Brunnermeier and Pedersen (2009) "partially linear" bifurcation boundary to the observable features engineered in our dataset. The full construction of these proxies is detailed in Section 3.

- **Hypothesis 2 (The Economic Driver Hypothesis):** The empirically estimated hyper-plane should be driven by those features that are proxies for the key economic mechanisms identified by the theory. Specifically, SHAP analysis should identify the empirically estimated hyperplane to be primarily determined by those features that relate to both the linear drivers (e.g., speculator fragility $\mathbf{x}_0$) and the non-linear threshold controls (e.g., margin sensitivity $\bar{\theta}$ and funding stress $C(\cdot)$).

We can frame the market's progression into a trough within the context of the B&P theoretical framework as a Two-Act Play:

1. **Act 1: The Fragile State:** The market becomes fragile as the preconditions worsen. The linear drivers (e.g., $\mathbf{x}_0$ grows due to increasing speculator losses) and the threshold controls (e.g., $\bar{\theta}$ increases as margin sensitivity rises) drive the market state $\mathbf{S}_t$ closer to the boundary.

2. **Act 2: The Dynamic Spiral:** A severe shock (e.g., to Z_1 or η_1) provides the final impetus, pushing the market state to exceed the bifurcation boundary $C(\mathbf{\Theta}_t, \mathbf{p}_1)$ and trigger the price collapse.

Before constructing the predictive models, we stylistically verify whether the Two-Act Play is plausible in our data. To this end, we conduct two preliminary verification tests for each of the acts described above.

- **Verification of the "Fragile State" (Act 1):** We perform a Level Test to examine the average state of the proxies in our dataset during the $T-5$ to $T-1$ time window. As depicted in Figure 2, the market is already in a highly fragile state. Market illiquidity proxies (amihud_illiquidity) and persistent selling pressure proxies (flow_concentratio n_10d_ewm_spread) are at statistically significant lows (t-statistics of -3.57 and -4.67, respectively). Additionally, the fragility of the speculator (unrealized losses upg_63d, the dex_oi_trend_z) and the sensitivity of margins (gex_volume_trend_z) are at very low levels (t-statistic of -7.71, -3.57, and -3.57, respectively). (See Section 3 for the formal definition of these indicators.)

- **Verification of the "Dynamic Spiral" (Act 2):** We perform an acceleration test to compare the time periods $T-5:T-1$ and $T-10:T-6$. The results (summarized in Table 2) indicate that the measures associated with the spiral are accelerating rapidly toward the trough. The measures of panic in the market (VIX), funding stress (rate-of-change in credit spread) and speculator's losses (unrealized profits) are accelerating rapidly (t-statistics of 5.61, 5.64, and -3.68, respectively), indicating the presence of an actively developing spiral that drives the market to cross the boundary.

We have established that the Two-Act Play, derived by the B&P framework, is consistent with the stylized facts in the empirical data. We now proceed to formally test Hypotheses 1 and 2.

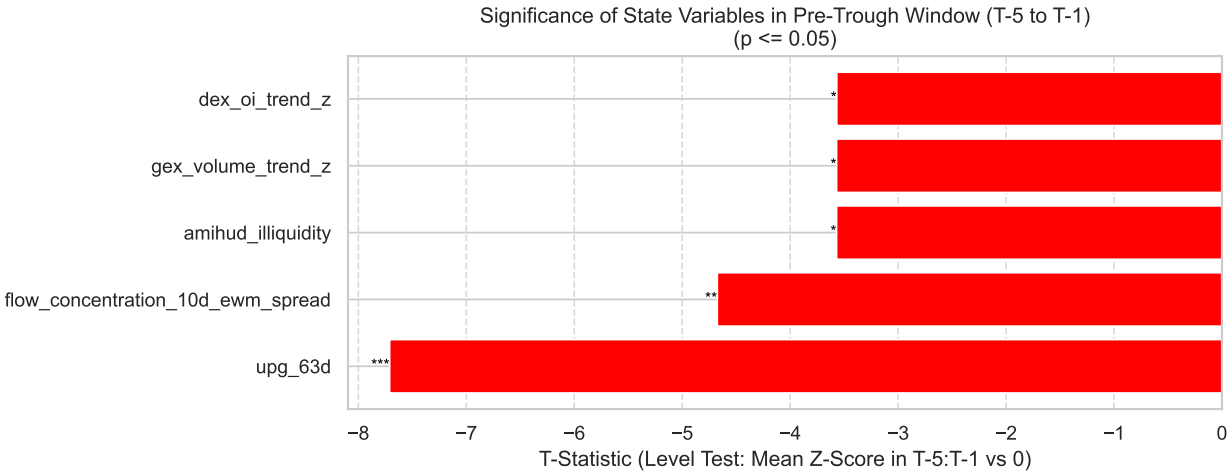


Figure 2: Validation of the "Fragile State" (Level Test)

Notes: This figure plots the t-statistics for the Level Test, which measures if the average normalized value of a proxy in the $T-5$ to $T-1$ window is significantly different from zero. The results confirm that the market is in a state of extreme, statistically significant fragility before the trough date. Significance levels are denoted by '*' ($p < 0.05$), '**' ($p < 0.01$), and '***' ($p < 0.001$). The specific proxies shown are formally constructed in Section 3.

3 Data and Empirical Construction of the State Vector

This section details the construction of our empirical framework. We first present our objective, rule-based definition of the target event—the market trough. We then detail the core of our empirical exercise: the construction of a high-dimensional state vector $\mathbf{X}_t$ designed to serve as an observable proxy for the latent theoretical variables derived in Section 2. We conclude by describing the raw data sources and the statistical transformations applied to finalize the feature set.

Table 2: Acceleration Test Results for Key Proxies

Proxy	**t-statistic**
(Panic Proxies)	
vix	5.61***
vix_roc63	6.01***
realized_volatility	2.25*
(Funding Stress Proxies)	
credit_spread_roc63	5.64***
credit_spread_ewm_slow	6.19***
(Speculator Fragility Proxies)	
upg_63d	-3.68**
upg_63d_ewm_slow	-12.05***
upg_63d_roc63	-5.20***
dex_oi_ewm_slow	-8.54***
gex_oi_ewm_slow	-5.97***

Notes: This table reports the t-statistics for the Acceleration Test, comparing the mean of each proxy in the $T-5:T-1$ window against the $T-10:T-6$ window. A positive t-statistic indicates acceleration (or worsening) of the variable into the trough. Significance levels: '*' $p < 0.1$, '**' $p < 0.05$, '***' $p < 0.01$. The construction of the specific proxies listed in this table is detailed in Section 3.

3.1 Defining the Target Event: Trough Identification

Our empirical objective is to test whether the current market state, $\mathbf{X}_t$, aligns with the theoretical conditions of a bifurcation boundary. This requires an objective, non-arbitrary label for what constitutes a trough event ($\mathbf{y}_t = 1$). We explicitly frame this task as a nowcast, not a forecast. This is analogous to the real-time detection of economic recessions, where the official NBER labels are also confirmed with a long ex-post lag. Our model's goal is to estimate $P(\mathbf{y}_t = 1|\mathcal{X}_t)$—that is, the probability that the current state $\mathcal{X}_t$ will be ex-post confirmed by an objective algorithm as a market trough. This framework is crucial for avoiding data leakage. The predictive features $\mathcal{X}_t$ are strictly pre-determined and based only on information available at or before time t. The label $\mathbf{y}_t$ is the only component derived from ex-post analysis. This design allows us to test our hypothesis about the current state's properties without circularity.

To generate these objective, ex-post labels, we adapt the Bry and Boschan (1971) algorithm. The Bry-Boschan (BB) algorithm is a rule-based process to date business cycles. We adapt its methodology to identify significant peaks and troughs in the daily S&P 500 log adjusted closed daily price series (P_t) by systematiclly removing minor price movements. The overall high-level procedure is outlined in Algorithm 1. The main procedure consists of first identifying all potential turning points of the price series, and then applying censoring rules. The complete implementation, including all helper functions, are provided in Algorithm 2 in Appendix A. To ensure all turning points within our primary sample (ending June 2025) are algorithmically identified without end-of-sample ambiguity, we apply the BB algorithm to a price series extended to August 12, 2025. This extended data is used only for label generation and is excluded from all feature engineering and model evaluation.

Utilizing the modified BB algorithm, We identify turning points with economic significance. We show in Table 3 the peaks and troughs that are identified in the sample period. The troughs in this table form the positive class labels for the market trough prediction model. In Figure 3, we illustrated these troughs compared against the S&P 500 price series, and show

Algorithm 1 High-Level Bry-Boschan Algorithm

Require: Log price series P_t, window order, min_phase, min_cycle.
Ensure: A DataFrame turns of significant peaks and troughs.

```
procedure IdentifyTurns(P_t, order, min_phase, min_cycle)
    Initialize turns with all local peaks and troughs from P_t, sorted by date.
    turns ← EnforceAlternation(turns)          ▷ Enforce P-T-P-T sequence.
    turns ← CensorPhases(turns, min_phase)     ▷ Censor short phases.
    turns ← CensorCycles(turns, min_cycle, P_t)  ▷ Censor short cycles.
    return turns
end procedure
```

how the algorithms marks distinct main market troughs effectively.

3.2 Constructing the State Vector Proxies

Our theory, culminating in the bifurcation boundary $\boldsymbol{\alpha}'\mathbf{S}_t = C(\boldsymbol{\Theta}_t, \mathbf{p}_1)$, identifies the latent state variables that govern market fragility. The central task of our empirical design is to construct an observable, high-dimensional vector $\mathbf{X}_t$ that serves as an empirical proxy for this

Table 3: Identified S&P 500 Turning Points (2013-2025)

Date	Log Price	Type
2013-04-18	7.3406	Trough
2015-05-21	7.6643	Peak
2016-02-11	7.5116	Trough
2018-09-20	7.9830	Peak
2018-12-24	7.7626	Trough
2020-02-19	8.1274	Peak
2020-03-23	7.7131	Trough
2022-01-03	8.4757	Peak
2022-10-12	8.1823	Trough
2023-07-31	8.4314	Peak
2023-10-27	8.3230	Trough
2025-02-19	8.7233	Peak
2025-04-08	8.5137	Trough

Notes: This table lists the economically significant market peaks and troughs identified in the daily S&P 500 log price series. The turning points are determined using a modified version of the Bry and Boschan (1971) algorithm, as described in Section 3.1, with no manual intervention. The trough dates are used to generate the positive labels ($\mathbf{y}_t = 1$) for the classification model.

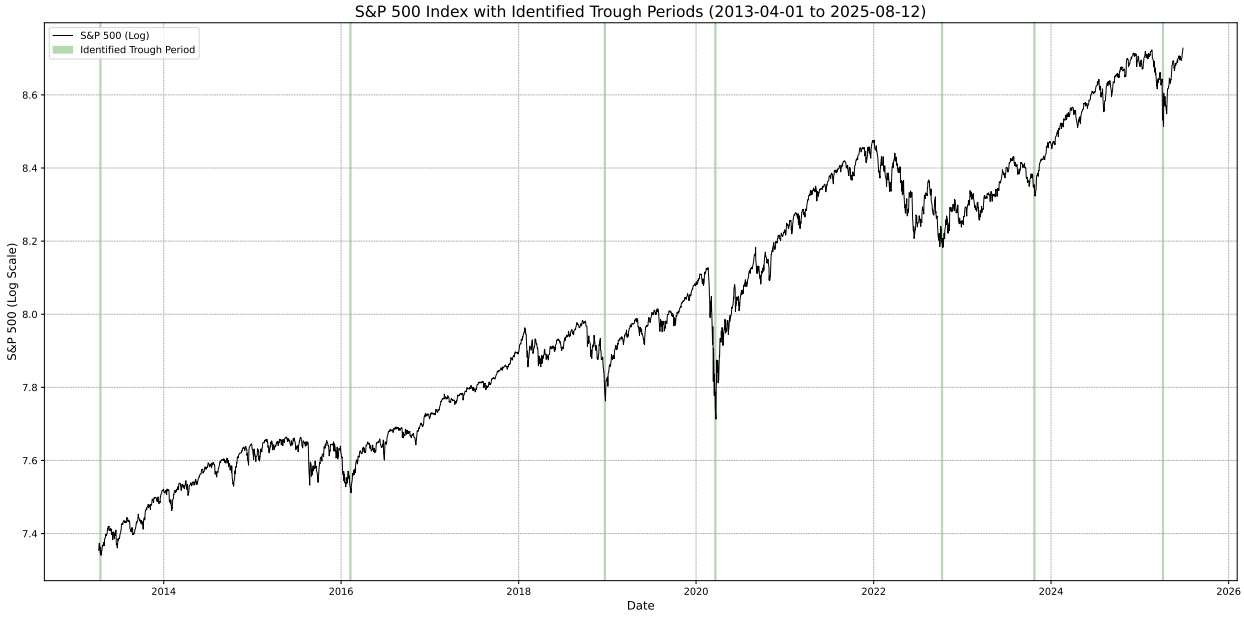


Figure 3: S&P 500 Log Price and Identified Market Troughs (April 2013 - June 2025)

Notes: This figure plots the daily log price of the S&P 500 index over our sample period. The vertical green lines indicate the dates of significant market troughs. These troughs are identified using a pure implementation of the modified Bry-Boschan algorithm, as detailed in Section 3.1. To ensure all turning points within the sample period are identified algorithmically without end-of-sample ambiguity, the algorithm is applied to a data series extending beyond June 2025. The S&P 500 price data is from Databento and Shiller's public database.

theoretical state. We map our indicators directly to the theoretical components identified in Table 1. Table 4 presents the key indicators used for this mapping, grouped by the theoretical concept they represent. We now briefly describe the rationale for each group.

3.2.1 Proxies for Customer Order Imbalance (Z_1)

The latent variable Z_1 represents the acute, one-sided selling pressure from customers that speculators must absorb. We proxy this using high-frequency measures of futures order flow. Our **Order Flow Imbalance (OFI)** measure captures the net sign-weighted volume from 1-minute bars, while **Flow Concentration** measures the persistence of this flow, distinguishing chaotic, balanced flow from the kind of sustained, unidirectional selling indicative of capitulation.

3.2.2 Proxies for Speculator Fragility ($\mathbf{x}_0, \mathbf{W}_1$)

The variables $\mathbf{x}_0$ and $\mathbf{W}_1$ represent the initial positions, wealth, and capital constraints of the speculator sector. We construct empirical proxies for this fragility. **Unrealized Profit** ($\frac{\mathbf{P}_t - \text{VWAP}_{63d}}{\text{VWAP}_{63d}}$) directly measures the paper losses of recent market participants; large negative values imply widespread financial stress. **Delta Exposure (DEX)** measures the net market delta positioning of options holders, where extremely low or negative values indicate bearish positioning and potential vulnerability.

3.2.3 Proxies for Margin Sensitivity ($\bar{\theta}$)

The parameter $\bar{\theta}$ governs the destabilizing margin spiral, where falling prices increase perceived volatility, causing financiers to raise margins. We proxy this latent sensitivity using observables that capture panic and hedging pressure. The **VIX** and **Realized Volatility (RV)** are direct measures of implied and historical panic, respectively. **Gamma Exposure (GEX)** measures dealer hedging pressure; a negative GEX state implies dealers must sell into a falling market, empirically amplifying the $\bar{\theta}$ parameter.

Table 4: Empirical Proxies for Theoretical State Variables

Theoretical Variable	Proxy Name	Mathematical Definition	Economic Intuition	Reference(s)
Proxies for Z_1	Order Flow Imbalance	$\sum_{k=1}^{N} \text{sign}(\mathbf{C}_k - \mathbf{O}_k) \cdot \text{Vol}_k$ over 1 minute	Proxy for net buying/selling pressure. Sustained negative OFI indicates aggressive selling preceding seller exhaustion.	Easley et al. (2012)
	Flow Concentration	$(\sum_{i=0}^{9} \mathbf{OFI}_{t-i}) \cdot \frac{\|\sum_{i=0}^{9} \mathbf{OFI}_{t-i}\|}{\sum_{i=0}^{9} \|\mathbf{OFI}_{t-i}\|}$	Measures persistence of order flow. High negative values suggest sustained, concentrated selling (capitulation).	
Proxies for $\mathbf{x}_0, \mathbf{W}_1$	Unrealized Profit	$\frac{\mathbf{P}_t - \text{VWAP}_{63d}}{\text{VWAP}_{63d}}$	Gauges P/L of recent participants. Large negative values mean participants are heavily underwater, increasing forced selling risk.	
	Delta Exposure	$\sum_i (\Delta_{C,i} \cdot \mathbf{OI}_{C,i} + \Delta_{P,i} \cdot \mathbf{OI}_{P,i}) \times 100$	Measures net market delta. Extremely low values indicate bearish positioning and potential for short covering near troughs.	SqueezeMetrics
Proxies for $\bar{\theta}$	GEX (OI)	$\sum_i (\Gamma_{C,i} \cdot \mathbf{OI}_{C,i} - \Gamma_{P,i} \cdot \mathbf{OI}_{P,i}) \times 100$	Measures dealer gamma exposure. Negative GEX regimes amplify volatility, empirically realizing the margin spiral.	SqueezeMetrics
	VIX	CBOE VIX Index methodology	Market's expectation of 30-day implied volatility. High VIX signals fear and demand for insurance, peaking at troughs.	Whaley (2000)
	Realized Volatility	$\sqrt{252 \cdot (\sum_{i=1}^{M-1} r_{i,intra}^2 + r_{overnight}^2)}$	Historical volatility from high-frequency data. Spikes in RV indicate panic and forced liquidation.	Andersen et al. (2003)
Proxies for $C(\cdot)$	Credit Spread	$\mathbf{Yld}_{HY} - \mathbf{Yld}_{RF}$	The premium for bearing credit risk. A widening spread signals deteriorating conditions and heightened risk aversion.	Fama and French (1989)
	Amihud Illiquidity	$\frac{\|\mathbf{R}_{daily}\|}{\mathbf{V}_{\$,\text{ daily}}}$	Measures price impact. High values indicate illiquidity (small volumes cause large price changes), which vanishes near troughs.	Amihud (2002)

Notes: This table details a selection of the key empirical indicators used to proxy the latent variables from the Brunnermeier and Pedersen (2009) bifurcation boundary, as mapped in Table 1. The indicators are grouped by the theoretical concept they are intended to capture. The full feature set $\mathbf{X}_t$ consists of transformations of these and other related parent indicators.

3.2.4 Proxies for the Stability Threshold ($C(\mathbf{\Theta}_t, \mathbf{p}_1)$)

The threshold $C(\cdot)$ represents the market's overall resilience, which is a function of broad funding conditions and market depth. We proxy this using two pillars of market stability. The **Credit Spread** (BofA US High Yield Index Yield minus a risk-free rate) measures system-wide funding stress and risk aversion. **Amihud Illiquidity** ($\frac{|\mathbf{R}|}{\mathbf{V}_\$}$) directly measures price impact, a proxy for market depth. A high and rising value for these proxies indicates a lower stability threshold, meaning the market is more fragile and closer to the bifurcation point.

To enhance the robustness of our study, we conduct a systematic treatment of outliers. We remove any value of Gamma Exposure indicators (gex_OI and gex_V) exceeding the 99.9th percentile threshold. In the case of the open interest-based Put/Call Ratio (pcr_OI), we treat any observation of exactly zero as data artifacts and remove it from the analysis.

Table 5: Data Sources and Characteristics

Source	Series Identifier	Description	Time Period	Native Freq.
Databento	SPX.EOD	SPX End-of-Day Option Chains	Apr 2013 - Jun 2025	Daily
Databento	CBOE.SPX.OPNINT	SPX Option Open Interest	Apr 2013 - Jun 2025	Daily
Databento	CME.ES.OHLCV.1M	E-mini S&P 500 Futures OHLCV	Apr 2013 - Jun 2025	1-Minute
Databento	CME.ES.BBO.1S	E-mini S&P 500 Futures BBO	Apr 2013 - Jun 2025	1-Second
CME DataMine	ZQ	30-Day Fed Funds Futures	Apr 2013 - Jun 2025	Daily
CME DataMine	6E,6J	EUR/USD, JPY/USD Futures	Apr 2013 - Jun 2025	Daily
FRED	BAMLH0A0HYM2EY	ICE BofA US High Yield Index Yield	Apr 2013 - Jun 2025	Daily
FRED	DGS1MO	1-Month Treasury Rate	Apr 2013 - Jun 2025	Daily
FRED	EFFR	Effective Federal Funds Rate	Apr 2013 - Jun 2025	Daily
Shiller Data	S&PComposite	S&P Composite Dividend Data	Apr 2013 - Jun 2025	Monthly
Yahoo Finance	^VIX	CBOE Volatility Index (VIX)	Apr 2013 - Jun 2025	Daily

Notes: This table details the raw data sources used for feature engineering in this study. The overall sample period for the analysis runs from April 2013 to June 2025. "Native Freq." refers to the highest frequency at which the data is natively available from the source before any aggregation or resampling.

