

# Sign-Invariant Cascading Threshold Classification with Ternary Moderation

A General Framework for Regime-Conditional Binary Prediction

S. Castaglia\*  
*Zoltar Predicts*

June 11, 2026

## Abstract

We introduce the *Sign-Invariant Cascading Threshold Classifier* (SICTC), a supervised classification framework for binary prediction problems where predictor–outcome relationships vary across identifiable regimes. SICTC extends binary logistic regression in three ways: (1) a ternary-state moderator partition that decomposes the sample space into  $3^n$  disjoint regimes before classification, (2) a cascading threshold architecture where Youden’s  $J$  statistic optimizes sequential binary gates on standardized features, creating a four-model conditional structure within each regime, and (3) a *sign-invariance* property whereby each sub-model autonomously detects and inverts contrarian predictors—predictors whose accuracy falls below 50% are automatically flipped, transforming reliably wrong signals into reliably right ones without manual intervention. We formalize the framework in domain-general terms, prove its completeness properties, derive a generalized Bayesian quality metric ( $O$ -value) with an impact-normalized variant that corrects for outcome-magnitude bias, and establish relationships to Hayes’ conditional process models, classical logistic regression, decision tree architectures, and modern deep learning. We demonstrate that SICTC occupies a distinct niche in the machine learning landscape: a structured, interpretable, data-efficient architecture for regime-conditional classification on tabular

data, complementary to rather than competitive with representation-learning approaches such as transformers.

**Keywords:** regime-conditional classification, sign-invariant prediction, Youden’s  $J$ , ternary moderation, impact-normalized scoring, walk-forward validation, interpretable machine learning

## 1 Introduction

Standard logistic regression models the probability of a binary outcome  $Y \in \{-1, +1\}$  as a smooth function of a predictor  $X$ :

$$\mathbb{P}(Y = 1 \mid X) = \sigma(\beta_0 + \beta_1 X), \quad (1)$$

where  $\sigma(\cdot)$  is the logistic sigmoid. This formulation assumes a single decision boundary, homogeneous effects across the sample, and a fixed sign relationship between predictor and outcome.

In many applied domains, all three assumptions are routinely violated:

- (a) **Regime-dependent relationships.** The relationship between predictors and outcomes changes with the state of the system—a biomarker that predicts treatment response in early-stage disease may mislead in advanced stages; macroeconomic variables that predict market direction during expansions may fail during contractions.
- (b) **Heterogeneous effects.** A predictor that works in one subgroup or system state may fail in another, not because it is poor, but because the generative process has changed.

---

\*Corresponding author.  
zoltarpredicts@gmail.com

Email:

- (c) **Contrarian indicators.** Some variables are predictive precisely because their relationship to the outcome is *inverse*. A variable that reliably predicts the wrong class carries the same information as one that reliably predicts the right class—it merely requires sign inversion.

We propose a framework that addresses all three violations while maintaining interpretability and walk-forward validity. The key insight is that a predictor which is reliably *wrong* carries the same information content as one that is reliably *right*—it merely requires sign inversion. The framework automatically detects and applies this inversion at each node of a cascading conditional structure, moderated by a ternary regime classification.

While we develop the framework in domain-general terms, we use financial prediction as the primary worked example, as it provides a natural setting where all three violations co-occur.

## 2 Formal Framework

### 2.1 Notation

Let  $\{(\mathbf{x}_t, y_t)\}_{t=1}^T$  denote  $T$  observations where  $\mathbf{x}_t \in \mathbb{R}^p$  is the predictor vector ( $p$  input variables, standardized),  $y_t \in \{-1, +1\}$  is the binary outcome, and  $w_t \in \mathbb{R}_{\geq 0}$  is the outcome magnitude (a weight reflecting the importance of observation  $t$ ). The framework is domain-agnostic; in the financial instantiation,  $y_t$  is the realized market direction and  $w_t = |r_t|$  is the absolute realized return.

For each predictor variable  $j$ , we construct rolling  $z$ -scored features at multiple timescales using only trailing data (walk-forward safe):

$$z_{j,t}^{(6)} = \frac{\Delta_j^{(6)}(t) - \bar{\mu}_{j,t}^{(6)}}{\bar{\sigma}_{j,t}^{(6)}}, \quad (2)$$

where  $\Delta_j^{(k)}(t)$  is the  $k$ -month percent change of variable  $j$  at time  $t$ , and  $\bar{\mu}, \bar{\sigma}$  are computed over a trailing window of  $w = 60$  months. Similarly for  $z^{(12)}, z^{(24)}$ , with velocity  $v_{j,t} = z_{j,t}^{(6)} - z_{j,t-1}^{(6)}$  and acceleration  $a_{j,t} = v_{j,t} - v_{j,t-1}$ .

### 2.2 Ternary Moderator Partition

**Definition 1** (Ternary Sign Function). For a continuous variable  $c \in \mathbb{R}$  and dead-zone threshold  $\delta > 0$ :

$$\text{sgn}_\delta(c) = \begin{cases} +1 & \text{if } c > \delta, \\ 0 & \text{if } |c| \leq \delta, \\ -1 & \text{if } c < -\delta. \end{cases} \quad (3)$$

**Definition 2** (Ternary Moderator Partition). Given  $n$  moderating variables  $(c_1, \dots, c_n)$  with dead-zone  $\delta$ , the moderator partition  $\mathcal{M}$  assigns each observation to a regime:

$$\tau = \mathcal{M}(c_1, \dots, c_n) = (\text{sgn}_\delta(c_1), \dots, \text{sgn}_\delta(c_n)) \in \{-1, 0, +1\}^n. \quad (4)$$

Each unique  $n$ -tuple  $\tau$  defines a **type**. For  $n$  moderating variables, the partition has  $|\{-1, 0, +1\}^n| = 3^n$  types.

**Proposition 1** (Exhaustive Coverage). *The partition  $\mathcal{M}$  is exhaustive: for any observation with finite, non-NaN moderating variables, exactly one type  $\tau$  is assigned.*

*Proof.* The ternary sign function  $\text{sgn}_\delta$  maps every real number to exactly one element of  $\{-1, 0, +1\}$ . The Cartesian product of  $n$  such mappings covers  $\mathbb{R}^n$ .  $\square$

*Remark 1.* When  $\delta = 0$ , the zero state requires exact equality ( $c = 0$ ), which has Lebesgue measure zero for continuous variables, effectively reducing the partition to  $2^n$  types. A positive  $\delta$  gives the zero state positive measure, enabling the ternary structure to function empirically. In our implementation,  $n = 3$  (CAPE valuation change, price change, earnings change) with  $\delta = 0.005$  (corresponding to  $\pm 0.5\%$  over six months), yielding  $3^3 = 27$  types.

### 2.3 The Four-Model Cascade

Within each type  $\tau$ , let  $D_\tau \subset \{1, \dots, T\}$  denote the index set of observations assigned to type  $\tau$ .

**Definition 3** (Youden’s  $J$  Threshold). For a predictor feature  $z$  and binary labels  $\{y_t\}_{t \in S}$  on

subset  $S$ , the optimal threshold is:

$$(\theta^*, d^*) = \arg \max_{\theta \in \Theta, d \in \{>, <\}} [\text{Sensitivity}(\theta, d) + \text{Specificity}(\theta, d)] \quad (5)$$

where  $\Theta$  is a set of candidate thresholds (quantiles of  $z$ ), and:

$$\text{Sensitivity}(\theta, d) = \mathbb{P}(\hat{y} = +1 \mid y = +1), \quad (6)$$

$$\text{Specificity}(\theta, d) = \mathbb{P}(\hat{y} = -1 \mid y = -1), \quad (7)$$

with prediction rule  $\hat{y}_t = +1$  if  $z_t \geq d$ , else  $\hat{y}_t = -1$ . Youden’s  $J$  is equivalent to maximizing the vertical distance between the ROC curve and the diagonal [Youden, 1950, Fluss et al., 2005].

The cascade partitions  $D_\tau$  into four disjoint models via sequential application of Youden-optimized thresholds:

---

**Algorithm 1** Four-Model Cascade within Type  $\tau$

---

- 1: **Stage 1:** Select primary variable  $j^*$  by effect size (Welch’s  $t$  or Mann–Whitney  $U$ ). Find  $(\theta_1^*, d_1^*)$  via Eq. (5).
  - 2: Split:  $P = \{t \in D_\tau : z_{j^*,t} \geq d_1^*\}$ ,  $F = D_\tau \setminus P$
  - 3: **Stage 2a** (on  $P$ ): Select secondary variable  $j_2^*$ , find  $(\theta_2^*, d_2^*)$ .
  - 4:  $M_1 = \{t \in P : z_{j_2^*,t} \geq d_2^*\}$ ,  $M_2 = P \setminus M_1$
  - 5: **Stage 2b** (on  $F$ ): Select tertiary variable  $j_3^*$ , find  $(\theta_3^*, d_3^*)$ .
  - 6:  $M_3 = \{t \in F : z_{j_3^*,t} \geq d_3^*\}$ ,  $M_4 = F \setminus M_3$
  - 7: **return**  $D_\tau = M_1 \cup M_2 \cup M_3 \cup M_4$
- 