3.3 Raw Data Sources

The raw financial data used to construct the empirical proxies $\mathbf{X}_t$ described in Section 3.2 are drawn from several high-frequency and daily sources, as detailed in Table 5. We gather data from several high-quality sources from April 2013 to June 2025.

3.4 Feature Transformation and Finalization

The raw proxies constructed from our data exhibit statistical properties common to financial time series, such as high persistence and extreme non-normality. Table 6 highlights these properties; for example, the kurtosis of Gamma Exposure (gex_oi) is 1790.7, and the first-order autocorrelation ($\rho(1)$) for the VIX is 0.970.

The raw proxies in Table 6 are statistically ill-suited for a linear model and, more importantly, they are *economically incomplete*. A raw VIX level, for example, conflates two theoretically distinct phenomena: a persistent, high-anxiety regime (Act 1) and an acute, rapid panic (Act 2). Our "Two-Act Play" hypothesis (Section 2.4) *requires* that we first decompose these raw signals into their distinct theoretical components.

Therefore, our feature transformation pipeline is not a statistical "black box" but a

Table 6: Descriptive Statistics for Parent Indicators (2013-2025)

Indicator	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max	$\rho(1)$
Panel A: Physical/Structural							
gex_oi	6.65e+04	2.21e+06	41.417	1790.665	-8.14e+06	1.03e+08	0.682
gex_volume	6.43e+04	1.81e+06	30.648	1001.631	-6.54e+06	6.47e+07	0.013
dex_oi	3.06e+07	7.07e+07	-0.795	5.632	-4.45e+08	4.36e+08	0.943
ofi	-4373.653	9.46e+04	-0.582	3.428	-6.70e+05	3.79e+05	0.054
credit_spread	0.049	0.016	0.258	0.024	0.014	0.114	0.998
amihud_illiquidity	9.94e-12	3.47e-11	0.000	0.000	0.000	1.17e-09	-0.049
ffr_slope	0.066	0.237	1.807	6.425	-0.615	1.203	0.995
ffr_basis	0.003	0.044	11.005	152.079	-0.128	0.715	0.900
Panel B: Psychological/Sentiment							
RV	12.700	9.840	4.105	32.356	0.758	133.842	0.669
VIX	17.812	6.942	2.730	13.973	9.140	82.690	0.970
VRP	3.336	5.042	-3.574	30.486	-58.725	16.729	0.662
PCR_OI	1.819	0.185	0.178	-0.554	1.389	2.489	0.987
PCR_V	1.390	0.318	0.356	0.335	0.533	3.092	0.729

Notes: This table reports summary statistics for the untransformed "parent" indicators at a daily frequency for the sample period April 2013 to June 2025. The final column, $\rho(1)$, reports the first-order autocorrelation coefficient. The pronounced non-normality (e.g., kurtosis of 1790 for gex_oi) and high persistence (e.g., $\rho(1) > 0.9$ for VIX and credit spreads) motivate the feature transformations detailed in Section 3.4.

theory-driven decomposition designed to construct observable proxies for our latent theoretical variables.

- **Causal Trend Decomposition (EWMA):** The "Fragile State" (Act 1) is, by definition, a low-frequency, persistent condition, while the "Dynamic Spiral" (Act 2) is a high-frequency, acute deviation. We use a causal trend decomposition based on Exponentially Weighted Moving Averages (EWMA) to isolate these components. The slow-moving EWMA (e.g., the '_ewm_slow' component) acts as a low-pass filter to capture the slow-moving buildup of systemic fragility (Act 1), while the residual from this trend (e.g., the '_ewm_resid' component) captures the high-frequency, acute dynamics of the spiral itself (Act 2).

- **Rate-of-Change (ROC) and Trend:** The "Dynamic Spiral" (Act 2) is a high-frequency phenomenon defined by acceleration. We use Rate-of-Change (ROC) and

Trend features to explicitly capture this acceleration (e.g., the change in credit spreads), providing the model with a direct proxy for the "Act 2" dynamic.

- **Rolling Rank Scaling:** Finally, we apply a rolling percentile rank transformation (252-day window) to all features. This serves a dual purpose. **Econometrically**, it is a crucial non-parametric step that transforms all proxies onto a common, stationary, unitless scale $[-1, 1]$, which is a prerequisite for a stable linear estimation. **Economically**, it aligns the features with market participants' adaptive expectations. The model learns from a feature's *relative state* (e.g., "Is VIX high relative to the past year?") rather than its absolute, non-stationary level, which is a more realistic representation of how participants assess risk.

4 Empirical Framework for Estimating the Bifurcation Boundary

4.1 Empirical Strategy: From Predictive Puzzle to Model Selection

This section details our empirical framework for building a robust nowcasting model of market troughs, $P(\mathbf{y}_t = 1|\mathcal{X}_t)$. A central, unaddressed question in crisis modeling is the nature of the boundary separating stable states from fragile ones. Are the V-shaped troughs documented in Section 2.1 complex, non-linear events that require flexible, universal approximators (like non-linear kernels or deep learning) to detect? Or do these complex, high-frequency dynamics resolve to a simpler, perhaps linear, boundary?

To answer this empirical question, we design a 'horse race' using the Support Vector Machine (SVM). The SVM is the ideal tool for this test as it allows a direct, clean comparison between a linear classifier (a linear-kernel SVM, which estimates a hyperplane $\mathbf{w}'\mathbf{x}_t + b = 0$) and a flexible, non-linear classifier (an RBF-kernel SVM, which can model boundaries of arbitrary complexity).

This empirical horse race has a sharp theoretical interpretation. As we demonstrated in Section 2, the canonical Brunnermeier and Pedersen (2009) liquidity spiral framework implies that the transition to a fragile, multi-equilibrium regime is a bifurcation event governed by a linear boundary. Therefore, this test is not just a statistical bake-off; it is a direct empirical validation of the B&P framework against its more complex, non-linear alternatives.

We argue that any "black box" critique of this approach is inverted. The feature transformations in Section 3.4 are not an opaque, non-linear model themselves. They are a *necessary, theory-driven data decomposition* required to isolate the "Fragile State" (Act 1) from the "Dynamic Spiral" (Act 2). The subsequent linear SVM is then the appropriate, parsimonious tool to estimate the simple boundary on these properly-specified theoretical components.

In this framework, the SVM's fitted weight vector $\mathbf{w}$ serves as an empirical estimate of the theoretical hyperplane's normal vector $\boldsymbol{\alpha}$, and the observable feature vector $\mathbf{x}_t$ serves as the empirical proxy for the latent theoretical state vector $\mathbf{S}_t$.

4.2 Nowcasting Objective and State Vector Definition

We formulate the task as a binary classification. As defined in Section 3.1, our objective is a nowcast of the ex-post confirmed trough label $\mathbf{y}_t = 1$ based on the current state $\mathcal{X}_t$. This framework, analogous to the real-time detection of NBER recessions, avoids data leakage as the features $\mathcal{X}_t$ are strictly pre-determined.

We define the positive class $\mathbf{y}_t = 1$ (the "fragile state") using a labeling window W_L. As illustrated in Figure 4, for any trough date T, a positive label is assigned to all time steps t in the W_L-day window immediately before the trough, such that $t \in [T - W_L, T]$. This window represents the period where the market state has definitively crossed the bifurcation boundary and is collapsing into the "bad" equilibrium—a dynamic consistent with the acute "V-shaped anatomy" documented in Section 2.1. All other days are assigned a negative label ($\mathbf{y}_t = 0$).

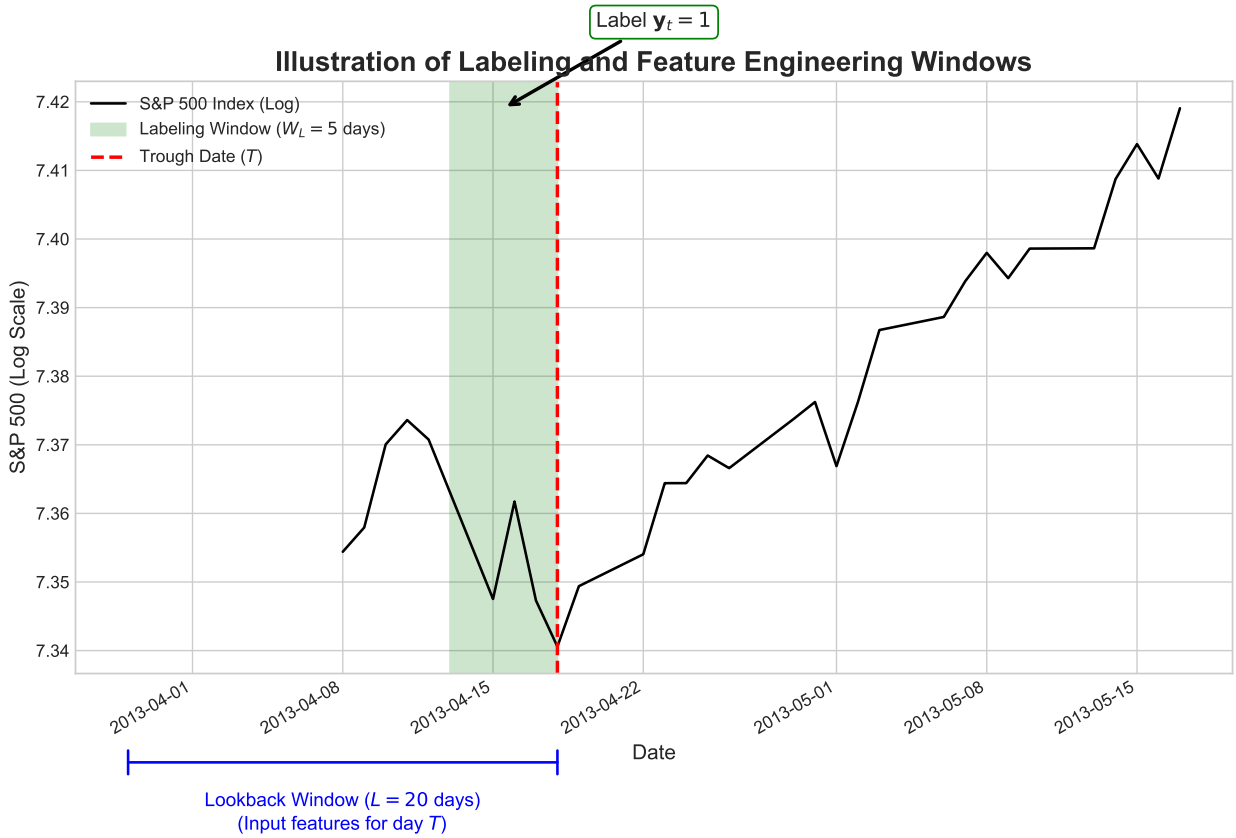


Figure 4: Illustration of the Labeling and Feature Engineering Methodology. The figure plots the daily log price of the S&P 500 Index for an illustrative period around the April 18, 2013 market trough. The trough date is denoted as T. For our classification task, a positive label ($\mathbf{y}_t = 1$) is assigned to all days within the shaded green *labeling window* ($W_L = 5$ days), defined as the period $t \in [T - W_L, T]$. The model's prediction for any given day uses input features derived from the data in the preceding blue *lookback window* ($L = 20$ days).

The theoretical state $\mathbf{S}_t$ is not static; it encompasses the dynamics and persistence of its components (e.g., the buildup of speculator losses, the acceleration of panic). To capture these path-dependent properties, our empirical proxy $\mathbf{x}_t$ is not a single snapshot. For a prediction at time t, we first assemble an input tensor $\mathcal{X}_t \in \mathbb{R}^{L \times D}$, corresponding to the D feature vectors from the last L time steps (the lookback window). This sequence tensor $\mathcal{X}_t$ is then aggregated to a final feature vector $\mathbf{x}_t \in \mathbb{R}^{4D}$ by calculating four statistics for each of the D features over the lookback window L: Mean, Standard Deviation, Trend (slope of linear regression), and Last Value. This aggregation captures both the level (Mean, Last Value) and momentum (Trend, Std. Dev.) of the market state.

This transformation is also econometrically crucial, as it converts non-stationary parent indicators into stationary features suitable for stable estimation. An Augmented Dickey-Fuller (ADF) test demonstrated that the percentage of stationary features increased from 90.6% in the parent set to 100% in the final aggregated set, increasing the stability of the hyperplane estimation.

4.3 Robust Estimation Pipeline

Estimating this boundary from financial data presents two primary econometric challenges: (1) extreme class imbalance, as trough states ($\mathbf{y}_t = 1$) are rare, and (2) high dimensionality, as our observable state vector proxy $\mathbf{x}_t$ is large. We employ a robust pipeline within a nested 'TimeSeriesSplit' cross-validation to address these challenges and avoid data leakage. The pipeline within each fold is:

1. **Data Augmentation**: To address class imbalance, we apply the Synthetic Minority Over-sampling Technique (SMOTE) only to the training set. This ensures the estimator can learn the properties of the rare "fragile" state without biasing the model evaluation.

2. **Feature Scaling**: We fit a 'StandardScaler' only on the augmented training data.

3. **Feature Selection**: To address high dimensionality and ensure the estimated hy-

perplane $\mathbf{w}$ is robust, we employ a regularized estimation procedure. This involves a pre-selection step using a Random Forest classifier, which selects the top N features from the training data based on Gini importance. This step reduces the feature space to the most relevant proxies for the theoretical drivers.

4. **Model Fitting**: We train the final SVM on the processed training data.

4.4 Core Empirical Finding: The Primacy of the Linear Boundary

We now conduct the central experiment of our framework, designed to answer the "linear vs. non-linear" puzzle posed in Section 4.1. We use the robust pipeline from Section 4.3 to tune and compare the out-of-sample performance of the linear-kernel SVM against its primary non-linear alternative, the Radial Basis Function (RBF) kernel SVM. The RBF kernel is a universal approximator capable of modeling boundaries of arbitrary complexity; if a more complex, non-linear boundary existed, the RBF model would be expected to find it and achieve superior performance.

We conduct a rigorous "horse race" using the nested time-series cross-validation pipeline described in Section 4.3. We tune the full set of hyperparameters (including the kernel, C, W_L, L, and $N_{features}$) independently for both models, selecting the optimal hyperparameters for each based on a combined score (a 50/50 blend of ROC AUC and 1-Brier score) to balance discrimination and calibration, applying the one-standard-error rule to favor parsimony.

The results of this 'horse race' provide a strong, nuanced validation of our hypothesis. During the cross-validation for our best-performing parameter set ($W_L = 5, L = 20$), the top-performing linear-kernel model (CV combined score 0.9395) and the best-performing RBF-kernel model (CV combined score 0.9395) achieved statistically indistinguishable scores. Following the one-standard-error rule to favor parsimony, our pipeline selected the simpler, more robust linear-kernel model (with $C = 0.01$, CV score 0.9371), as the complex non-linear model offered no significant performance benefit. This failure of the flexible RBF-kernel to outperform the simpler linear model is the key statistical finding. It provides economic

evidence that a more complex boundary is not necessary and that the parsimonious linear model is the most appropriate choice, aligning with our B&P-driven hypothesis. This winning linear model, when advanced to the final hold-out test set, achieved the final performance (ROC AUC of 0.9435) reported in Section 5.

Table 7 presents the search space and the final values selected for our "winning" linear model. Table 13 lists the final 15 features selected by the Random Forest step when the production model was trained on the full dataset, representing the final empirical estimate of the state vector's key components.

Table 7: Hyperparameter Tuning and Final Values

Hyperparameter	**Search Space**	**Final Value**
Labeling Window W_L (days)	{5, 10, 15, 20, 25, 30}	5
Lookback Window L (days)	{5, 10, 15, 20, 25, 30}	20
Number of Features $N_{features}$	{10, 15, 20, 25, 30}	15
Augmentation Method	{none, SMOTE, jitter, mixup}	SMOTE
Oversample Factor	{1.0, 5.0}	5.0
SVM Kernel	{linear, rbf}	**linear**
SVM C (Regularization)	{0.01, 0.1, 1.0}	0.01

Notes: This table presents the hyperparameter search space and the optimal values selected for the primary classification model. The selection is performed using a nested time-series cross-validation procedure on the training data. The final values are chosen to maximize the out-of-fold ROC AUC score, applying the one-standard-error rule. The selection of the **linear** kernel over the **rbf** kernel is a key empirical finding supporting our hypothesis.

4.5 Final Model Calibration

The raw output of the SVM, $f(\mathbf{x}_t) = \mathbf{w}'\mathbf{x}_t + b$, represents the signed distance from the estimated hyperplane. This is not a well-calibrated probability $P(\mathbf{y}_t = 1|\mathbf{x}_t)$. To transform this distance into an economically meaningful and reliable nowcast, we pass the out-of-sample scores from the cross-validation folds through an 'IsotonicRegression' model. This non-parametric step calibrates the model's output without data leakage. The upper panel

in Figure 5 demonstrates the resulting model's strong calibration on the hold-out test set, confirming its reliability as a probability-based nowcasting tool.

5 Empirical Results I: The Linear Boundary and its Economic Drivers

This section presents our two primary empirical contributions. First, we evaluate the out-of-sample performance of the linear SVM, confirming its robustness as an estimation tool. Second, we deconstruct this winning linear model to identify the specific economic state variables that drive market fragility.

5.1 Empirical Finding 1: A Linear Boundary Robustly Predicts Troughs

5.1.1 Decision on Performance Metrics

Due to the extreme class imbalance in predicting rare market troughs, standard classification metrics are ill-posed for this problem. Precision, Recall, and F1-Score are all contingent on assigning a fixed decision threshold (e.g. 0.5) to a model's probabilistic output. When the positive class (a trough) is incredibly rare, we would expect a very well-calibrated model to assign a very low probability to this event on most days. That means the predicted probability is almost never going to cross the 0.5 mark, resulting in zero positive predictions. If the number of true positives is zero, Precision, Recall, and F1 all go to zero and the default threshold is not measuring the predictive power of the model itself, but the inappropriateness of using a default threshold. Thus, our assessment is based on two metrics that are threshold insensitive and directly assess the quality of the nowcasting market capitulation warning system:

- **ROC AUC**: The area under the receiver operator characteristic curve quantifies how

well a model *ranks* the observations correctly. Specifically, if we are to randomly present both a positive instance and a negative instance, how often would the model rank the positive higher? A high AUC value indicates good discrimination.

- **Brier Score**: The Brier score measures the *accuracy and calibration* of the probability forecast itself. It is the mean-squared error of each of the predicted probabilities against the truth of what happened (0 or 1). A lower Brier score means that the model's probability output is more trustworthy and closer to the true likelihood of the event.

We choose ROC AUC to measure discrimination, and Brier score to measure probabilistic reliability, because they provided the best representation of the practical value of the model.

5.1.2 Model Performance and Benchmarks

The primary linear-kernel SVM, identified as the optimal estimator in Section 4.4, demonstrates strong out-of-sample predictability and reliability on the hold-out test set. It attained a ROC AUC of **0.9435**, showing strong discriminatory power, and a Brier score of **0.0165**, indicating reliable probabilities. This result confirms that the theoretically-grounded linear model provides a robust fit to the data.

A visualized summary of the results is shown in Figure 5. The top portion has the calibration curve, and the bottom shows practical utility of the model, showing that the green spikes in predicted trough probability act as the timely and accurate signal to actual market troughs. In non-capitulating stable market regimes, the model's green probability line remain around zero, demonstrating its ability to largely avoid false alarms.