## 2.4 Sign Invariance: The Flip Mechanism

**Definition 4** (Sign-Invariant Classifier). A binary classifier is *sign-invariant* if it achieves effective accuracy:

$$\alpha_{\text{eff}}(M_k) = \max(\alpha_{\text{raw}}(M_k), 1 - \alpha_{\text{raw}}(M_k)), \quad (8)$$

where  $\alpha_{\text{raw}}$  is the fraction of observations where the unflipped prediction matches the realized direction. When  $\alpha_{\text{raw}} < 0.5$ , predictions are inverted:

$$\hat{y}_t^{\text{final}} = \begin{cases} -\hat{y}_t^{\text{raw}} & \text{if } \alpha_{\text{raw}}(M_k) < 0.5, \\ \hat{y}_t^{\text{raw}} & \text{otherwise.} \end{cases} \quad (9)$$

**Proposition 2** (Information Equivalence). A predictor with raw accuracy  $\alpha < 0.5$  carries identical mutual information  $I(X; Y)$  as a predictor with accuracy  $(1 - \alpha) > 0.5$ .

*Proof.* Mutual information satisfies  $I(X; Y) = I(X; -Y)$ . The transformation  $\hat{y} \mapsto -\hat{y}$  is equivalent to label negation  $Y \mapsto -Y$ . A predictor correct with probability  $\alpha$  under original labels is correct with probability  $(1 - \alpha)$  under negated labels. The flip at 0.5 selects whichever labeling yields  $\alpha_{\text{eff}} \geq 0.5$ , recovering the full mutual information.  $\square$

*Remark 2* (Relationship to Logistic Regression). In standard logistic regression (1), a negative coefficient  $\beta_1 < 0$  automatically encodes an inverse predictor–outcome relationship. The model learns this continuously through maximum likelihood estimation. SICTC achieves the same effect through the discrete flip at 0.5, but with a critical advantage: the flip is applied *independently per type and per cascade model*. In logistic regression, achieving the same per-regime sign flexibility would require a fully interacted model:

$$\mathbb{P}(Y = 1 \mid X, \tau) = \sigma \left( \sum_{\tau'} \mathbf{1}_{\tau=\tau'} (\beta_{0,\tau'} + \beta_{1,\tau'} X) \right), \quad (10)$$

with  $3^n \times K \times p$  parameters (types  $\times$  cascade models  $\times$  predictors), a combinatorial explosion that SICTC avoids through its sequential architecture.

## 2.5 The $O$ -Value: Bayesian Profit-Weighted Quality

**Definition 5** ( $O$ -Value). For model  $M_k$  with  $n_k$  observations, of which  $c_k$  are correctly predicted after flip adjustment, and with realized returns  $\{r_t\}_{t \in M_k}$ :

$$O(M_k) = (\tilde{\alpha}_k)^{5/2} \times 2 \times \bar{r}_{\text{net},k}, \quad (11)$$

where:

$$\tilde{\alpha}_k = \frac{c_k + 1}{n_k + 2} \quad (12)$$

is the Bayesian (Laplace-smoothed) accuracy estimate—equivalent to the posterior mean under

Table 1: Effect of the 5/2 exponent on accuracy weighting. A 55%-accurate model receives 41% of the weight a linear metric would assign.

Accuracy	$\tilde{\alpha}$	$\hat{\alpha}^1$ (linear)	$\hat{\alpha}^{5/2}$
55%	0.549	0.549	0.224
60%	0.598	0.598	0.277
65%	0.647	0.647	0.337
70%	0.696	0.696	0.404
75%	0.745	0.745	0.479
80%	0.794	0.794	0.562
85%	0.843	0.843	0.653
90%	0.892	0.892	0.752

a Beta(1, 1) prior [Laplace, 1814]—and:

$$\bar{r}_{\text{net},k} = \frac{\sum_{t \in \text{correct}} |r_t| - \sum_{t \in \text{wrong}} |r_t|}{n_k} \quad (13)$$

is the net return per observation.

**Proposition 3** (Properties of the  $O$ -Value).

- (i) **Non-negativity:**  $O(M_k) \geq 0$  when correct predictions capture more return magnitude than incorrect predictions ( $\bar{r}_{\text{net}} \geq 0$ ).
- (ii) **Super-linear accuracy penalization:** For  $\tilde{\alpha} \in (0, 1)$ , the mapping  $\tilde{\alpha} \mapsto \tilde{\alpha}^{5/2}$  is convex and maps  $[0.5, 1.0] \rightarrow [0.177, 1.0]$ , compressing marginal-accuracy signals relative to a linear metric (Table 1).
- (iii) **Profit alignment:** Two models with identical accuracy can have different  $O$ -values if one is correct on large-magnitude moves while the other is correct on small ones.

*Remark 3* (Choice of Exponent). The exponent  $\gamma = 5/2$  was selected empirically through a sensitivity analysis over  $\gamma \in \{1, 3/2, 2, 5/2, 3, 7/2, 4\}$  on walk-forward backtests of S&P 500 monthly returns. The selection criterion was cumulative net return under position sizing proportional to  $O$ -value, evaluated over rolling 20-year out-of-sample windows.

The decision-theoretic motivation connects to the Kelly criterion [Kelly, 1956]. The Kelly-optimal bet fraction  $f^* = (bp - q)/b$ , where  $p$  is the win probability,  $q = 1 - p$ , and  $b$  is the payoff ratio, approaches zero linearly as  $p \rightarrow 0.5$ . A

linear weighting ( $\gamma = 1$ ) still assigns substantial weight to models with marginal accuracy (e.g., 55%). The super-linear exponent compresses the accuracy axis to better approximate the Kelly fraction’s behavior: aggressive penalization of marginal accuracy, with weight concentrating on high-accuracy models.

Empirically, the sensitivity analysis found that  $\gamma \in [2, 3]$  produced stable performance, with  $\gamma = 5/2$  yielding the best Sharpe ratio in the median backtest window. Values below  $3/2$  over-allocated to marginal models (portfolio dominated by noise), while values above  $7/2$  excessively penalized moderate-accuracy models (65–75%), concentrating weight in too few types and increasing variance. The choice  $\gamma = 5/2$  is therefore empirically calibrated rather than analytically derived; alternative applications may benefit from re-calibrating  $\gamma$  to their specific accuracy distribution and payoff structure.

**Definition 6** ( $O$ -Score). The coverage-weighted aggregate across all four models within type  $\tau$ :

$$O_{\text{score}}(\tau) = \sum_{k=1}^4 O(M_k) \cdot \frac{n_k}{|D_\tau|}. \quad (14)$$

## 2.6 Impact-Normalized $O$ -Value

The  $O$ -value as defined uses raw outcome magnitudes  $w_t$  in the net return calculation. This introduces a **magnitude bias**: models operating during high-variance regimes receive inflated  $O$ -values regardless of directional skill, because both correct and incorrect predictions involve large magnitudes.

**Example 1** (Volatility Bias). Consider a predictor for equity returns that also captures VIX-related movements. When the VIX spikes, monthly returns can exceed  $\pm 10\%$ . A model with 55% accuracy during these episodes receives a far higher  $O$ -value than a model with 75% accuracy during calm periods where  $|r| \approx 1\%$ . The raw  $O$ -value rewards the first, despite the second being the better model.

**Definition 7** (Impact Function). An *impact function*  $\phi : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  maps each outcome

Table 2: Impact function variants.

$\phi(w, \sigma)$	Name	Effect
$w$	Raw	Magnitude-sensitive; biased toward volatile regimes
$w/\sigma$	Vol-norm.	Scale-invariant; rewards skill vs. local variability
$\text{rank}(w)/n$	Rank	Fully nonparametric
1	Uniform	Pure accuracy weighting
$\min(w, \kappa)$	Winsorized	Caps extreme magnitudes at $\kappa$

magnitude and a local scale parameter to a normalized impact score. The local scale  $\sigma_t$  is a trailing estimate of outcome variability.

**Definition 8** (Normalized  $O$ -Value). The impact-normalized  $O$ -value is:

$$O^\phi(M_k) = (\tilde{\alpha}_k)^\gamma \times 2 \times \bar{\phi}_{\text{net},k}, \quad (15)$$

where  $\bar{\phi}_{\text{net},k} = (\sum_{t \in \text{correct}} \phi(w_t, \sigma_t) - \sum_{t \in \text{wrong}} \phi(w_t, \sigma_t)) / n_k$ .