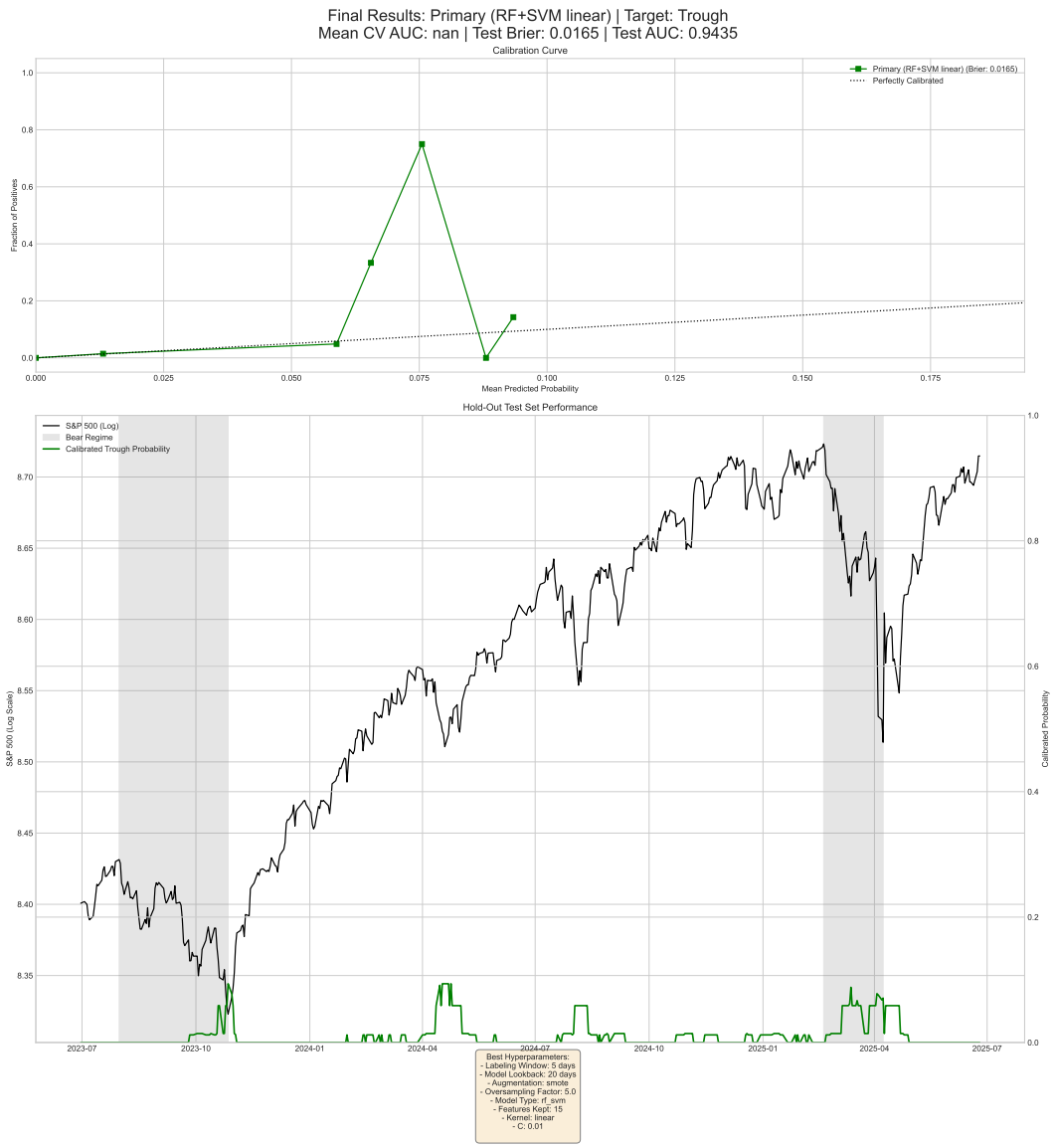


Figure 5: Out-of-Sample Predictive Performance and Trough Probabilities

Notes: This figure evaluates the primary model's performance on the hold-out test set from July 2023 to June 2025. The model is a Support Vector Machine (SVM) using 15 features selected via Random Forest Gini importance, with outputs calibrated post-hoc via Isotonic Regression (see Section 4.5 for details).

Panel A (Top): Probability Calibration. The panel plots the model's calibration curve. The dashed diagonal line represents perfect calibration. The solid line shows the model's reliability, achieving a Brier score of 0.0165. The model's overall discriminatory power, measured by the Area Under the ROC Curve (AUC), is 0.9435.

Panel B (Bottom): Predicted Trough Probability. This panel plots the model's calibrated daily probability of being in a trough state (green line, right y-axis) against the S&P 500 log price series (black line, left y-axis). The shaded vertical bars denote the actual market trough periods identified in Table 3. The hyperparameters for this model specification are listed in the embedded text box.

We compare our nowcast model with a range of benchmarks in Table 8, which provides important context for our model's performance. The benchmark results provide strong, albeit nuanced, support for our linear boundary hypothesis. The Vanilla SVM model—which uses the same linear kernel but all features without selection—achieves the highest discriminatory power (ROC AUC 0.9605). This strongly suggests the underlying problem is, at its core, linearly separable. Our Primary (RF+SVM linear) model, which adds feature selection, achieves a slightly lower AUC (0.9435) but a slightly better Brier score (0.0165), suggesting the feature selection step is valuable for calibration and parsimony.

This finding fully validates our 'linear boundary' hypothesis at the discriminatory level. The strong performance of the non-linear RF+SVM RBF (AUC 0.9141) shows that a non-linear model can also find the signal, but it fails to outperform the simpler, theoretically-grounded linear models. The Lasso model's poor discriminatory power (AUC 0.8482) suggests that simple L1 regularization (with $C = 0.01$) is less effective than the SVM's hinge loss or our RF-based pre-selection.

The other benchmarks confirm the value of this framework. The naive heuristic (VIX > 40) had low discriminatory power (AUC of 0.6656). Lastly, the Gaussian Naive Bayes model performed no better than random guessing (AUC 0.5000), indicating that the feature independence assumption is strongly violated. Thusly, our primary SVM model provided an excellent balance of high discriminatory power and trustworthy probability nowcasting.

Our primary finding—that a linear SVM effectively predicts market troughs—is robust to various alternative specifications. We test different feature subsets, alternative hyperparameter choices, and variations in the lookback and labeling windows used for model construction. Across these variations, the linear kernel consistently performs well, often outperforming non-linear alternatives. Detailed results of these methodological robustness checks, including model stability and covariate shift analysis, are provided in Appendix B.

Table 8: Out-of-Sample Performance Comparison on the Hold-Out Test Set

Model	ROC AUC	Brier Score
Primary (RF+SVM linear)	**0.9435**	**0.0165**
Benchmark Models		
Vanilla SVM	0.9605	0.0171
Lasso	0.8482	0.0173
Gaussian Naive Bayes	0.5000	0.0181
RF+SVM RBF	0.9141	0.0166
Heuristic (VIX $>$ 40)	0.6656	0.0141

Notes: This table compares the out-of-sample performance of the primary SVM model against several benchmarks on the hold-out test set (July 2023 - June 2025). Performance is measured by the Area Under the ROC Curve (ROC AUC), which assesses discriminatory power (higher is better), and the Brier Score, which measures the accuracy of probability forecasts (lower is better). The "Primary SVM (RF Select)" model is the main model from Section 4, with features selected by a Random Forest. The benchmark models are trained and evaluated under an identical time-series cross-validation framework for a fair comparison. The "Heuristic (VIX $>$ 40)" is a simple rule-based benchmark. The poor Brier score of the LassoCV model highlights its lack of probability calibration.

5.2 Empirical Finding 2: Deconstructing the Linear Boundary: What Drives Fragility?

Having established that a linear boundary provides the best fit (Section 4.4) and that it performs well out-of-sample (Section 5.1), we now deconstruct this hyperplane to answer what it represents. What are the specific economic state variables that define the boundary between stability and fragility? We use SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) to interpret the model's drivers.

This analysis is not a tautology. Our full feature set contains proxies for our B&P theory (e.g., gex_oi, upg_63d) as well as proxies for simpler, alternative theories, such as general macro conditions (e.g., ffr_slope) or standard market illiquidity (e.g., amihud_illiquidity). The SHAP analysis is performed on the 15 features selected by the Random Forest pre-selection step, forcing the interpretation to focus on the most predictively dominant state variables.

5.2.1 The Core Drivers of the Hyperplane

The SHAP analysis provides a clear verdict. The model's predictive power (ROC AUC 0.9435) is not diffuse but is highly concentrated in features that are direct proxies for our theoretical pillars. Figure 6 ranks the 15 selected features by their global importance.

Table 9 provides the "Rosetta Stone" mapping these SHAP-ranked features to their theoretical concepts. The results are decisive: of the top five most important predictors, three are direct measures of Speculator Fragility ($\mathbf{x}_0, \mathbf{W}_1$) (proxied by upg_63d... and dex_oi...). The remaining two are proxies for market Panic and Margin Sensitivity ($\bar{\theta}$) (proxied by vix_scaled_last and pcr_volume...).

This is a key economic insight: the model's estimate of the boundary is overwhelmingly defined by the pre-existing conditions of high speculator fragility and elevated market panic—the core components of the "Fragile State" (Act 1). This confirms Hypothesis 2.

Notably, the model's estimate of the boundary is not driven by the acute 'trigger' (Z_1), as the model did not select any proxies for Order Imbalance (e.g., 'ofi' or 'flow_concentration') for its top 15 features. Instead, the boundary is defined by the pre-existing conditions rather than the acute "trigger" (Act 2). This provides strong empirical support for the 'Two-Act Play' narrative, where the market becomes vulnerable long before the final push.

This distinction between pre-conditions (like speculator fragility) defining the boundary and triggers (like order flow) crossing it is the core of our 'Two-Act Play' hypothesis. The dominance of pre-condition proxies (like upg_63d...) in our predictive model is a key finding, suggesting the boundary's location is a function of the 'Fragile State' (Act 1). This sets up a crucial test for our structural analysis in Section 7: we will investigate if the 'Dynamic Spiral' (Act 2), proxied by acute order flow, has a robust partial effect that is distinct from the predictive power of the pre-conditions.

It is important to distinguish this economic interpretation (using SHAP on the final SVM) from the feature selection pipeline (using Random Forest Gini importance) that feeds the model. The stability and robustness of that Gini-based selection pipeline, which informs our

Table 9: Mapping of Top SHAP Features to Theoretical Concepts (New)

SHAP Rank	Final Feature Name (from SHAP)	Parent Proxy	Theoretical Concept	Economic Rationale (from Section 3)
1	upg_63d_ewm_resid_scaled_last	upg_63d	Speculator Fragility $(\mathbf{x}_0, \mathbf{W}_1)$	Acute deviation in unrealized losses.
2	upg_63d_ewm_spread_scaled_last	upg_63d	Speculator Fragility $(\mathbf{x}_0, \mathbf{W}_1)$	Divergence in short/long trends of unrealized losses.
3	vix_scaled_last	vix	Margin Sensitivity $(\bar{\theta})$	Level of market panic / implied volatility.
4	pcr_volume_ewm_fast_scaled_last	pcr_volume	Margin Sensitivity $(\bar{\theta})$	Fast-moving trend in put/call volume (panic hedging).
5	dex_oi_scaled_mean	dex_oi	Speculator Fragility $(\mathbf{x}_0, \mathbf{W}_1)$	Mean level of options delta exposure (positioning).

Notes: This table provides the explicit mapping for the five most important features from the new SHAP analysis (Figure 6) back to their theoretical drivers (Table 1). This mapping serves as the primary evidence for validating Hypothesis 2. The model's top features are clearly dominated by proxies for Speculator Fragility and Margin Sensitivity/Panic.

Table 10: Descriptive Statistics for Key Predictive Features

Feature	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max	$\rho(1)$
'upg_63d_ewm_resid_scaled_last'	-0.001	0.366	0.231	0.730	-1.419	1.831	0.052
'upg_63d_ewm_spread_scaled_last'	-0.001	0.354	0.046	0.540	-1.583	1.543	0.055
'vix_scaled_last'	-0.027	0.601	0.063	-1.189	-0.992	1.000	0.950
'pcr_volume_ewm_fast_scaled_last'	0.003	0.490	0.078	-0.835	-1.619	1.701	0.686
'dex_oi_scaled_mean'	0.021	0.545	-0.117	-0.920	-1.742	1.340	0.985

Notes: This table presents summary statistics for the five most important final features used in the predictive model, as determined by the global SHAP analysis shown in Figure 6. Statistics are for the full sample period (N=3068). These features are transformations of the parent indicators (Table 6) and have been scaled. The final column, $\rho(1)$, is the first-order autocorrelation coefficient.

covariate shift analysis, are validated separately in Appendix B.

5.2.2 Feature Dependence and Interaction

Having confirmed what the model looks at, we now explore how it interprets these features. The SHAP beeswarm plot (Figure 7) reveals the clear, intuitive relationships the linear model has learned for its top predictors.

- For the top driver, Speculator Fragility (upg_63d_ewm_resid_scaled_last), there is a clear negative relationship: low values (blue dots, indicating high unrealized losses)

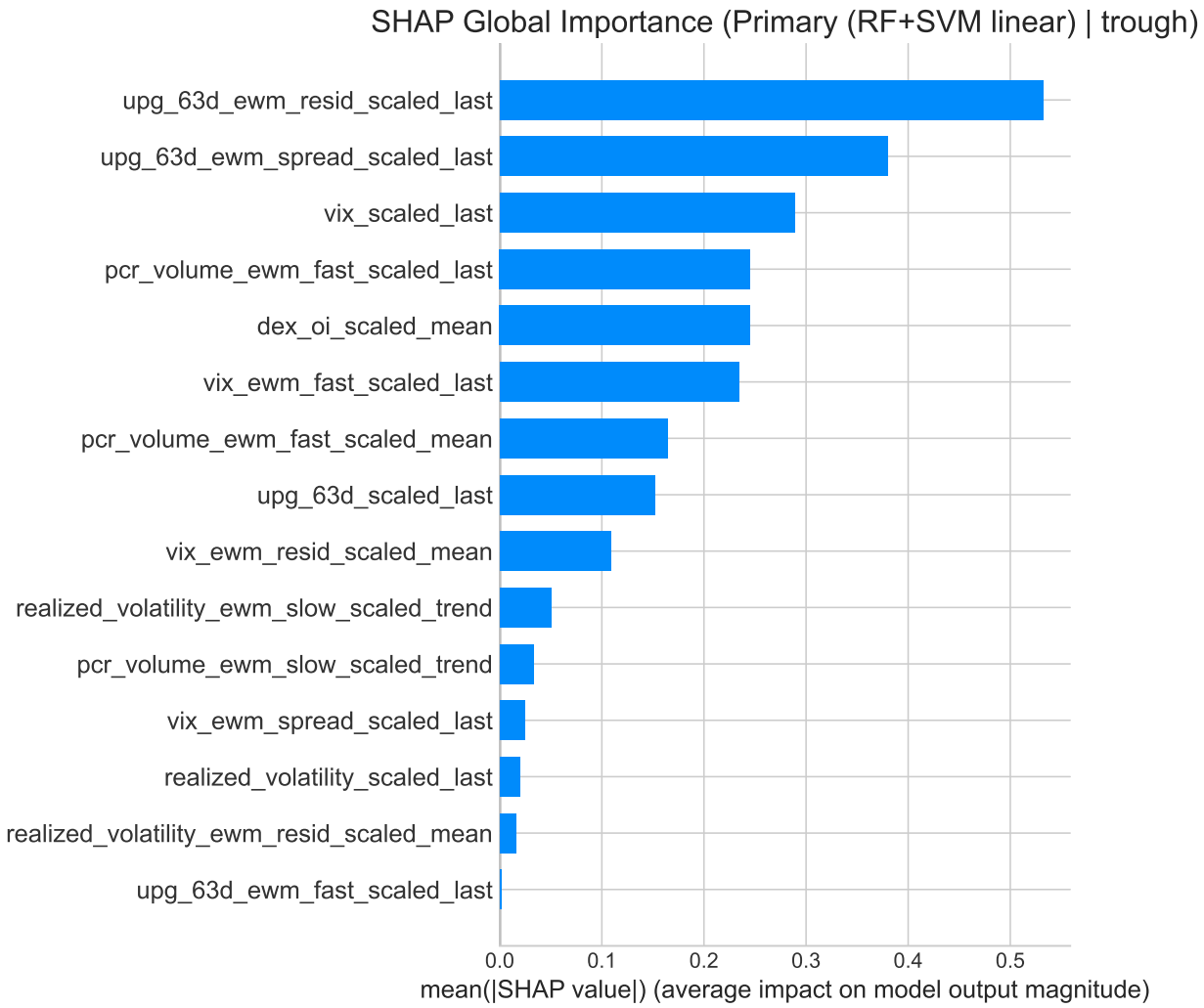


Figure 6: SHAP Global Feature Importance for the SVM Model on the Hold-Out Test Set. *Notes:* The figure displays the mean absolute SHAP (SHapley Additive exPlanations) value for the 15 features selected for our primary SVM classification model, evaluated on the hold-out test set. The x-axis represents the average magnitude of a feature's impact on the model's log-odds output. The plot clearly shows that the predictions are dominated by features related to Speculator Fragility (e.g., upg_63d...) and Panic/Margin Sensitivity (e.g., vix..., pcr_volume...), validating Hypothesis 2.

are associated with large positive SHAP values, strongly pushing the model toward a trough.

- Conversely, for Panic (vix_scaled_last), there is a clear positive relationship: high values (red dots, indicating high VIX) are associated with positive SHAP values, also pushing the model toward a trough.

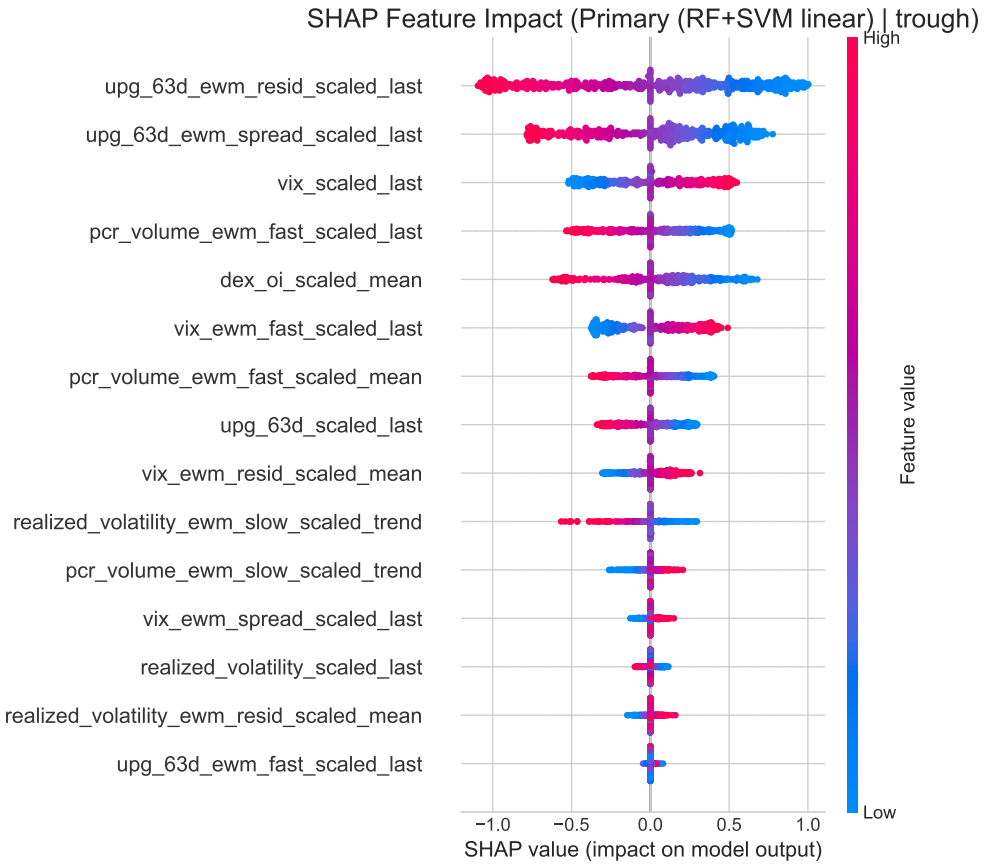


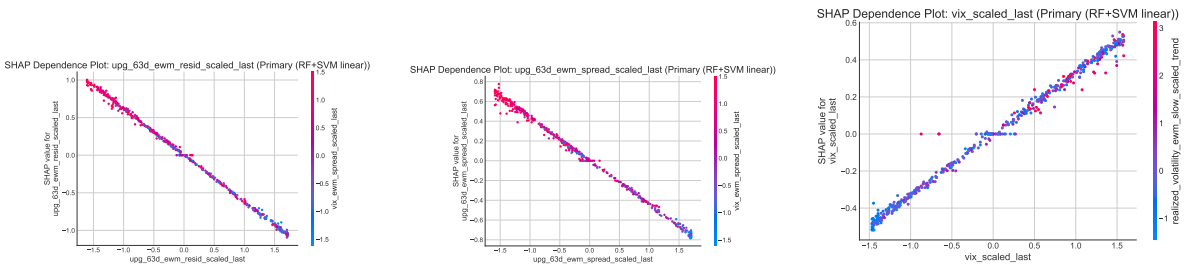
Figure 7: SHAP Feature Dependence Beeswarm Plot on the Hold-Out Test Set.
Notes: The figure illustrates both the magnitude and direction of feature impacts on the primary SVM model's predictions for the hold-out test set. Each dot is a daily observation. The horizontal position indicates its SHAP value (positive pushes toward "Trough"). The color represents the feature's value (Red=High, Blue=Low). The plot shows clear, intuitive drivers: low unrealized profits (e.g., upg_63d..., blue dots) and high VIX (e.g., vix_scaled_last, red dots) both strongly predict a trough.

The dependence plots (Figure 8) isolate these effects. For the 15 selected features of the Primary Model, the main drivers exhibit strongly linear relationships with the model's

output.

- **Panels (a) and (b): Speculator Fragility.** These plots show a powerful, negative, and almost perfectly linear relationship. As speculator fragility worsens (the feature value moves from positive to negative on the x-axis), the impact on predicting a trough (the SHAP value on the y-axis) increases linearly. The interaction effect (color) is weak, suggesting these features are strong, standalone linear predictors.

- **Panel (c): Panic / Margin Sensitivity.** This plot shows a strong, positive, and linear relationship between vix_scaled_last and its SHAP value. This confirms the intuition that higher VIX linearly increases the trough probability. The plot also reveals an interaction: the impact of VIX is dampened (dots are lower) when the trend in realized volatility (realized_volatility...trend, red dots) is already high.



(a) Speculator Fragility (Resid) (b) Speculator Fragility (Spread) (c) Panic / Margin Sensitivity (VIX)

Figure 8: SHAP Dependence Plots for Top Theoretical Drivers (Primary Model)

Notes: The figure shows the relationship between a feature's value (x-axis) and its impact on the Primary SVM model's prediction (SHAP value, y-axis). (a) Plots upg_63d_ewm_resid_scaled_las t. (b) Plots upg_63d_ewm_spread_scaled_last. (c) Plots vix_scaled_last.

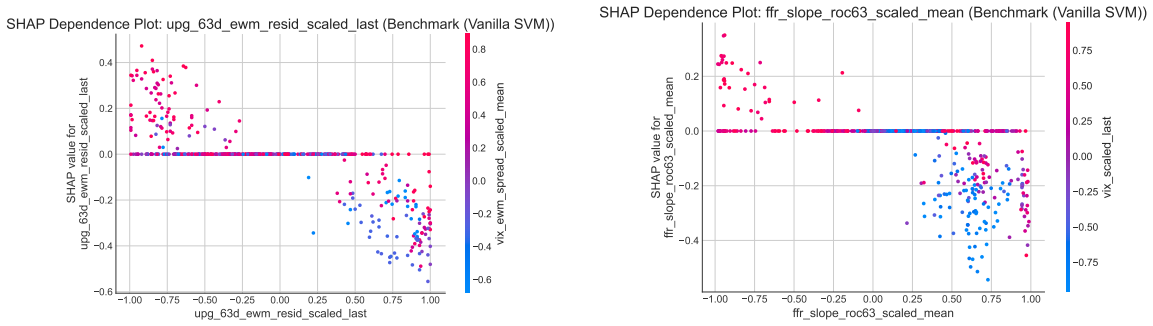
The linearity of these plots seems to challenge the non-linear B&P spiral theory. However, this linearity is an artifact of the feature selection step: the Random Forest has correctly identified the 15 features with the most powerful, direct, and linear relationships to the outcome.

To test this, we conduct the same SHAP analysis on the Benchmark (Vanilla SVM)

model, which uses all 500+ features without pre-selection. This reveals the true, underlying complexity of the feature space. As shown in Figure 9, the relationships are no longer linear.

- **Panel (a):** The impact of the top feature, upg_63d..., is highly non-linear. It has a positive impact in the low-value zone, no impact in the middle zone, and a negative impact in the high-value zone. This is a clear threshold effect.
- **Panel (b):** The impact of ffr_slope... (Fed Funds slope) is also non-linear and interactive. It pushes toward a trough at very low values (and only when VIX is high, red dots) and away from a trough at high values (especially when VIX is low, blue dots).

This discovery—that the underlying feature space is highly non-linear and interactive, even if a curated subset of features is linear—is the critical motivation for the structural analysis in Section 7. If simple correlation is non-linear and interactive (as seen in the Vanilla model), then a simple linear-causal model (like DML-PLR) is theoretically misspecified. This motivates our use of the DML-APE framework, which is designed to estimate partial effects robustly in the presence of precisely these non-linear interactions.



(a) Speculator Fragility (Non-Linear) (b) Policy Expectations (Non-Linear & Interactive)

Figure 9: SHAP Dependence Plots for Benchmark (Vanilla SVM) Model

Notes: The figure shows SHAP dependence plots for the Benchmark SVM model, which uses all 500+ features. Unlike the linear plots in Figure 8, these plots reveal the underlying non-linear and interactive nature of the full feature space. (a) shows a clear non-linear threshold effect for speculator fragility. (b) shows a non-linear and interactive effect for monetary policy expectations. This motivates the use of a non-linear DML-APE model in Section 7.

6 Economic Validation: The V-Shaped Reversal Dynamic

Sections 4 and 5 established that a linear boundary, estimated by our B&P-grounded SVM, effectively nowcasts troughs. This section provides the *third core empirical contribution* of our paper by answering the central question of theoretical choice: what is the precise nature of the reversal our signal detects?

The Brunnermeier and Pedersen (2009) framework, taken to its theoretical extreme, predicts an instantaneous V-shaped reversal as frictions (the order imbalance) are resolved in the next time step. In contrast, "slow-moving" IAP models (He and Krishnamurthy, 2013) describe a protracted, multi-month recovery. Our experiment tests where on this spectrum our empirical signal lies. Does it capture only an instantaneous, "fast-only" reversal, or does it identify the start of a durable, multi-week recovery trend?

To test this, we conduct a stylized backtest. This is not to propose a production-ready trading strategy, but to serve as a *diagnostic experiment* to characterize the *temporal dynamics* and *information decay* of the signal. The key experimental variable is the **holding period**. This design yields two sharp, competing predictions:

- **The 'Fast-Only' Hypothesis:** If the signal captures a purely instantaneous, B&P-style event, its predictive power should be highly concentrated in a very short-term window (e.g., 1-5 days) and should decay rapidly toward zero thereafter.
- **The 'Durable Bottom' Hypothesis:** If the signal identifies the start of a more durable recovery trend, its predictive power should be strong in the short-term and remain robust over longer holding periods (e.g., 20 trading days).

6.1 Experimental Design

We simulate a simple trading strategy from the model's daily, out-of-sample calibrated trough probabilities on the hold-out test set. We use the E-mini S&P 500 (ticker: ES) futures contract. The rules are as follows:

1. **Signal Generation**: A long position signal is created at the end of any day t where the calibrated trough probability exceeds the threshold of 5%, i.e., $P(\text{Trough})_t > 0.05$.

2. **Trade Entry**: The long position is entered at the closing price of the ES contract on day t. All trades are valued at \$50 per point movement.

3. **Holding Period Sensitivity**: To test the signal's information decay, each position is held for periods ranging from 5 to 20 trading days.

4. **Transaction Costs**: A round-trip commission and slippage cost of \$5.00 is deducted from the profit or loss of each contract traded.

We evaluate two strategies to diagnose the signal's properties:

- **Fixed-Size Strategy**: Serves as the baseline test, trading one contract per signal to measure the raw economic value of the signal at different time horizons.

- **Pyramiding Strategy**: Serves as a diagnostic for signal persistence. It places a position of N contracts on the N^{th} consecutive day the signal is active. This tests the hypothesis that the signals cluster at a definitive, lasting trend change versus a rapid, acute reversal.

6.2 Empirical Results: Validation of a Durable "V-Shaped" Recovery

The results of this experiment, presented in Table 11, provide strong validation for the 'Durable Bottom' hypothesis.

First, the Fixed-Size strategy confirms the signal captures the "fast" V-shape. Performance is strong in the short term, with an annualized Sharpe Ratio of **2.09** at 7 days.

Second, the results decisively reject the 'Fast-Only' hypothesis. This is the central economic finding: the signal's profitability does *not* degrade after the initial reversal. Instead,

it remains highly robust at longer horizons, with a Sharpe Ratio of **2.29** at 10 days and **1.91** at 20 days. The Total Net P&L is, in fact, highest at the 20-day hold ($392,957.50). This non-decay directly contradicts the idea of a purely instantaneous reversal and confirms the signal identifies the start of a durable, multi-week recovery trend.

Third, the Pyramiding strategy's results confirm this interpretation. This strategy, which tests the signal's persistence by leveraging into it, is *highly successful*. As shown in Table 11, it achieves a peak annualized Sharpe Ratio of **2.71** at 7 days, with a manageable Max Drawdown of **19.04%**. This success is a key feature, demonstrating that the model is not just a "local" capitulation detector for a noisy, mean-reverting bounce. It proves the model is robustly identifying the *global*, durable market bottom that begins a sustained positive trend, making it a powerful *trend-following validator*.

Table 11: Economic Significance: Holding Period Sensitivity Analysis

Holding Period	Strategy	Total Net P&L	Sharpe Ratio (Ann.)	Profit Factor	Max Drawdown	Max Drawdown (%)
5 Days	Fixed-Size	$116,145.00	1.01	1.73	($99,117.50)	40.03%
	Pyramiding	$1,287,622.50	2.18	4.04	($289,960.00)	25.26%
7 Days	Fixed-Size	$229,582.50	2.09	3.29	($77,715.00)	24.59%
	Pyramiding	$1,639,022.50	2.71	6.40	($267,727.50)	19.04%
10 Days	Fixed-Size	$295,357.50	2.29	3.33	($116,705.00)	29.95%
	Pyramiding	$1,466,260.00	1.53	2.67	($855,247.50)	48.27%
12 Days	Fixed-Size	$285,532.50	1.75	2.41	($193,787.50)	46.50%
	Pyramiding	$1,193,872.50	0.92	1.78	($1,508,412.50)	78.90%
15 Days	Fixed-Size	$328,320.00	1.82	2.44	($222,267.50)	50.06%
	Pyramiding	$1,548,897.50	1.24	2.01	($1,518,477.50)	74.09%
17 Days	Fixed-Size	$371,295.00	2.04	2.72	($216,100.00)	47.10%
	Pyramiding	$1,790,810.00	1.47	2.24	($1,439,345.00)	69.20%
20 Days	Fixed-Size	$392,957.50	1.91	2.57	($250,537.50)	55.17%
	Pyramiding	$1,863,497.50	1.44	2.22	($1,527,945.00)	78.52%

Notes: This table summarizes the performance of two stylized trading strategies on the hold-out test set (July 2023 - June 2025), evaluated across different holding periods. Both strategies trade E-mini S&P 500 futures based on the model's out-of-sample trough probability forecasts. The "Fixed-Size" strategy (termed "Pyramiding Disabled" in the raw output) trades one contract per signal. The "Pyramiding" strategy ("Pyramiding Enabled") increases position size with each consecutive signal day. The results show that performance is strong in the short-term (e.g., Sharpe 2.09 at 7 days) and remains robust for longer holds (e.g., Sharpe 1.91 at 20 days), confirming the signal captures a durable recovery. The success of the Pyramiding strategy (e.g., Sharpe 2.71, MDD 19.04%) further validates the signal's persistence.

6.3 Summary of Economic Validation

In summary, this diagnostic experiment provides the third pillar of our paper's empirical contribution. It validates our choice of the B&P framework as the correct microfoundation for our linear model by showing its signal is both fast and durable. The results deliver two new empirical facts about the signal's economic properties:

1. **The Signal is "Fast" and "Durable":** The signal's profitability is strong in the short-term (7-10 day Sharpe 2.09-2.29) and, critically, remains robust for 20-day holding periods. This demonstrates the signal captures *both* the "fast" V-shaped reversal and the start of a durable, multi-week recovery trend, refuting a "fast-only" instantaneous reversal hypothesis.

2. **The Signal is "Global," Not "Local":** The *success* of the pyramiding strategy (e.g., Sharpe 2.71, manageable drawdown) confirms the model is not just identifying local spirals but is robustly identifying the *global*, durable market bottom. This confirms the model acts as a trend-validating tool.

Having now validated the bifurcation boundary's existence (Hypothesis 1, Section 4), its economic drivers (Hypothesis 2, Section 5), and its "fast-moving" B&P dynamics (Section 6), we proceed to the final step: quantifying the **structural potency** of these drivers in pushing the market toward this fragile state, using a framework that respects the theory's non-linear interactions.

7 Empirical Results II: Testing the Non-Linear Spiral Mechanism

7.1 Resolving the Linear-Nonlinear Puzzle: From Boundary to Mechanism

In Section 5, we established that the bifurcation boundary separating stable from fragile states is (approximately) linear, validating the first part of our B&P-driven hypothesis. We now test the second, equally critical component of that theory: the non-linear, interactive spirals that push the market state towards this boundary.

This distinction is crucial and resolves the paper's central "linear vs. non-linear" puzzle.

- **The Boundary (Classification):** Estimating the boundary $\mathbf{w}'\mathbf{x}_t + b = 0$ is a classification task. A linear SVM won because the boundary itself is (partially) linear.
- **The Mechanism (Structural Estimation):** Quantifying the drivers is a structural estimation task. The B&P mechanism (the spirals) is non-linear by definition (e.g., the partial effect of an illiquidity shock D is multiplied by the level of speculator fragility X_i).

We hypothesize that a structural model that ignores these non-linear interactions (like a linear DML-PLR) will be misspecified and fail. In contrast, a model that allows for these interactions (DML-APE) will correctly quantify the structural drivers. This section, therefore, completes our validation of the B&P framework by testing its non-linear mechanism, just as Section 5 validated its linear boundary.

7.2 Baseline Model: DML of the Partially Linear Model (DML-PLR)

Our analysis starts with a baseline: the DML approach for Partially Linear Regression (PLR) models, as outlined by Chernozhukov et al. (2018). This specification provides a point of reference, with the assumption that the treatment effect is constant and additively separable. We write the structural form as:

$$\mathbf{Y} = \theta\mathbf{D} + g(\mathbf{X}) + \epsilon$$

where $\mathbf{Y}$ is the trough outcome, $\mathbf{D}$ is the treatment variable (a single indicator of interest), $\mathbf{X}$ is a high-dimensional vector of all other features that are potentially confounding, and $g(\cdot)$ is an unknown nonlinear function. The DML2 algorithm with cross-fitting provides a $\sqrt{N}$-consistent and asymptotically normal estimate for the constant treatment effect θ by flexibly modeling two nuisance functions: the outcome model $\hat{l}_0(\mathbf{X}) = \mathbb{E}[\mathbf{Y}|\mathbf{X}]$, and the treatment model $\hat{m}_0(\mathbf{X}) = \mathbb{E}[\mathbf{D}|\mathbf{X}]$.

In order to avoid the arbitrary choice of a single machine learning model for the nuisance functions, we use a data-driven selection process in each cross-fitting fold. In estimating the conditional mean of the treatment, $\hat{m}_0(\mathbf{X})$, we conduct a 'horse race'. A 'GradientBoostingRegressor' and a 'LassoCV' model are trained on the training portion of the fold. The model that exhibits better predictive performance, as assessed by out-of-sample R-squared on the validation portion of the fold, is dynamically selected for the predictions. This automatic selection improves the robustness of the DML procedure.

While the PLR specification is a standard benchmark, it has two important limitations for this setting. First, for our binary outcome $\mathbf{Y} \in \{0, 1\}$, PLR is a Linear Probability Model (LPM), which could generate predictive probabilities outside the logical [0, 1] range. Second, and more critically, it imposes the restrictive assumption that the treatment effect θ is constant and additively separable. This assumption is in direct theoretical conflict with

the Brunnermeier and Pedersen (2009) framework, which is, by definition, one of non-linear "Loss Spirals" and "Margin Spirals" where the effect of one variable depends on the state of others. The PLR model is, by design, incapable of capturing these essential interactive dynamics.

7.3 Main Model: DML for the Average Partial Effect (DML-APE)

To address the theoretical misspecification of the PLR model, we adopt the DML for Average Partial Effect (APE) as our primary specification. This more flexible DML estimator, based on an interactive model, circumvents the limitations of the PLR framework by defining the conditional probability of a trough as a non-linear interactive function:

$$P(\mathbf{Y} = 1|\mathbf{D} = d, \mathbf{X} = x) = l(d, x)$$

This specification $l(d, x)$ is not just a statistical choice; it is theoretically aligned with the B&P framework. It explicitly allows the partial effect of a treatment d (e.g., a funding shock) to depend on the state of the confounders x (e.g., the level of speculator fragility), thus capturing the interactive nature of the spirals we aim to measure. This specification is theoretically valid for a binary outcome and enables the treatment effect to vary with the state of the high dimensional confounders. The parameter that we are interested in is the Average Partial Effect (APE), θ_0, defined as the expected gradient of the conditional probability function with respect to the treatment:

$$\theta_0 = \mathbb{E}_{\mathbf{D},\mathbf{X}} \left[\frac{\partial l(\mathbf{D}, \mathbf{X})}{\partial \mathbf{D}} \right]$$

APE measures the average change in probability of a market trough for a one unit increase in treatment, averaged across the entire data distribution. In order to estimate APE reliably, this framework requires learning these three nuisance functions:

1. The outcome model (classification): $l(d, x) = \mathbb{E}[\mathbf{Y}|\mathbf{D} = d, \mathbf{X} = x]$.

2. The treatment mean model (regression): $m(x) = \mathbb{E}[\mathbf{D}|\mathbf{X} = x]$.

3. The treatment conditional variance model (regression): $v(x) = \mathbb{E}[(\mathbf{D} - m(\mathbf{X}))^2|\mathbf{X} = x]$.

Similar to the PLR approach, we run both 'GradientBoostingRegressor' and 'LassoCV' models in each cross-fitting fold as a "horse race" to determine the best performing estimator of the conditional mean $\hat{m}_0(\mathbf{X})$ and conditional variance $\hat{v}_0(\mathbf{X})$, based on out-of-sample R-squared.

Given a standard and flexible assumption that the treatment is a heteroskedastic Gaussian process conditional on confounders, the Neyman-orthogonal score function for APE is:

$$\psi(\mathbf{W}; \theta, \eta) = \underbrace{\frac{\partial l(\mathbf{D}, \mathbf{X})}{\partial \mathbf{D}} - \theta}_{\text{Naive Score}} + \underbrace{\frac{\mathbf{D} - m(\mathbf{X})}{v(\mathbf{X})} (\mathbf{Y} - l(\mathbf{D}, \mathbf{X}))}_{\text{Bias Correction}}$$

where $\mathbf{W} = (\mathbf{Y}, \mathbf{D}, \mathbf{X})$ and $\eta = (l, m, v)$. A complete derivation of this score function is provided in D. The partial derivative term $\partial l(\mathbf{D}, \mathbf{X})/\partial \mathbf{D}$ is numerically computed with a standard finite difference approach in the fitted outcome model $\hat{l}$. The bias correction term ensures the final estimate of θ is robust to first-order estimation errors in the nuisance models. To be robust against estimation "noise", especially possible outliers when $\hat{v}(\mathbf{X})$ is close to zero, our final point estimate $\hat{\theta}$ is the median of the scores computed on the out-of-sample fold. Inference proceeds with the non-parametric bootstrap of those scores.

7.4 Identification Strategy and Robustness

Any credible estimate depends on good model specification. Our analysis is not a "fishing expedition" across all 588 features, but a targeted quantification of the structural potency of the theoretical pillars derived from the B&P model in Section 2. Our treatments D are the empirical proxies for these pillars:

- **Speculator Fragility ($D \approx \mathbf{x}_0, \mathbf{W}_1$):** Proxies include 'upg_63d' (Unrealized Profit), 'dex_oi' (Delta Exposure), and 'vrp...' (Variance Risk Premium).

- **Margin Sensitivity ($D \approx \bar{\theta}$):** Proxies include 'gex_oi' (Gamma Exposure), 'vix', 'realized_volatility', and the 'risk_neutral_skewness/kurtosis' and 'pcr_oi...' families.
- **Stability Threshold ($D \approx C(\cdot)$):** Proxies include 'credit_spread' (Funding Stress) and 'amihud_illiquidity' (Market Illiquidity).
- **Order Imbalance Trigger ($D \approx Z_1$):** Proxies include 'flow_concentration' and 'ofi'.