The choice of  $\phi$  determines what the  $O$ -value measures:

**Proposition 4** (Scale Invariance). *Under the volatility-normalized impact function  $\phi(w, \sigma) = w/\sigma$ , two models with identical accuracy and identical risk-adjusted net impact have identical  $O$ -values, regardless of the absolute magnitude of outcomes in their respective regimes.*

*Proof.* The normalization  $w/\sigma$  transforms each outcome to units of local standard deviation. If model  $A$  operates with  $\sigma = 8\%$  and model  $B$  with  $\sigma = 1\%$ , a correct prediction capturing a 10% move in regime  $A$  contributes  $10/8 = 1.25$  normalized units, while a 1.5% move in regime  $B$  contributes  $1.5/1.0 = 1.50$  units. The normalized  $O$ -value correctly identifies  $B$  as having greater risk-adjusted skill.  $\square$

Table 3: Structural comparison with Hayes’ PROCESS framework.

Concept	Hayes	SICTC
Moderator	Continuous $W$	Discrete $\tau \in \{-1, 0, +1\}^n$
Interaction	$XW$ product	Separate model/type
Estimation	Simultaneous	Sequential cascade
Mediator	Continuous path	Binary threshold
Effect sign	Fixed $\beta$	Adaptive (flip)

### 3 Relationship to Existing Frameworks

#### 3.1 Hayes’ PROCESS Models

Hayes’ conditional process framework [Hayes, 2022] models moderated mediation as:

$$Y = b_0 + b_1X + b_2W + b_3XW + e, \quad (16)$$

where  $W$  is a continuous moderator and  $XW$  is the interaction. Table 3 summarizes the correspondence.

The critical structural difference is estimation strategy. Hayes estimates all paths simultaneously, yielding proper standard errors for conditional indirect effects. SICTC estimates sequentially—Model 3’s search space is conditioned on the boundary chosen by Model 1’s threshold. This dependency is compensated through walk-forward validation rather than analytical standard errors.

#### 3.2 Decision Trees (CART/CHAID)

SICTC shares structural similarity with classification trees: both use recursive binary splits optimizing a criterion at each node. Key differences include:

- CART selects the best variable independently at each node and is free to re-select the same variable at successive depths—often producing correlated splits on a single dominant feature. SICTC *requires* a different variable at each cascade stage, which functions as a structural regularizer that forces the cascade to incorporate diverse predictor information and avoids collinear splits.

- CART does not invert the target; SICTC’s flip mechanism has no standard analogue.
- SICTC’s cascade is fixed at depth 2, preventing overfitting through deep trees, at the cost of reduced expressiveness relative to fully grown CART.
- Youden’s  $J$  maximizes Sensitivity + Specificity  $-1$ , equivalent to maximizing vertical distance from the ROC diagonal, whereas CART minimizes Gini impurity or cross-entropy. Unlike impurity measures, Youden’s  $J$  explicitly balances sensitivity and specificity, making it better suited to directional prediction where both false positives and false negatives carry economic cost.

### 3.3 Logistic Regression

Standard logistic regression is a limiting case of SICTC with  $n = 0$  moderating variables (single regime), no cascade, no flip mechanism, and a smooth sigmoid boundary replacing the step-function threshold. The equivalence:

$$\sigma(\beta_0 + \beta_1 X) = 0.5 \iff X = -\frac{\beta_0}{\beta_1} \equiv \theta^*, \tag{17}$$

shows that logistic regression’s decision boundary is a single Youden threshold with  $d^* = \text{sgn}(\beta_1)$ . SICTC generalizes this to  $3^n \times K$  independent thresholds with adaptive signs.

### 3.4 Deep Learning and Transformers

SICTC and modern deep learning architectures occupy complementary regions of the bias–variance–data tradeoff space.

**Inductive bias.** Transformers [Vaswani et al., 2017] and deep networks are low-bias, high-variance architectures that learn representations with minimal structural assumptions. Given sufficient data ( $10^4$ – $10^8$  examples), a transformer can discover regime structure, sign inversions, and conditional effects without being told they exist. SICTC is high-bias, low-variance: it assumes identifiable regimes (ternary partition), conditionally

independent predictor effects within regimes (cascade), and possible sign inversion (flip). When these assumptions hold, SICTC extracts the relevant structure with far fewer observations.