Our DML framework provides robust estimates of the Average Partial Effect under the assumption of selection on observables—that is, conditional on our high-dimensional state vector $\mathbf{X}$, there are no unobserved confounders that jointly determine the treatment $\mathbf{D}$ and the outcome $\mathbf{Y}$. While this assumption is strong in observational time-series data, our goal is not to claim full causal identification as one would from a quasi-experiment. Instead, our objective is to 1) estimate the most robust, debiased partial effects possible given the data, and 2) demonstrate the critical importance of correct, theory-guided model specification (APE vs. PLR).

To keep comparisons between DML-PLR and DML-APE fair, the methods adopted to mitigate endogeneity are applied equally to both frameworks.

7.4.1 Bad Controls and Multicollinearity

In order to eliminate spurious results from multicollinearity or "bad controls", we have an explicit exclusion map. When an aggregated variable (e.g. vrp_scaled_mean) is selected to be the treatment variable $\mathbf{D}$; we exclude all other aggregated features with the same parent indicator (i.e. vrp_scaled_std, vrp_scaled_trend from the set of potential confounders $\mathbf{X}$. We also used an exclusion map to remove any features that would be mechanistic components of the treatment. For example, because Variance Risk Premium (VRP) is defined by VIX and Realized Volatility (RV), we eliminated all features based on VIX or RV from $\mathbf{X}$ when VRP related features are the treatment $\mathbf{D}$. This procedure is essential to estimate the total partial effect.

7.4.2 Sensitivity to Unobserved Confounders

To test the robustness of our estimates to potential violations of the selection on observables assumption (i.e., the presence of unobserved confounders), we subject all statistically significant DML estimates to a formal sensitivity analysis based on Cinelli and Hazlett (2020). It quantifies how large an unobserved confounder must be (expressed by its partial R^2 with the treatment and the outcome) to reject the estimated effect. We use this as a pragmatic guide to validate our structural claims against unobserved "worst-case" linear confounders. Only findings that are statistically significant ($p < 0.05$) and pass this sensitivity check are deemed "robust."

7.5 DML Results: The Primacy of Non-Linear Spirals

Our comparative structural analysis confirms that robust economic insight depends on correct theoretical specification. The restrictive DML-PLR model identified 71 robust partial effects, while the flexible, theory-aligned DML-APE model identified 52. The switch to the APE model causes a massive re-evaluation of the market's structural drivers: 58 findings from the linear model are revealed to be non-robust ("Lost Robustness"), while 39 new drivers, invisible to the linear model, are uncovered ("Gained Robustness").

Most critically, we identify 17 variables where the linear PLR model was not just wrong, but reported a sign that is the opposite of the robust effect found by the APE model ("Sign Flips"). This demonstrates the profound danger of using a misspecified linear-causal model. The results yield four core insights, summarized in Table 12.

7.5.1 Finding 1: Correcting Theoretically-Inverted Effects (The "Sign Flip" List)

This finding is the central empirical contribution of our DML analysis. The restrictive PLR model is not just imprecise; it is actively misleading. The "Variables with Coefficient Sign Flips" table (see Appendix C) shows 17 instances where the APE model finds a robust effect

with the opposite sign of the non-robust PLR estimate.

- **Market Panic:** The 'pcr_oi_ewm_slow_scaled_std' (Put/Call Ratio volatility) is negative in the PLR (suggesting panic-buying volatility is stabilizing) but becomes robustly positive in the APE model ($\hat{\theta} = 0.0134$). The APE model correctly identifies that instability in panic-driven put-buying is, as expected, a destabilizing driver.

- **Risk Pricing:** The 'risk_neutral_kurtosis_ewm_slow_scaled_std' (Kurtosis volatility) is negative in PLR but robustly positive in APE ($\hat{\theta} = 0.0084$). This confirms that instability in tail-risk pricing is, as B&P theory predicts, a destabilizing driver—a fact the linear model completely misses and inverts.

- **Speculator Fragility:** The 'vrp_ewm_slow_scaled_mean' (level of Variance Risk Premium) is negative in PLR but robustly positive in APE ($\hat{\theta} = 0.0380$). This suggests that a higher risk premium, a proxy for intermediary constraints and fragility, is a destabilizing driver, which only the interactive APE model correctly identifies.

7.5.2 Finding 2: Identifying Key Non-Linear Drivers (The "Gained Robustness" List)

The APE model's primary contribution is its discovery of the true non-linear structural drivers, which were invisible to the linear model. These findings align perfectly with the B&P mechanisms.

- **Order Imbalance Trigger (Z_1):** This is a major new finding. The "Gained Robustness" list is filled with proxies for the B&P "trigger" (Z_1). Variables like 'flow_concentration_10d_ewm_slow_scaled_mean' ($\hat{\theta} = 0.1015$) and ofi_ewm_slow_scaled_trend ($\hat{\theta} = 0.1937$) are non-robust in PLR but emerge as powerful, positive drivers in APE. The flexible APE model is the only one to robustly identify the acute, destabilizing partial effect of one-sided order flow.

- **Foreign Exchange Risk:** The APE model robustly identifies proxies for FX volatility (e.g., ‘fx_rv_6e_21d_scaled_mean‘ $\hat{\theta} = -0.0225$) and momentum (e.g., fx_momentum_6j_21d_roc63_scaled_std $\hat{\theta} = 0.0082$) as structural drivers, pointing to international capital flow risks that the linear model misses.

7.5.3 Finding 3: Filtering Spurious Linear Effects (The "Lost Robustness" List)

The APE model acts as a crucial filter, discarding 58 spurious linear relationships. This demonstrates that the PLR model is noisy and over-confident.

- **Credit Spreads:** The ‘credit_spread_ewm_fast_scaled_last‘ variable was robust in PLR ($\hat{\theta} = 0.0177$) but is not robust in APE. This confirms that the relationship between credit spreads and troughs is more complex and interactive than the linear model can capture. Gnbsp;
- **Market Illiquidity:** A key proxy, ‘amihud_illiquidity_ewm_slow_scaled_std‘, was robustly positive in the PLR model ($\hat{\theta} = 0.0611$) but loses robustness in the APE model. This demonstrates that the simple linear partial effect of illiquidity is not structurally robust once non-linear interactions are permitted.

7.5.4 Finding 4: Refining the "Two-Act Play" and the Stable Policy Channel

Our analysis confirms and refines the "Two-Act Play" narrative (Sec 2.4) by distinguishing predictive correlation from direct partial effects.

- **The ’Fragile State’ (Act 1):** This finding empirically confirms the distinction we hypothesized in Sections 2.4 and 5. In our predictive SHAP analysis (Section 5.2, Figure 6), speculator fragility (upg_63d) was the #1 predictor. This is because the ’Fragile State’ is a necessary pre-condition (high correlation) for a trough. However, the DML analysis here confirms its direct partial effect is not robust. It is the ’tinder,’ not the ’spark.’

Table 12: Comparative DML Partial Effect Estimates: PLR vs. APE Models

Theme	Treatment Variable (D)	Model	Coeff. ($\hat{\theta}$)	p-value	Robust?
Finding 1: Sign Reversal for Panic and Risk Pricing					
Market Panic	pcr_oi_ewm_slow_scaled_std	PLR	-0.0168	-	No
		APE	0.0134	<0.0001	Yes
Tail-Risk Pricing	risk_neutral_kurtosis_ewm_slow_scaled_std	PLR	-0.0094	-	No
		APE	0.0084	<0.0001	Yes
Speculator Fragility	vrp_ewm_slow_scaled_mean	PLR	-0.0026	-	No
		APE	0.0380	<0.0001	Yes
Finding 2: New Non-Linear Drivers (Gained Robustness)					
Order Imbalance (Z_1)	flow_concentration_10d_ewm_slow_scaled_mean	PLR	-	-	No
		APE	0.1015	<0.0001	Yes
Order Imbalance (Z_1)	ofi_ewm_slow_scaled_trend	PLR	-	-	No
		APE	0.1937	<0.0001	Yes
Finding 3: Spurious Linear Effects (Lost Robustness)					
Credit Conditions	credit_spread_ewm_fast_scaled_last	PLR	0.0177	<0.0001	Yes
		APE	-	-	No
Market Illiquidity	amihud_illiquidity_ewm_slow_scaled_std	PLR	0.0611	0.0138	Yes
		APE	-	-	No
Finding 4: Consistent Stabilizing Effect (Remained Robust)					
Monetary Policy	ffr_slope_trend_z_scaled_trend	PLR	-0.3425	<0.0001	Yes
		APE	-0.0000	<0.0001	Yes

Notes: This table compares robust partial effect estimates for selected variables from the DML-PLR (Partially Linear Regression) and DML-APE (Average Partial Effect) models, highlighting the new findings. Robustness is determined by a formal sensitivity analysis à la Cinelli and Hazlett (2020). A "Yes" indicates $p < 0.05$ and passing the sensitivity check. A "No" indicates the finding is not robust. Finding 1 explicitly shows the non-robust PLR coefficient to demonstrate the sign flip. Full results are in Appendix C.

- **The 'Dynamic Spiral' (Act 2):** Conversely, as foreshadowed in Section 5.2.1, the DML results (Finding 2) show the true, structurally potent 'spark' comes from the 'Dynamic Spirals.' The acute order flow (ofi..., flow_concentration...) and tail-risk pricing (risk_neutral_kurtosis...)—features that were absent from the top 15 predictors—are now revealed to possess the robust partial effect that ignites the crisis.

- **The Stable Policy Channel:** Finally, the "Remained Robust" list shows that proxies for monetary policy easing expectations (e.g., 'ffr_slope_trend_z_scaled_trend') are robustly negative in both PLR ($\hat{\theta} = -0.3425$) and APE ($\hat{\theta} \approx -0.0$). This provides powerful, specification-invariant evidence that easing expectations act as a structurally stabilizing force.

7.5.5 Hypotheses Discarded: the VIX Trend Example

The formal DML and sensitivity analysis allows us to formally discard plausible but non-robust hypotheses. A prime example is the trend of VIX. In the initial PLR DML procedure, vix_roc63_scaled_trend produced a statistically significant coefficient ($p = 0.0161$). However, the nuisance model R^2 is 0.78, meaning that much of its variation is already explained by other controls. The formal sensitivity analysis ("Notable Non-Robust Findings" table) shows that the finding is highly sensitive to unobserved confounding. Therefore, we cannot make a robust structural claim about the VIX trend and discard it from our primary findings. This filtering procedure is essential for credible economic interpretations.

7.6 Summary: Structural Drivers and the Importance of Non-Linearity

Our structural analysis completes the paper's narrative arc, validating its central "dual structure" thesis. We have demonstrated that respecting the Brunnermeier and Pedersen (2009) theory's non-linear, interactive mechanism is essential for correct empirical inference. The restrictive DML-PLR model, which assumes linear-additive effects, failed. It produced dozens of spurious results and, more critically, theoretically-inverted signs for key B&P mechanisms like tail-risk pricing ('risk_neutral_kurtosis...') and market panic ('pcr_oi...'). In contrast, our flexible DML-APE model, which is theoretically aligned with the spirals, corrected these errors, filtered spurious linear findings, and discovered the true, non-linear structural drivers: namely, the instability in tail-risk pricing ('risk_neutral_kurtosis...') and, critically, the acute order-flow dynamics ('ofi...', 'flow_concentration...') that are the empirical signature of the B&P spiral's "trigger" (Z_1). This provides robust, theory-grounded empirical evidence that the B&P framework—with both its linear boundary and non-linear mechanism—is the correct lens for V-shaped market troughs.

8 Conclusions

This paper addresses two central puzzles in the study of financial crises. The first is economic: what mechanism drives the sharp, V-shaped price dynamics frequently observed in market troughs, which suggests a "fast-moving" driver? The second is econometric: why do simple, linear classifiers often exhibit surprisingly strong predictive power for these apparently complex, non-linear events? We argue the econometric puzzle (linearity) is the key to resolving the economic puzzle (the 'fast-moving' V-shaped driver). This paper's core premise is that this observed linearity is not a statistical fluke but a deep economic finding. To provide a microfoundation for this puzzle, we revisit the canonical 'fast-moving' market and funding liquidity (ML/FL) framework of Brunnermeier and Pedersen (2009). We show that the model's stability conditions imply that the transition to a fragile, multi-equilibrium regime—which we model as a bifurcation event—is governed by a linear hyperplane in the model's key state variables. This theoretical implication, which we are the first to empirically test at high-frequency, provides a direct, testable microfoundation for why a simple linear classifier exhibits such strong performance.

This theory-grounded hypothesis provides a direct bridge to our empirical analysis and reframes the task from statistical "prediction" to theory-driven "estimation": the objective is to empirically estimate this theoretical bifurcation boundary. This framework immediately suggests a rigorous empirical test: if the boundary is truly linear, a linear classifier should outperform a flexible non-linear one. A linear-kernel Support Vector Machine (SVM) is not merely a simplistic choice but is the most theoretically appropriate estimator for the linear boundary we derived. Our validation of Hypothesis 1 (Section 4) confirmed this. The linear-kernel SVM achieved high out-of-sample discriminatory power (ROC AUC 0.9435) and, crucially, decisively outperformed its more complex, non-linear (RBF-kernel) alternative. Furthermore, our validation of Hypothesis 2 (Section 5) used SHAP analysis to deconstruct the estimated hyperplane. This confirmed it is not a statistical 'black box' but is driven by the specific economic levers predicted by the B&P theory: proxies for margin sensitivity ($\bar{\theta}$)

and speculator fragility ($\mathbf{x}_0$).

Having validated the boundary's existence and drivers, we then validated its mechanism to resolve the economic puzzle of the V-shaped trough. First, a diagnostic backtest (Section 6) served as an experiment on the signal's temporal dynamics. The model's predictive power was not only concentrated in a 7-10 day window (capturing the "fast-moving" V-shape) but remained robust and profitable out to 20 days. This provided the empirical signature of a signal that captures *both* the "fast" B&P reversal and the subsequent "slow" U-shaped recovery implied by Intermediary Asset Pricing models, effectively linking the two frameworks. Second, we completed our validation of the theory's *dual structure*. We moved from classifying the linear *boundary* (where the linear SVM won in Section 5) to structurally estimating the non-linear *mechanism* (the spirals themselves). This DML analysis (Section 7) revealed our most important methodological finding: respecting the theory's non-linear, interactive nature is essential for correct empirical inference. A restrictive, linear-additive DML-PLR model failed, producing dozens of spurious findings and, in many cases, reporting the opposite sign of the true effect for key drivers like tail-risk pricing ('risk_neutral_kurtosis...') and market panic ('pcr_oi...'). In contrast, the flexible, theory-aligned DML-APE model (Section 7) corrected these errors and discovered the true structural drivers: the instability in tail-risk pricing ('risk_neutral_kurtosis...') and, critically, the acute order-flow dynamics ('ofi...', 'flow_concentration...') that are the empirical signature of the B&P spiral's 'trigger' (Z_1).

This paper's contribution is threefold.

- First, we identify and resolve an **empirical puzzle**: we demonstrate that linear classifiers *outperform* non-linear classifiers in nowcasting market troughs. We provide the microfoundation for this finding by showing the canonical Brunnermeier and Pedersen (2009) model *implies* a linear bifurcation boundary. This implies that non-linear interaction terms are a first-order component of the *mechanism* (the spiral), not the *boundary*, suggesting future asset pricing models should focus on modeling these non-linearities in the structural dynamics, not necessarily in the crisis threshold itself.

- Second, we provide a novel, two-part validation of the liquidity spiral framework. We show (a) the model's signal predicts both the "fast" V-shaped reversal and the "slow" U-shaped recovery, implying IAP and B&P models are complementary, describing a single, linked phenomenon (a B&P trigger for an IAP recovery). We also show (b) its estimated hyperplane is driven by the B&P-proxies for margin sensitivity ($\bar{\theta}$) and speculator fragility ($\mathbf{x}_0$). This provides a clear directive for risk managers: monitoring should focus not just on simple levels, but on their instability (e.g., `risk_neutral_kurtosis_..._std`) and on acute *order-flow dynamics* (`ofi...`, `flow_concentration...`), which we identify as key structural drivers of the fragile state.

- Third, we show that respecting this theory's non-linear interactions is *essential* for correct empirical inference. We use DML to show that restrictive linear-structural models (PLR) fail, producing spurious findings and theoretically-inverted signs, while a flexible model (APE) aligned with the theory's interactive nature successfully quantifies the key drivers of systemic fragility. For practitioners, this serves as a direct warning that standard linear factor models for risk management are misspecified and will fail during the build-up to a crisis, as they ignore the potent, non-linear instabilities (like in *order-flow* and *tail-risk pricing*) that our model identifies as primary drivers.

Our work moves the analysis of market troughs from statistical pattern recognition to the empirical estimation of a core economic mechanism.

This research opens several avenues for future work. A natural extension is to apply this "bifurcation" framework to the symmetric but distinct challenge of predicting market peaks, which are likely driven by different mechanisms than the panic-driven trough. Methodologically, our structural analysis could be extended from Average Partial Effects (APE) to Conditional Average Partial Effects (CAPE), allowing for a richer understanding of how the partial effects of drivers (like order-flow or risk-pricing instability) changes contingent on the market state. While obtaining high-frequency options data prior to 2013 is challenging, testing

the framework's robustness on the 2008 Global Financial Crisis would be an important out-of-sample validation across different macroeconomic regimes. Finally, a promising frontier is to integrate our high-frequency, empirically-estimated boundary conditions with continuous-time structural macro-finance models, such as those governed by the Hamilton-Jacobi-Bellman (HJB) equation, to formally link short-term spirals to long-term asset pricing dynamics.

9 Acknowledgements

References

Adrian, Tobias, Erkko Etula, and Tyler Muir, "Financial Intermediaries and the Cross-Section of Asset Returns," *The Journal of Finance*, 2014, *69* (6), 2557–2596.

Amihud, Yakov, "Illiquidity and stock returns: cross-section and time-series effects," *Journal of Financial Markets*, jan 2002, *5* (1), 31–56.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, "Modeling and Forecasting Realized Volatility," *Econometrica*, mar 2003, *71* (2), 579–625.

Bluwstein, Kristina, Marilyne Buckmann, Anei Joseph, Weimin Kang, Sujit Kapadia, and Özge O'Neill, "Credit growth, the yield curve and financial crisis prediction: evidence from a machine learning approach," *Bank of England working paper*, 2020, (856).

Brunnermeier, Markus K. and Lasse Heje Pedersen, "Market Liquidity and Funding Liquidity," *The Review of Financial Studies*, jun 2009, *22* (6), 2201–2238.

Bry, Gerhard and Charlotte Boschan, *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, New York: National Bureau of Economic Research, 1971.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, feb 2018, *21* (1), C1–C68.

Cinelli, Carlos and Chad Hazlett, "Making Sense of Sensitivity: Extending Omitted Variable Bias," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, feb 2020, *82* (1), 39–67.

Easley, David, Marcos M. López de Prado, and Maureen O'Hara, "Flow Toxicity and Liquidity in a High-Frequency World," *The Review of Financial Studies*, may 2012, *25* (5), 1457–1493.

Fama, Eugene F. and Kenneth R. French, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, nov 1989, *25* (1), 23–49.

Flood, Mark D., John C. Liechty, and Thomas Piontek, "Macroeconomic Patterns in System-Wide Liquidity," *Journal of Financial Econometrics*, 2014, *12* (4), 683–719.

Gromb, Denis and Dimitri Vayanos, "Intermediary Asset Pricing and the Firing Squad," *The Review of Financial Studies*, 2018, *31* (11), 4111–4151.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, may 2020, *33* (5), 2223–2273.

Gualdi, Stanislao, Martin Tarzia, Francesco Zamponi, and Jean-Philippe Bouchaud, "V-, U-, L- or W-shaped economic recovery after Covid-19: Insights from an Agent Based Model," *PLoS ONE*, 2021, *16* (3), e0247823.

He, Zhiguo and Arvind Krishnamurthy, "Intermediary Asset Pricing," *The American Economic Review*, apr 2013, *103* (2), 732–770.

_ , Bryan Kelly, and Asaf Manela, "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, 2017, *126* (1), 1–35.

Lundberg, Scott M. and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," in I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc. 2017, pp. 4765–4774.

Whaley, Robert E., "The Investor Fear Gauge," *The Journal of Portfolio Management*, spring 2000, *26* (3), 12–17.

A Full Bry-Boschan Algorithm Implementation

This appendix provides the full, unabridged pseudocode for the modified Bry-Boschan algorithm used to identify market turning points, as referenced in Section 2.2.

Algorithm 2: Full Modified Bry-Boschan Algorithm

Require: Log price series P_t, window order, min_phase, min_cycle.

Ensure: A DataFrame turns of significant peaks and troughs.

```
procedure IdentifyTurns(P_t, order, min_phase, min_cycle)
    Initialize turns with all local peaks and troughs from P_t, sorted by date.
    turns ← EnforceAlternation(turns)
    turns ← CensorPhases(turns, min_phase)
    turns ← CensorCycles(turns, min_cycle, P_t)
    return turns
end procedure

function EnforceAlternation(turns)
    Let processed_turns be an empty list.
    for each current_turn in turns do
        if processed_turns is empty or current_turn.type ≠ last_turn.type then
            Add current_turn to processed_turns.
        else                                  ▷ If same type, keep the more extreme one.
            if (current_turn.type is 'Peak' and current_turn.value > last_turn.value) or
     (current_turn.type is 'Trough' and current_turn.value < last_turn.value) then
                Replace last element of processed_turns with current_turn.
            end if
        end if
    end for
```

```
    return processed_turns
end function

function CensorPhases(turns, min_phase)
    repeat
        has_changed ← false
        for i from 0 to length of turns - 2 do
            if Duration(turns[i], turns[i+1]) < min_phase then
                Drop the less extreme of turns[i] and turns[i+1].
                turns ← EnforceAlternation(turns)
                has_changed ← true; break                    ▷ Restart scan.
            end if
        end for
    until has_changed is false
    return turns
end function

function CensorCycles(turns, min_cycle, P_t)
    repeat
        has_changed ← false
        for i from 0 to length of turns - 3 do
            if Duration(turns[i], turns[i+2]) < min_cycle then
                Let (t1, t2, t3) be (turns[i], turns[i+1], turns[i+2]).
                if t2 is the true extremum between t1 and t3 then
                    Drop t1 and t3.
                else
                    Drop t2.
```

```
                end if
                turns ← EnforceAlternation(turns)
                has_changed ← true; break                    ▷ Restart scan.
            end if
        end for
    until has_changed is false
    return turns
end function
```

B Robustness and Stability Evaluation

One of the primary difficulties for any predictive model in finance is dealing with structural breaks: a fundamental change in the data-generating process in the market, which can disrupt relationships that are learned from historical data. Conventional econometric models (e.g., OLS, VAR) with fixed parameters are particularly susceptible to structural breaks, whereas our machine learning pipeline is more robust in being designed for flexibility. Our SVM model is non-parametric and implements adaptive feature engineering over rolling windows, further enhanced by the robust time-series cross-validation protocol, so it should be robust to evolving market environments.

Nonetheless, we conduct a series of diagnostic evaluations on the hold-out sample in order to justify the robustness of our model. These evaluations are used to identify common failure modes in machine learning models, such as degraded performance, covariate shift (when the distributions of input data change), and concept drift (when the relationship between inputs and outcome change).

B.1 Stability of Model Performance Over Time

- **Rationale**: The simplest way to evaluate model robustness is to look at how performance changes over time. A stable model should be able to maintain its predictive power, while a model that has suffered from a structural break would demonstrate abrupt and sustained decline in performance. The Brier score is a useful metric for assessing performance, as it captures the accuracy of probabilistic forecasts.

- **Performance Evaluation**: We estimate the Brier score on the hold-out test set over a **63-day rolling window** (approximately one trading quarter), rather than estimating it as a single number. This allows us to assess the model's calibration and accuracy chronologically.

- **Results**: Our findings reveal that the model is highly stable. The rolling Brier score stays extremely low (near 0.00) for the vast majority of the test period, suggesting consistently measured probabilities that are accurate and well-calibrated during a stable market. The brier score shows two elevated periods that maps to the two actual market trough events identified by the BB algorithm. These spikes should not be interpreted as the failure of the model, but merely a reflection of the inherent difficulty and uncertainty of those specific moments. Importantly, the Brier score falls back to its low baseline quickly after the spikes, indicating that model performance is not predictably worse after a crisis event. This verifies the model is not "broken" by market capitulation; it recognizes it, and then stabilizes back to near zero, as illustrated in Figure 10.

B.2 Input Feature Stability: Covariate Shift Analysis

- **Rationale**: A model trained on data with a particular distribution perform poorly when it is asked to make predictions using data with a markedly different distribution—this is called covariate shift. To test for this, we compare the distributions of the most

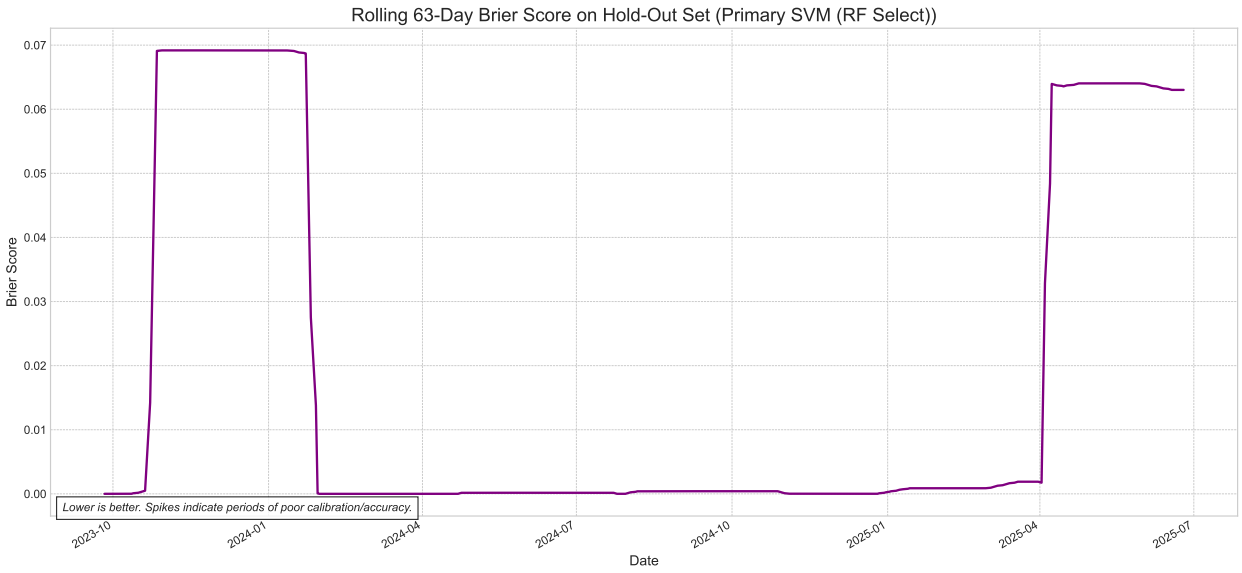


Figure 10: Model Performance Stability on the Hold-Out Test Set

Notes: This figure plots the Brier score of the primary SVM model's calibrated probability forecasts, calculated over a 63-day rolling window. The sample is the hold-out test set, covering the period from July 2023 to June 2025. The Brier score measures the mean squared error between predicted probabilities and actual outcomes; a lower score indicates better forecast accuracy and calibration. The sharp spikes in the score around October 2023 and April 2025 coincide with the actual market troughs identified in Table 3. The score's rapid return to a near-zero baseline following these events demonstrates that the model's performance is stable and does not persistently degrade after periods of market stress.

Table 13: Final 15 Features Selected by Random Forest for SVM Input

Feature Name
upg_63d_ewm_resid_scaled_last
upg_63d_ewm_spread_scaled_last
vix_scaled_last
pcr_volume_ewm_fast_scaled_last
dex_oi_scaled_mean
vix_ewm_fast_scaled_last
pcr_volume_ewm_fast_scaled_mean
upg_63d_scaled_last
vix_ewm_resid_scaled_mean
realized_volatility_ewm_slow_scaled_trend
pcr_volume_ewm_slow_scaled_trend
vix_ewm_spread_scaled_last
realized_volatility_scaled_last
realized_volatility_ewm_resid_scaled_mean
upg_63d_ewm_fast_scaled_last

Notes: This table lists the 15 features selected by the Random Forest classifier when the final production model is trained on the entire main dataset (Apr 2013 - Jun 2023). These features constitute the final empirical proxy $\mathbf{x}_t$ for the theoretical state vector. The feature selection step is performed dynamically within each CV fold for model evaluation. The list shown here represents the specific inputs for the final model that is evaluated on the hold-out test set.

important input features selected by the SVM nowcasting model between the training and testing time periods.

- **Implementation**: We plot kernel-density estimates (KDEs) for the five features with the highest SHAP importance, as identified in Section 5.2 (Figure 6). It is methodologically critical to use the SHAP-ranked features for this test. The purpose of covariate shift analysis is to verify the stability of the direct inputs to the final classifier. Since our paper's interpretation rests on SHAP analysis (which defines the final model's logic), we must check for drift in the features that the final model deems most important.

- **Results**: The results of this analysis are shown in Figure 11. The distributions for the features plotted overlap closely. The lack of significant covariate shift provides strong evidence that the statistical properties of the important predictors did not change qualitatively over the hold-out period. It therefore supports the hypothesis that the model is operating within a similar data regime, which provides further credibility of the test set performance.

B.3 Model Interpretation Stability: Concept Drift Analysis

- **Rationale**: The most subtle and important type of structural break is concept drift, when the distributional relationship between the features and the outcome has fundamentally changed. We test this by looking at the stability of the model's own interpretation of feature importance over time.

- **Implementation**: We perform a SHAP stability analysis. We split the hold-out test set chronologically into 2 halves, and assess global SHAP feature importance bar plots for the first half and for the second half. We conduct this test for both our **Primary (RF+SVM linear)** model and the **Benchmark (Vanilla SVM)** model.

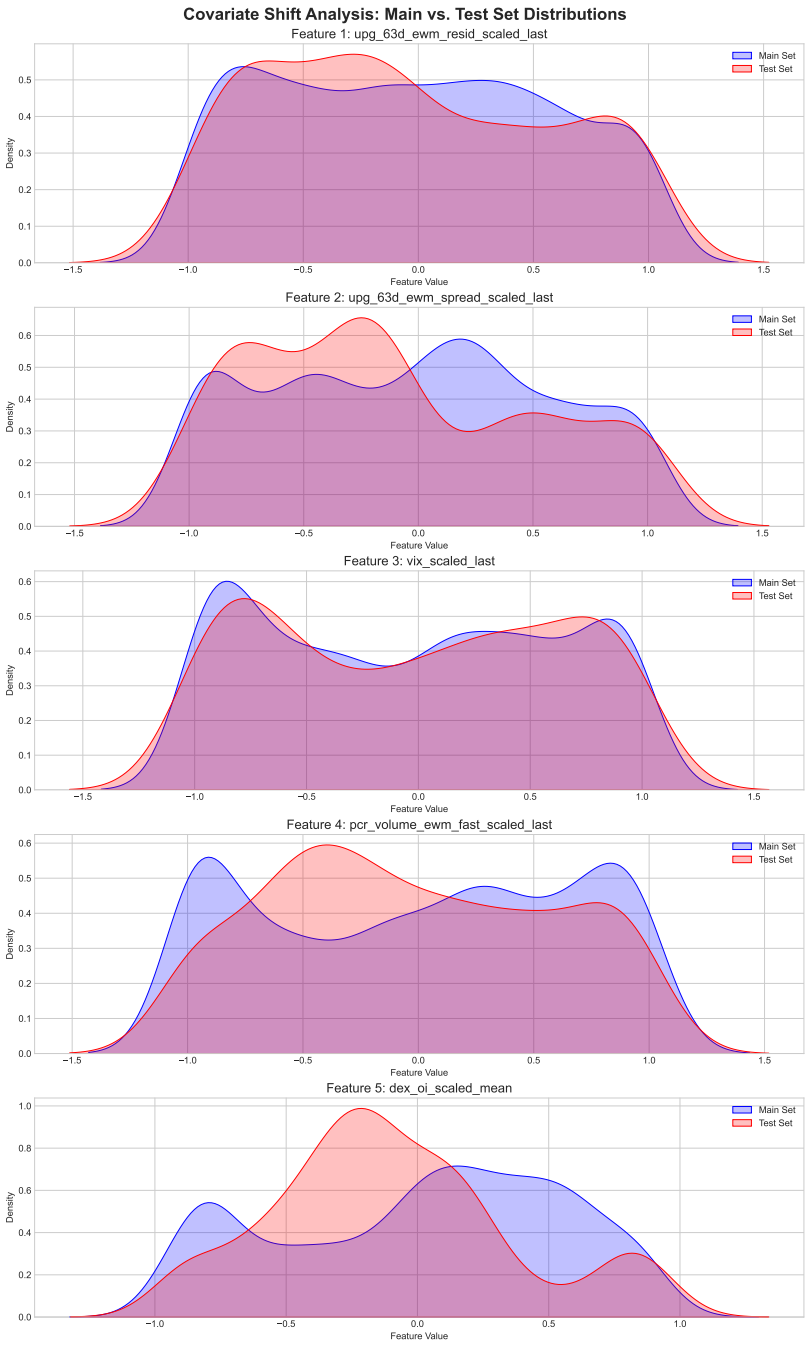


Figure 11: Covariate Shift Analysis for Top Predictive Features

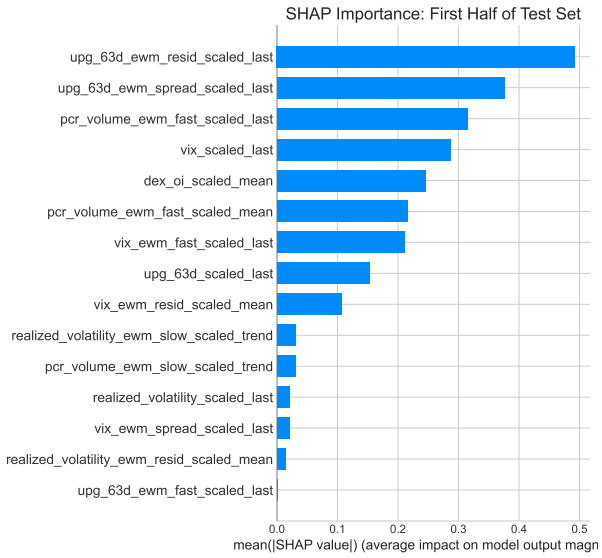
Notes: This figure visually inspects for covariate shift by comparing the distributions of key input features between the main dataset and the hold-out test set. The plots display kernel density estimates (KDEs). The "Main Set" (blue) comprises the training and validation data from April 2013 to June 2023. The "Test Set" (orange) is the hold-out sample from July 2023 to June 2025. The five features shown are the top five predictors as ranked by global SHAP importance (see Figure 6). The complete list of 15 selected features is available in Table 13. The high degree of overlap between the distributions suggests the absence of significant covariate shift.

Any considerable change in the rank or magnitude of importance would signal concept drift.

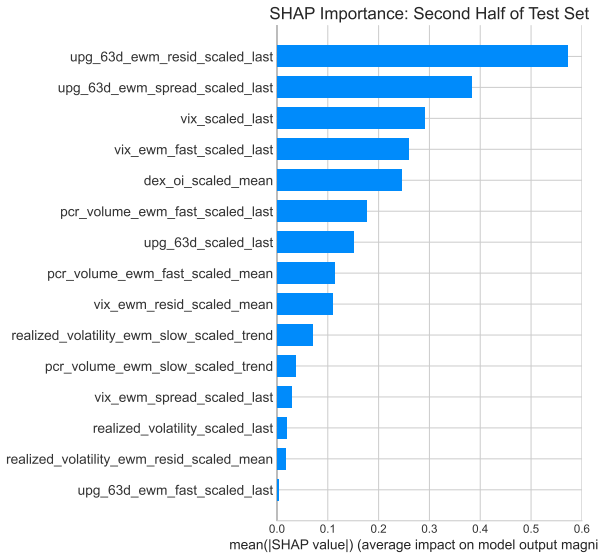
- **Results**: The findings provide strong validation for our feature-selected Primary model.
  - As seen in Figure 12, the Primary Model is highly stable. The top features in the first half (Panel (a))—such as upg_63d_ewm_resid..., upg_63d_ewm_spread.., and pcr_volume...—remain the top-ranked, dominant features in the second half (Panel (b)). This consistency gives us confidence that the model has learned durable, structural relationships.
  - In contrast, Figure 13 shows the **Benchmark (Vanilla SVM)** model exhibits moderate instability. While the top drivers (upg_63d..., ffr_slope...) remain consistent, other features show significant rank changes. For example, gex_oi_trend_z... and risk_neutral_kurtosis... are top-7 drivers in the first half but drop significantly in the second, replaced by vix_ewm_resid... and realized_volatility.... This instability reinforces the value of the feature selection step used in our Primary model, which successfully isolates the most stable predictive signals.

C Full DML Estimation and Sensitivity Analysis Results

This appendix contains the complete set of treatment variables for which the Double/Debiased Machine Learning analysis yielded a statistically significant causal estimate ($p < 0.05$) that is also robust to the formal sensitivity analysis of Cinelli and Hazlett (2020). Table 14 lists the 71 robust findings from the baseline DML-PLR model. Table 15 lists the 52 robust findings from our primary DML-APE model.



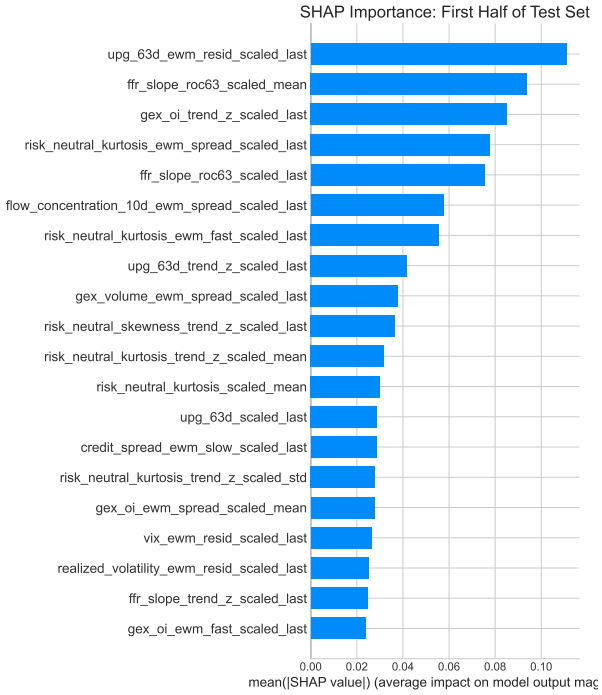
(a) Primary Model: First Half of Test Set



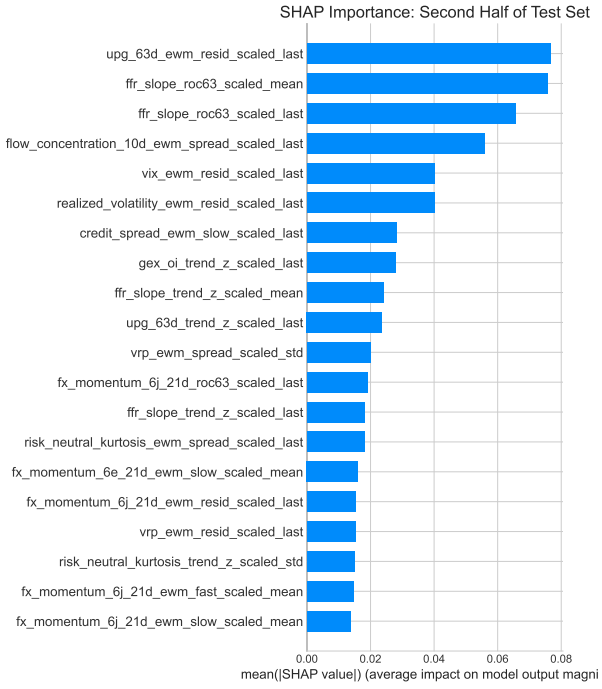
(b) Primary Model: Second Half of Test Set

Figure 12: SHAP Stability Analysis (Primary Model)

Notes: This figure compares the global feature importance rankings for the Primary (RF+SVM linear) model across the first and second halves of the test set. The rankings are highly consistent, demonstrating the model's logic is stable over time.



(a) Benchmark Model: First Half of Test Set



(b) Benchmark Model: Second Half of Test Set

Figure 13: SHAP Stability Analysis (Benchmark Model)

Notes: This figure compares the global feature importance rankings for the Benchmark (Vanilla SVM) model. The model shows moderate instability, with features like gex_oi... and risk_neutral_kurtosis... dropping in importance in the second half. This highlights the robustness gained from the Primary model's feature selection.

Table 14: Complete Robust Causal Estimates from the DML-PLR Model

Treatment Variable	Coeff. ($\hat{\theta}$)	p-value	bias_phi	Adj. 95% CI Lower	Adj. 95% CI Upper	Benchmark R_Y^2	Benchmark R_D^2
ffr_slope_trend_z_scaled_trend	-0.3425	0.0000	0.0000	-0.4740	-0.2111	0.0655	0.0000
credit_spread_ewm_fast_scaled_last	0.0177	0.0000	0.0091	0.0010	0.0343	0.0473	0.0734
fx_rv_6j_21d_trend_z_scaled_mean	-0.0261	0.0000	0.0000	-0.0375	-0.0147	0.0728	0.0000
ffr_basis_scaled_trend	0.1465	0.0000	0.0000	0.0808	0.2121	0.0655	0.0000
fx_rv_6e_21d_scaled_std	0.0534	0.0000	0.0000	0.0278	0.0790	0.0746	0.0000
ffr_slope_ewm_resid_scaled_trend	-0.2867	0.0000	0.0000	-0.4241	-0.1493	0.0655	0.0000
ffr_slope_scaled_trend	-0.3799	0.0001	0.0000	-0.5645	-0.1953	0.0655	0.0000
ffr_slope_roc63_scaled_last	-0.0172	0.0001	0.0000	-0.0258	-0.0086	0.0655	0.0000
ffr_slope_ewm_spread_scaled_trend	-0.1913	0.0002	0.0000	-0.2931	-0.0895	0.0655	0.0000
ffr_basis_trend_z_scaled_trend	0.0903	0.0002	0.0000	0.0428	0.1378	0.0655	0.0000
ffr_slope_ewm_resid_scaled_last	-0.0099	0.0003	0.0000	-0.0153	-0.0045	0.0655	0.0000
gex_volume_ewm_spread_scaled_trend	0.1660	0.0003	0.0000	0.0751	0.2569	0.0361	0.0000
risk_neutral_kurtosis_ewm_resid_scaled_std	0.0505	0.0004	0.0000	0.0225	0.0786	0.1493	0.0000
fx_rv_6j_21d_ewm_spread_scaled_trend	-0.1315	0.0005	0.0000	-0.2050	-0.0580	0.0728	0.0000
ffr_slope_trend_z_scaled_last	-0.0185	0.0005	0.0000	-0.0290	-0.0081	0.0655	0.0000
risk_neutral_kurtosis_ewm_spread_scaled_std	0.0648	0.0010	0.0000	0.0263	0.1033	0.1493	0.0000
risk_neutral_kurtosis_ewm_slow_scaled_mean	0.0098	0.0014	0.0033	0.0005	0.0192	0.1493	0.0031
flow_concentration_10d_ewm_fast_scaled_mean	0.0146	0.0016	0.0000	0.0055	0.0237	0.0721	0.0000
gex_volume_trend_z_scaled_std	0.0317	0.0018	0.0000	0.0118	0.0516	0.0361	0.0000
flow_concentration_10d_ewm_fast_scaled_std	0.0557	0.0023	0.0181	0.0017	0.1096	0.0721	0.0108
ffr_basis_trend_z_scaled_std	-0.0205	0.0026	0.0000	-0.0338	-0.0071	0.0655	0.0000
risk_neutral_skewness_ewm_fast_scaled_std	0.0337	0.0034	0.0000	0.0111	0.0563	0.1493	0.0000
flow_concentration_10d_ewm_slow_scaled_mean	0.0150	0.0035	0.0000	0.0049	0.0250	0.0721	0.0000
gex_volume_ewm_fast_scaled_trend	0.3701	0.0041	0.0000	0.1173	0.6228	0.0361	0.0000
ofi_ewm_fast_scaled_mean	0.0127	0.0044	0.0000	0.0040	0.0215	0.0721	0.0000
gex_oi_ewm_spread_scaled_std	-0.0337	0.0054	0.0000	-0.0574	-0.0099	0.0361	0.0000
risk_neutral_skewness_ewm_spread_scaled_std	0.0901	0.0054	0.0000	0.0266	0.1535	0.1493	0.0000
risk_neutral_kurtosis_trend_z_scaled_mean	0.0174	0.0061	0.0000	0.0050	0.0298	0.1493	0.0000
ofi_ewm_resid_scaled_mean	-0.0384	0.0064	0.0032	-0.0692	-0.0076	0.0721	0.0002
fx_momentum_6e_21d_scaled_mean	0.0100	0.0068	0.0000	0.0028	0.0172	0.0746	0.0000
risk_neutral_skewness_ewm_spread_scaled_trend	0.2599	0.0070	0.0000	0.0710	0.4489	0.1493	0.0000
ffr_basis_ewm_fast_scaled_last	-0.0099	0.0080	0.0000	-0.0172	-0.0026	0.0655	0.0000
ofi_ewm_slow_scaled_mean	0.0120	0.0094	0.0000	0.0030	0.0211	0.0721	0.0000
ffr_basis_scaled_last	-0.0097	0.0099	0.0000	-0.0171	-0.0023	0.0655	0.0000
risk_neutral_kurtosis_trend_z_scaled_std	0.0258	0.0106	0.0000	0.0060	0.0456	0.1493	0.0000
fx_momentum_6e_21d_trend_z_scaled_mean	0.0089	0.0109	0.0000	0.0020	0.0158	0.0746	0.0000
ffr_basis_ewm_slow_scaled_mean	-0.0086	0.0111	0.0000	-0.0152	-0.0020	0.0655	0.0000
credit_spread_trend_z_scaled_std	-0.0285	0.0121	0.0000	-0.0507	-0.0062	0.0473	0.0000
ffr_basis_trend_z_scaled_mean	-0.0117	0.0126	0.0000	-0.0210	-0.0025	0.0655	0.0000
amihud_illiquidity_ewm_slow_scaled_std	0.0611	0.0138	0.0000	0.0124	0.1097	0.0469	0.0000
ffr_basis_ewm_slow_scaled_last	-0.0084	0.0143	0.0000	-0.0150	-0.0017	0.0655	0.0000
fx_momentum_6j_21d_ewm_resid_scaled_trend	0.0907	0.0143	0.0000	0.0181	0.1633	0.0728	0.0000
gex_volume_ewm_spread_scaled_std	0.0187	0.0149	0.0000	0.0036	0.0337	0.0361	0.0000
ffr_basis_scaled_mean	-0.0115	0.0157	0.0000	-0.0208	-0.0022	0.0655	0.0000
fx_rv_6j_21d_scaled_std	-0.0356	0.0159	0.0000	-0.0646	-0.0067	0.0728	0.0000
pcr_volume_trend_z_scaled_std	0.0610	0.0165	0.0000	0.0111	0.1108	0.1234	0.0000
risk_neutral_skewness_roc63_scaled_std	0.0689	0.0166	0.0000	0.0125	0.1254	0.1493	0.0000
risk_neutral_kurtosis_scaled_mean	0.0186	0.0173	0.0000	0.0033	0.0338	0.1493	0.0000
gex_volume_ewm_resid_scaled_std	0.0210	0.0176	0.0000	0.0037	0.0384	0.0361	0.0000
pcr_volume_scaled_std	0.0584	0.0187	0.0000	0.0097	0.1072	0.1234	0.0000
flow_concentration_10d_scaled_mean	0.0138	0.0197	0.0000	0.0022	0.0254	0.0721	0.0000
fx_momentum_6j_21d_ewm_slow_scaled_last	-0.0074	0.0247	0.0000	-0.0139	-0.0009	0.0728	0.0000
ffr_slope_roc63_scaled_mean	-0.0093	0.0262	0.0000	-0.0175	-0.0011	0.0655	0.0000
flow_concentration_10d_roc63_scaled_std	0.0440	0.0274	0.0000	0.0049	0.0831	0.0721	0.0000
fx_momentum_6e_21d_ewm_slow_scaled_mean	-0.0086	0.0289	0.0000	-0.0164	-0.0009	0.0746	0.0000
ffr_basis_ewm_fast_scaled_mean	-0.0089	0.0294	0.0000	-0.0169	-0.0009	0.0655	0.0000
fx_rv_6j_21d_ewm_fast_scaled_std	-0.0368	0.0294	0.0000	-0.0700	-0.0037	0.0728	0.0000
fx_rv_6j_21d_ewm_fast_scaled_last	-0.0077	0.0342	0.0000	-0.0149	-0.0006	0.0728	0.0000
fx_rv_6e_21d_ewm_spread_scaled_std	0.0427	0.0354	0.0000	0.0029	0.0825	0.0746	0.0000
fx_rv_6e_21d_trend_z_scaled_std	0.0274	0.0356	0.0000	0.0018	0.0529	0.0746	0.0000
ffr_basis_trend_z_scaled_last	-0.0087	0.0377	0.0000	-0.0168	-0.0005	0.0655	0.0000
fx_rv_6j_21d_scaled_mean	-0.0107	0.0383	0.0000	-0.0208	-0.0006	0.0728	0.0000
fx_momentum_6e_21d_ewm_spread_scaled_trend	0.1075	0.0404	0.0000	0.0047	0.2102	0.0746	0.0000
risk_neutral_kurtosis_roc63_scaled_std	0.0397	0.0406	0.0000	0.0017	0.0778	0.1493	0.0000
vrp_trend_z_scaled_std	-0.0450	0.0440	0.0000	-0.0889	-0.0012	0.0354	0.0000
gex_volume_ewm_slow_scaled_trend	0.3513	0.0451	0.0000	0.0077	0.6949	0.0361	0.0000
gex_volume_trend_z_scaled_mean	0.0152	0.0455	0.0000	0.0003	0.0301	0.0361	0.0000
risk_neutral_kurtosis_trend_z_scaled_last	0.0077	0.0457	0.0000	0.0001	0.0153	0.1493	0.0000
ffr_slope_scaled_last	-0.0072	0.0490	0.0000	-0.0143	0.0000	0.0655	0.0000
risk_neutral_kurtosis_scaled_last	0.0069	0.0491	0.0000	0.0000	0.0138	0.1493	0.0000
risk_neutral_kurtosis_ewm_fast_scaled_mean	0.0105	0.0491	0.0000	0.0000	0.0210	0.1493	0.0000

Notes: This table reports the complete set of 71 statistically significant ($p < 0.05$) causal estimates from the Double/Debiased Machine Learning Partially Linear Regression (DML-PLR) model that are robust to unobserved confounding. The analysis uses daily data from April 2013 to June 2025 ($N = 3068$). The dependent variable is a binary indicator for a market trough. For each treatment variable listed, the model includes all other engineered features as high-dimensional controls, subject to the exclusion protocol described in Section 7.3.1. Coefficients ($\hat{\theta}$) represent the estimated linear causal effect of a one-unit change in the treatment variable on the probability of a market trough. Adjusted 95% confidence intervals and p-values are based on robust standard errors from the DML procedure. The 'bias_phi' column reports a measure of the estimation bias from the nuisance functions.

Robustness is assessed using the formal sensitivity analysis of Cinelli and Hazlett (2020). Bench-

D Deriving the Neyman Orthogonal Score for the APE

In this appendix we provide a formal derivation of the Neyman-orthogonal score function for the Average Partial Effect (APE) in a general non-parametric interactive model. We state the general form of the score, then provide a rigorous proof of its Neyman-orthogonality property with respect to both the outcome and treatment models, clarifying all necessary assumptions. Finally, we derive the practical and implementable score function that results from a semi-parametric Gaussian assumption for the treatment model, providing a complete theoretical basis for its use in debiased machine learning.

D.1 Model Setup, Assumptions, and Parameter of Interest

We consider a structural model where an outcome variable $\mathbf{Y}$ is determined by a continuous scalar treatment $\mathbf{D}$ and a vector of confounding variables $\mathbf{X}$ through a general, non-separable function g_0:

$$\mathbf{Y} = g_0(\mathbf{D}, \mathbf{X}) + \mathbf{U}$$

The core causal assumption is unconfoundedness, which posits that the treatment assignment is independent of the potential outcomes, conditional on the covariates $\mathbf{X}$. This implies that $\mathbf{U}$ is mean-independent of $\mathbf{D}$ conditional on $\mathbf{X}$, which gives the key moment condition $\mathbb{E}[\mathbf{U}|\mathbf{D}, \mathbf{X}] = 0$. This justifies defining the conditional expectation function as $l_0(\mathbf{D}, \mathbf{X}) = \mathbb{E}[\mathbf{Y}|\mathbf{D}, \mathbf{X}]$. Let $p_0(\mathbf{D}|\mathbf{X})$ denote the true conditional probability density function (PDF) of the treatment $\mathbf{D}$ given the confounders $\mathbf{X}$.

For the derivation that follows, we make the following assumptions:

- **Regularity Conditions**:
  - The conditional expectation function $l_0(d, \mathbf{X})$ is differentiable with respect to its first argument d for all $(\mathbf{D}, \mathbf{X})$ in the support of the data distribution.
  - All expectations presented in this document exist and are finite.

- For any valid perturbation function $r_l(d, \mathbf{X})$ in the Gateaux derivative path, the boundary condition $\lim_{d\to\pm\infty} r_l(d, \mathbf{X})p_0(d|\mathbf{X}) = 0$ holds. This ensures the validity of the integration by parts used in the proof of orthogonality.
- The perturbed density paths are sufficiently smooth to allow for the interchange of partial derivatives (Clairaut's Theorem).

The causal parameter of interest is the **Average Partial Effect (APE)**, denoted θ_0, which is the expected partial derivative of the outcome model with respect to the treatment:

$$\theta_0 = \mathbb{E}\left[\frac{\partial l_0(\mathbf{D}, \mathbf{X})}{\partial \mathbf{D}}\right]$$

The expectation is taken over the joint distribution of $(\mathbf{D}, \mathbf{X})$.

D.2 The General Neyman-Orthogonal Score

To achieve a $\sqrt{n}$-consistent and asymptotically normal estimator for θ_0, we use a debiased estimation approach centered on a Neyman-orthogonal score function. This score corrects for the bias introduced by regularization in high-dimensional or non-parametric estimation of the nuisance functions.

Proposition 1. *The Neyman-orthogonal score function for the APE parameter θ_0 is given by:*

$$\psi(\mathbf{W}; \theta, \eta) = \frac{\partial l(\mathbf{D}, \mathbf{X})}{\partial \mathbf{D}} - \theta - \frac{1}{p(\mathbf{D}|\mathbf{X})}\frac{\partial p(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}}\left(\mathbf{Y} - l(\mathbf{D}, \mathbf{X})\right)$$

where $\mathbf{W} = (\mathbf{Y}, \mathbf{D}, \mathbf{X})$ is the observable data and $\eta = (l, p)$ is the set of nuisance functions.

Proof of Neyman-Orthogonality. A score function is Neyman-orthogonal if its Gateaux derivative with respect to the nuisance functions, evaluated at the true parameters, is zero. Let $\eta_0 = (l_0, p_0)$ and let θ_0 be the true parameter values. We must show that for valid directional

perturbations $[r_l]$ and $[r_p]$, the following conditions hold:

$$\frac{\partial}{\partial t}\mathbb{E}\left[\psi(\mathbf{W};\theta_0, l_0 + t[r_l], p_0)\right]\Big|_{t=0} = 0$$

$$\frac{\partial}{\partial t}\mathbb{E}\left[\psi(\mathbf{W};\theta_0, l_0, p_0 + t[r_p])\right]\Big|_{t=0} = 0$$

- **1. Orthogonality with respect to the outcome model** l: Differentiating the expected score with respect to the path parameter t and evaluating at $t = 0$ yields:

$$\mathbb{E}\left[\frac{\partial r_l(\mathbf{D},\mathbf{X})}{\partial \mathbf{D}} - \frac{1}{p_0(\mathbf{D}|\mathbf{X})}\frac{\partial p_0(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}}\left(-r_l(\mathbf{D},\mathbf{X})\right)\right] = \mathbb{E}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right] + \mathbb{E}\left[\frac{p'_{0,\mathbf{D}}}{p_0} r_l\right]$$

  We analyze the second term by conditioning on $\mathbf{X}$ and using integration by parts over the domain of $\mathbf{D}$. Let d be a realization of $\mathbf{D}$.

$$\mathbb{E}_{\mathbf{D}|\mathbf{X}}\left[\frac{p'_{0,\mathbf{D}}(d|\mathbf{X})}{p_0(d|\mathbf{X})} r_l(d,\mathbf{X})\right] = \int \frac{p'_{0,\mathbf{D}}(d|\mathbf{X})}{p_0(d|\mathbf{X})} r_l(d,\mathbf{X}) p_0(d|\mathbf{X}) dd = \int p'_{0,\mathbf{D}}(d|\mathbf{X}) r_l(d,\mathbf{X}) dd$$

  Applying integration by parts, $\int u\, dv = uv - \int v\, du$, with $u = r_l(d,\mathbf{X})$ and $dv = p'_{0,\mathbf{D}}(d|\mathbf{X})dd$, gives:

$$[r_l(d,\mathbf{X}) p_0(d|\mathbf{X})]_{-\infty}^{\infty} - \int p_0(d|\mathbf{X}) \frac{\partial r_l(d,\mathbf{X})}{\partial d} dd$$

  Per our regularity conditions, the boundary term is zero. The expression simplifies to:

$$-\int p_0(d|\mathbf{X}) \frac{\partial r_l}{\partial d} dd = -\mathbb{E}_{\mathbf{D}|\mathbf{X}}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right]$$

  Substituting this back into the full expectation, the Gateaux derivative is:

$$\mathbb{E}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right] - \mathbb{E}\left[\mathbb{E}_{\mathbf{D}|\mathbf{X}}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right]\right] = \mathbb{E}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right] - \mathbb{E}\left[\frac{\partial r_l}{\partial \mathbf{D}}\right] = 0$$

  The score is therefore orthogonal to perturbations in l.

- **2. Orthogonality with respect to the density model** p: The Gateaux derivative must be computed by taking the derivative with respect to t inside the expectation, as the expectation is over the true, fixed data distribution. Let $p_t = p_0 + t[r_p]$.

$$\mathcal{D}_p = \frac{\partial}{\partial t}\mathbb{E}_{\mathbf{W}}\left[\psi(\mathbf{W}; \theta_0, l_0, p_t)\right]\Big|_{t=0}$$

The only term in ψ that depends on the path p_t is the correction term.

$$\mathcal{D}_p = \mathbb{E}_{\mathbf{W}}\left[\frac{\partial}{\partial t}\left(-\frac{\partial \log p_t(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}}\left(\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})\right)\right)\Big|_{t=0}\right]$$

Assuming the perturbed path is sufficiently smooth, we can interchange the derivatives with respect to t and $\mathbf{D}$ (Clairaut's Theorem):

$$\frac{\partial}{\partial t}\left(\frac{\partial \log p_t}{\partial \mathbf{D}}\right) = \frac{\partial}{\partial \mathbf{D}}\left(\frac{\partial \log p_t}{\partial t}\right) = \frac{\partial}{\partial \mathbf{D}}\left(\frac{r_p}{p_t}\right)$$

Evaluating this at $t = 0$ (where $p_t = p_0$):

$$\frac{\partial}{\partial t}\left(\frac{\partial \log p_t}{\partial \mathbf{D}}\right)\Big|_{t=0} = \frac{\partial}{\partial \mathbf{D}}\left(\frac{r_p(\mathbf{D}|\mathbf{X})}{p_0(\mathbf{D}|\mathbf{X})}\right)$$

Substituting back into the expectation for the Gateaux derivative gives:

$$\mathcal{D}_p = \mathbb{E}_{\mathbf{W}}\left[-\frac{\partial}{\partial \mathbf{D}}\left(\frac{r_p(\mathbf{D}|\mathbf{X})}{p_0(\mathbf{D}|\mathbf{X})}\right)\left(\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})\right)\right]$$

We now apply the Law of Iterated Expectations, conditioning on $(\mathbf{D}, \mathbf{X})$:

$$\mathcal{D}_p = \mathbb{E}_{(\mathbf{D},\mathbf{X})}\left[-\frac{\partial}{\partial \mathbf{D}}\left(\frac{r_p}{p_0}\right)\mathbb{E}[\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})|\mathbf{D}, \mathbf{X}]\right]$$

From the model setup, $\mathbb{E}[\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})|\mathbf{D}, \mathbf{X}] = \mathbb{E}[\mathbf{U}|\mathbf{D}, \mathbf{X}] = 0$. Thus:

$$\mathcal{D}_p = \mathbb{E}_{(\mathbf{D},\mathbf{X})}\left[-\frac{\partial}{\partial \mathbf{D}}\left(\frac{r_p}{p_0}\right) \cdot 0\right] = 0$$

The score is therefore orthogonal to perturbations in p.

□

D.3 A Practical Score via a Semi-Parametric Assumption

The general score in Proposition 1 is difficult to implement, as it requires non-parametric estimation of a conditional density and its derivative. A practical score is obtainable by imposing a semi-parametric structure on the treatment model.

Assumption 1 (Heteroskedastic Gaussian Treatment Model)**.** *The conditional distribution of the treatment* $\mathbf{D}$ *given the confounders* $\mathbf{X}$ *is Gaussian, with mean function* $m_0(\mathbf{X})$ *and variance function* $v_0(\mathbf{X})$*:*

$$\mathbf{D}|\mathbf{X} \sim \mathcal{N}(m_0(\mathbf{X}), v_0(\mathbf{X}))$$

Proposition 2. *Under Assumption 1, the general Neyman-orthogonal score from Proposition 1 simplifies to:*

$$\psi(\mathbf{W}; \theta, \eta) = \frac{\partial l(\mathbf{D}, \mathbf{X})}{\partial \mathbf{D}} - \theta + \frac{(\mathbf{D} - m(\mathbf{X}))}{v(\mathbf{X})}(\mathbf{Y} - l(\mathbf{D}, \mathbf{X}))$$

where the set of nuisance functions is now $\eta = (l, m, v)$*.*

Proof of Score Simplification. We need to evaluate the term $\frac{1}{p(\mathbf{D}|\mathbf{X})}\frac{\partial p(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}}$, which is the score of the log-likelihood, $\frac{\partial \log p(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}}$. Under Assumption 1, the conditional log-density is:

$$\log p(\mathbf{D}|\mathbf{X}) = -\frac{1}{2}\log(2\pi v(\mathbf{X})) - \frac{(\mathbf{D} - m(\mathbf{X}))^2}{2v(\mathbf{X})}$$

Taking the partial derivative with respect to $\mathbf{D}$ (treating $m(\mathbf{X})$ and $v(\mathbf{X})$ as constant with

respect to $\mathbf{D}$):

$$\frac{\partial \log p(\mathbf{D}|\mathbf{X})}{\partial \mathbf{D}} = -\frac{2(\mathbf{D} - m(\mathbf{X}))}{2v(\mathbf{X})} = -\frac{\mathbf{D} - m(\mathbf{X})}{v(\mathbf{X})}$$

Substituting this result into the general score function from Proposition 1 yields the desired practical score:

$$\psi(\dots) = \frac{\partial l}{\partial \mathbf{D}} - \theta - \left(-\frac{\mathbf{D} - m(\mathbf{X})}{v(\mathbf{X})}\right)(\mathbf{Y} - l(\mathbf{D}, \mathbf{X})) = \frac{\partial l}{\partial \mathbf{D}} - \theta + \frac{\mathbf{D} - m(\mathbf{X})}{v(\mathbf{X})}(\mathbf{Y} - l(\mathbf{D}, \mathbf{X}))$$

□

Proof of Orthogonality for the Practical Score. The score in Proposition 2 is orthogonal to perturbations in l by the same argument as in Proposition 1. We must also show it is orthogonal to perturbations in the new nuisance functions m and v. The argument is as follows:

- **Orthogonality w.r.t.** m: Let $m_t = m_0 + t[r_m]$ be a perturbed path for the true function m_0. The Gateaux derivative of the expected score is:

  $$\frac{\partial}{\partial t}\mathbb{E}\left[\psi(\mathbf{W}; \theta_0, l_0, m_t, v_0)\right]\Big|_{t=0} = \mathbb{E}\left[\frac{\partial}{\partial t}\left(\frac{\mathbf{D} - m_t(\mathbf{X})}{v_0(\mathbf{X})}\right)\Big|_{t=0}(\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X}))\right] = \mathbb{E}\left[\frac{-r_m(\mathbf{X})}{v_0(\mathbf{X})}(\mathbf{Y} - l_0)\right]$$

  By the Law of Iterated Expectations, conditioning on $\mathbf{X}$:

  $$\mathbb{E}_{\mathbf{X}}\left[\frac{-r_m(\mathbf{X})}{v_0(\mathbf{X})}\mathbb{E}[\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})|\mathbf{X}]\right]$$

  Since $\mathbb{E}[\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X})|\mathbf{X}] = \mathbb{E}[\mathbb{E}[\mathbf{Y} - l_0|\mathbf{D}, \mathbf{X}]|\mathbf{X}] = \mathbb{E}[0|\mathbf{X}] = 0$, the entire expression is zero.

- **Orthogonality w.r.t.** v: Similarly, for a path $v_t = v_0 + t[r_v]$, the Gateaux derivative is:

  $$\frac{\partial}{\partial t}\mathbb{E}\left[\psi(\mathbf{W}; \theta_0, l_0, m_0, v_t)\right]\Big|_{t=0} = \mathbb{E}\left[\frac{\partial}{\partial t}\left(\frac{\mathbf{D} - m_0(\mathbf{X})}{v_t(\mathbf{X})}\right)\Big|_{t=0}(\mathbf{Y} - l_0)\right]$$

$$= \mathbb{E}\left[-\frac{\mathbf{D} - m_0(\mathbf{X})}{v_0(\mathbf{X})^2} r_v(\mathbf{X})(\mathbf{Y} - l_0)\right]$$

Again, conditioning on $\mathbf{X}$:

$$\mathbb{E}_{\mathbf{X}}\left[-\frac{r_v(\mathbf{X})}{v_0(\mathbf{X})^2}\mathbb{E}[(\mathbf{D} - m_0(\mathbf{X}))(\mathbf{Y} - l_0(\mathbf{D}, \mathbf{X}))|\mathbf{X}]\right]$$

The inner expectation is zero because $\mathbb{E}[(\mathbf{D} - m_0)(\mathbf{Y} - l_0)|\mathbf{X}] = \mathbb{E}_{\mathbf{D}|\mathbf{X}}[(\mathbf{D} - m_0)\mathbb{E}[\mathbf{Y} - l_0|\mathbf{D}, \mathbf{X}]|\mathbf{X}] = \mathbb{E}_{\mathbf{D}|\mathbf{X}}[(\mathbf{D} - m_0) \cdot 0|\mathbf{X}] = 0$. Thus, the score is orthogonal to perturbations in both m and v.

□

E Robustness of the Median Estimator for the APE

In this appendix, we describe the justification for using a median-of-scores estimator and non-parametric bootstrap for inference, instead of the sample mean of the scores and its analytical variance. It is essential to make this choice in order ensure our causal estimates are robust against the estimation noise from the machine learning nuisance functions, which could cause score distributions to have fat tails and high skewness.

E.1 The Problem: Unreliable Estimation with Heavy-Tailed Scores

The practical score function for APE has a bias correction term that is inversely proportional to the conditional variance of treatment: $\hat{v}(\mathbf{X})$ (this treatment nuisance function is fitted with a machine learning algorithm). When $\hat{v}(\mathbf{X})$ is predicted to be small or close to zero for some observations, the bias correction term produces extreme outliers in the scores distribution $\hat{\psi}i$. When these outliers are present, the sample mean is no longer a reliable way to estimate the central tendency of the distribution. In Figure 14, we demonstrate this problem for the treatment variable amihud_illiquidity_trend_z_scaled_std. The distribution is shown to be skewed and heavy-tailed. There are a few extreme positive outliers that almost cancel the

statistical weight of the much more numerous scores with small negative values. Thus, we get a sample mean (0.0010) that is deceivingly close to zero. A naive interpretation of the mean would have resulted in a mischaracterization of a null causal effect.

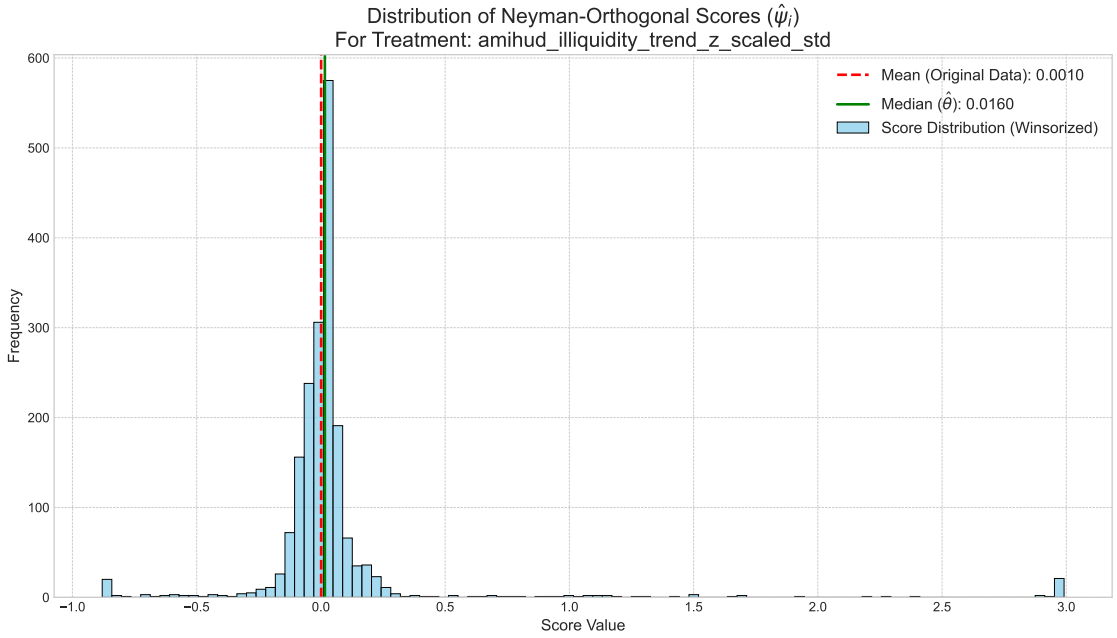


Figure 14: Distribution of Neyman-Orthogonal Scores for the APE

Notes: This figure shows the histogram of the calculated Neyman-orthogonal scores ($\hat{\psi}_i$) for the DML-APE model where the treatment variable is amihud_illiquidity_trend_z_scaled_std. The distribution is heavy-tailed and skewed. The red dashed line indicates the sample mean of the original scores (0.0010). The green solid line marks the sample median (0.0160), which is used as the robust point estimate for the APE ($\hat{\theta}$). This plot demonstrates how the sample mean can be a misleading measure of central tendency; in this case, it is driven toward zero by the cancelling effects of outliers, while the median robustly captures the positive shift in the core of the distribution. For visualization, the scores have been winsorized at the 1st and 99th percentiles.

E.2 The Solution: Median Estimator & Bootstrap Inference

Instead of calculating the mean, we can use the sample median as a robust point estimator. The median is the 50th percentile, and its actual value only depends on all of the scores at the center of the distribution, immune to the outliers are in the tails. Therefore, by using the median of the $\hat{\psi}_i$ scores as a point estimate for θ_0, we have an estimator that is robust to noisy outputs from the nuisance models.

Since analytical formulas for the standard error are also not suitable for heavy-tailed distribution, we use a non-parametric bootstrap to estimate the standard-error for the median. We resample our calculated scores repeatedly and compute the median from each bootstrap sample, producing an empirical sampling distribution for our estimator. This bootstrapping allows us to produce a more credible standard-error.

E.3 Significance

This procedure is beneficial for credible inference. As shown in Figure 14, if we calculated the sample mean, we would get a point estimate around zero and would erroneously concluded a null causal effect (Type II error). The sample median, on the other hand, identifies the positive and statically significant signal. The median/bootstrap median framework helps defend our analysis against the influence of outliers, enhancing the ability to identify credible economic conclusions, and systematically minimize the likelihood of erroneous null findings.

F Online Appendix

A supplementary online appendix, containing the complete results of the Double / Debiased Machine Learning (DML) estimation and the full sensitivity analysis for all variables tested, is available at the following persistent URL: github.com/jackraorpl/market-trough-prediction-appendix.

Table 15: Complete Robust Causal Estimates from the DML-APE Model

Treatment Variable	Coeff. ($\hat{\theta}$)	p-value	bias_phi	Adj. 95% CI Lower	Adj. 95% CI Upper	Benchmark R_Y^2	Benchmark R_D^2
pcr_oi_ewm_slow_scaled_std	0.0134	0.0000	0.0000	0.0114	0.0154	0.1234	0.0000
gex_volume_ewm_fast_scaled_trend	0.0000	0.0000	0.0000	0.0000	0.0000	0.0361	0.0000
vrp_ewm_slow_scaled_mean	0.0380	0.0000	0.0195	0.0008	0.0752	0.0354	0.2311
flow_concentration_10d_scaled_mean	0.0032	0.0000	0.0000	0.0017	0.0047	0.0721	0.0000
fx_rv_6e_21d_scaled_last	-0.0010	0.0000	0.0000	-0.0014	-0.0005	0.0746	0.0000
ffr_slope_ewm_fast_scaled_trend	-0.0001	0.0000	0.0000	-0.0001	0.0000	0.0655	0.0000
ffr_slope_trend_z_scaled_trend	0.0000	0.0000	0.0000	0.0000	0.0000	0.0655	0.0000
ffr_slope_ewm_slow_scaled_trend	-0.0002	0.0000	0.0000	-0.0002	-0.0001	0.0655	0.0000
fx_momentum_6e_21d_trend_z_scaled_std	0.0028	0.0000	0.0000	0.0017	0.0040	0.0746	0.0000
fx_momentum_6j_21d_trend_z_scaled_std	0.0018	0.0000	0.0000	0.0011	0.0025	0.0728	0.0000
pcr_oi_ewm_slow_scaled_last	-0.0011	0.0000	0.0000	-0.0016	-0.0007	0.1234	0.0000
fx_rv_6e_21d_trend_z_scaled_mean	-0.0010	0.0000	0.0000	-0.0013	-0.0006	0.0746	0.0000
ofi_scaled_mean	0.0051	0.0000	0.0000	0.0033	0.0070	0.0721	0.0000
fx_momentum_6e_21d_scaled_std	0.0029	0.0000	0.0000	0.0019	0.0039	0.0746	0.0000
ofi_ewm_fast_scaled_mean	0.0470	0.0000	0.0000	0.0320	0.0620	0.0721	0.0000
risk_neutral_kurtosis_ewm_slow_scaled_std	0.0084	0.0000	0.0000	0.0058	0.0110	0.1493	0.0000
ofi_ewm_slow_scaled_trend	0.1937	0.0000	0.0000	0.1121	0.2752	0.0721	0.0000
vrp_roc63_scaled_std	0.0000	0.0000	0.0000	0.0000	0.0000	0.0354	0.0000
ofi_ewm_slow_scaled_mean	0.0317	0.0000	0.0000	0.0231	0.0403	0.0721	0.0000
fx_rv_6e_21d_roc63_scaled_mean	-0.0026	0.0000	0.0000	-0.0033	-0.0020	0.0746	0.0000
fx_momentum_6j_21d_roc63_scaled_std	0.0082	0.0000	0.0000	0.0062	0.0102	0.0728	0.0000
fx_momentum_6e_21d_ewm_slow_scaled_std	0.0044	0.0000	0.0000	0.0038	0.0051	0.0746	0.0000
flow_concentration_10d_ewm_fast_scaled_mean	0.0456	0.0000	0.0000	0.0362	0.0550	0.0721	0.0000
flow_concentration_10d_ewm_slow_scaled_mean	0.1015	0.0000	0.0000	0.0834	0.1195	0.0721	0.0000
flow_concentration_10d_ewm_slow_scaled_last	0.0496	0.0000	0.0000	0.0408	0.0585	0.0721	0.0000
flow_concentration_10d_ewm_fast_scaled_last	0.0176	0.0000	0.0000	0.0124	0.0229	0.0721	0.0000
ffr_slope_ewm_fast_scaled_std	0.0113	0.0000	0.0000	0.0103	0.0124	0.0655	0.0000
fx_rv_6e_21d_ewm_spread_scaled_mean	-0.0028	0.0000	0.0000	-0.0035	-0.0022	0.0746	0.0000
risk_neutral_kurtosis_roc63_scaled_std	0.0236	0.0000	0.0000	0.0215	0.0258	0.1493	0.0000
fx_rv_6e_21d_ewm_spread_scaled_last	-0.0020	0.0000	0.0000	-0.0025	-0.0016	0.0746	0.0000
pcr_volume_ewm_slow_scaled_mean	0.0747	0.0000	0.0272	0.0338	0.1157	0.1234	0.1532
ofi_ewm_slow_scaled_last	0.0377	0.0000	0.0000	0.0333	0.0420	0.0721	0.0000
fx_rv_6e_21d_ewm_fast_scaled_std	0.0000	0.0001	0.0000	0.0000	0.0000	0.0746	0.0000
fx_rv_6e_21d_trend_z_scaled_last	-0.0009	0.0001	0.0000	-0.0014	-0.0004	0.0746	0.0000
fx_momentum_6e_21d_ewm_fast_scaled_trend	0.0000	0.0001	0.0000	0.0000	0.0000	0.0746	0.0000
ffr_basis_ewm_spread_scaled_last	0.0000	0.0002	0.0000	0.0000	0.0000	0.0655	0.0000
fx_momentum_6e_21d_ewm_spread_scaled_trend	0.0000	0.0003	0.0000	0.0000	0.0000	0.0746	0.0000
ffr_basis_ewm_spread_scaled_mean	0.0000	0.0004	0.0000	0.0000	0.0000	0.0655	0.0000
fx_rv_6e_21d_scaled_mean	-0.0225	0.0008	0.0000	-0.0357	-0.0094	0.0746	0.0000
ffr_basis_ewm_spread_scaled_trend	0.0000	0.0010	0.0000	0.0000	0.0000	0.0655	0.0000
gex_volume_ewm_spread_scaled_std	0.0000	0.0011	0.0000	0.0000	0.0000	0.0361	0.0000
fx_rv_6e_21d_ewm_slow_scaled_std	0.0000	0.0016	0.0000	0.0000	0.0000	0.0746	0.0000
fx_rv_6e_21d_ewm_fast_scaled_last	-0.0138	0.0019	0.0000	-0.0225	-0.0051	0.0746	0.0000
fx_rv_6j_21d_ewm_fast_scaled_mean	-0.0007	0.0040	0.0000	-0.0012	-0.0002	0.0728	0.0000
pcr_oi_ewm_slow_scaled_mean	-0.0008	0.0068	0.0000	-0.0014	-0.0002	0.1234	0.0000
fx_momentum_6j_21d_ewm_resid_scaled_std	0.0016	0.0074	0.0000	0.0004	0.0027	0.0728	0.0000
fx_rv_6e_21d_ewm_resid_scaled_last	-0.0006	0.0096	0.0000	-0.0011	-0.0002	0.0746	0.0000
gex_volume_ewm_slow_scaled_last	0.0009	0.0096	0.0000	0.0002	0.0016	0.0361	0.0000
ffr_basis_roc63_scaled_mean	0.0000	0.0100	0.0000	0.0000	0.0000	0.0655	0.0000
ffr_slope_roc63_scaled_mean	0.0000	0.0126	0.0000	0.0000	0.0000	0.0655	0.0000
gex_volume_ewm_slow_scaled_trend	0.0000	0.0218	0.0000	0.0000	0.0000	0.0361	0.0000
risk_neutral_skewness_ewm_fast_scaled_std	0.0000	0.0260	0.0000	0.0000	0.0000	0.1493	0.0000

Notes: This table reports the complete set of 52 statistically significant ($p < 0.05$) causal estimates from the Double/Debiased Machine Learning Average Partial Effect (DML-APE) model that are robust to unobserved confounding. The analysis uses daily data from April 2013 to June 2025 ($N = 3068$). The dependent variable is a binary indicator for a market trough. For each treatment variable, all other engineered features are included as high-dimensional controls, subject to the exclusion protocol in Section 7.3.1.

The coefficient ($\hat{\theta}$) is the Average Partial Effect (APE), representing the average change in trough probability for a one-unit change in the treatment, averaged over the data distribution. Following the procedure in Section 7.2, the point estimate is the median of the Neyman-orthogonal scores, and the 95% confidence intervals and p-values are derived from a non-parametric bootstrap of these scores. This method is chosen for its robustness to outliers in the score function, as justified in E.

All findings are validated using the sensitivity analysis framework of Cinelli and Hazlett (2020). Benchmark R_Y^2 and Benchmark R_D^2 report the out-of-sample partial R^2 of the outcome and the treatment explained by the observed confounders, respectively. These values serve as a benchmark for the plausible strength of an unobserved confounder. The results are deemed robust if the adjusted 95% confidence interval, which accounts for potential bias from a hypothetical confounder as strong as the observed ones, still excludes zero.