**Data efficiency.** SICTC requires  $10^2$ – $10^3$  observations per regime for stable estimates. A transformer on the same tabular task would typically require  $10^4$ + observations to learn comparable conditional structure—and the learned structure would be opaque. For domains with small-to-moderate datasets and known regime structure, SICTC’s assumptions convert prior knowledge into statistical efficiency.

**Interpretability.** Every SICTC component is inspectable: regime assignment, variable selection, threshold direction, and flip status. This transparency is unavailable in deep architectures.

**Where each excels.** Transformers dominate for sequential data (language, long-range time series), high-dimensional inputs (images, audio), and massive datasets. SICTC offers advantages for tabular data with moderate dimensionality ( $p < 100$ ), known or suspected regime structure, small datasets, contrarian indicators, interpretability requirements, and walk-forward deployment without full retraining. SICTC is not an alternative to transformers but a structured inductive bias for problems transformers handle inefficiently—analogous to a Kalman filter being the right tool when state-space assumptions hold, rather than a competitor to neural sequence models.

**Hybrid architectures.** A natural extension uses a neural network to learn the regime-embedding function (replacing the hand-specified ternary partition with a learned partition) while retaining the cascade and flip within each learned regime. This preserves interpretability at the regime and cascade level while allowing regime boundaries to be learned from data.

## 4 Walk-Forward Validity

All components of SICTC are designed for genuine out-of-sample prediction:

1. **Z-score normalization** uses rolling trailing windows. No future data enters the standardization. Savitzky–Golay velocity features use trailing coefficients ( $\text{pos} = w - 1$ ).
2. **Threshold optimization:** Youden’s  $J$  at time  $t$  uses only  $\{(\mathbf{x}_s, y_s)\}_{s=1}^{t-1}$ .
3. **Flip decision:** the accuracy check ( $\alpha_{\text{raw}} < 0.5?$ ) uses only realized outcomes up to  $t - 1$ .
4. **O-value:** Bayesian accuracy and net return use only historical observations.

This ensures the framework produces walk-forward valid predictions. The cost is wider uncertainty during early periods, reflected in the Laplace prior’s regularization toward 50%.

### 4.1 Walk-Forward Validation as Sequential Bias Mitigation

The cascade architecture introduces a structural dependency between stages: the fail set  $F = D_\tau \setminus P$  on which Model 3 is estimated is itself defined by the primary threshold  $\theta_1^*$ , which is estimated on  $D_\tau$ . In simultaneous estimation frameworks such as Hayes’ PROCESS models, all paths are estimated jointly and standard errors account for the full covariance structure. In SICTC, the sequential architecture means that analytical standard errors underestimate the true system variance, because they condition on the cascade structure as fixed when it is in fact estimated.

Walk-forward validation provides the appropriate mitigation. Rather than deriving analytical confidence intervals that assume independence between cascade stages, the framework evaluates each threshold on data that was not used to estimate it. Concretely: the threshold  $\theta_1^*$  is estimated on training observations  $\{(\mathbf{x}_s, y_s)\}_{s=1}^{t-1}$ , and its discriminative power is evaluated on subsequent observations  $\{(\mathbf{x}_s, y_s)\}_{s=t}^T$  that played no role in the optimization. This out-of-sample evaluation absorbs the sequential dependency: if a primary threshold overfits, the secondary and tertiary thresholds estimated on its residuals will

show degraded walk-forward performance, and the composite quality score will reflect this degradation. The bias is not eliminated analytically—it is revealed empirically.

This is the same logic that motivates cross-validation in any multi-stage pipeline (e.g., feature selection followed by model fitting), but applied to the specific structure of the cascade. The key property is that walk-forward evaluation penalizes exactly the failure mode that sequential estimation bias creates: thresholds that appear discriminative in-sample but whose apparent power derives from conditioning on an overfitted upstream split.

### 4.2 Statistical Inference via Bootstrap

SICTC’s sequential architecture precludes standard analytical inference: the cascade introduces dependencies between stages, and the flip mechanism is a discrete operation with no smooth analogue for which standard errors can be derived. We therefore adopt a nonparametric bootstrap procedure [Efron and Tibshirani, 1993] to construct confidence intervals for the  $O$ -value and  $O$ -score.

#### Bootstrap confidence intervals for $O$ -value.

For each model  $M_k$  within type  $\tau$ , with  $n_k$  assigned observations:

1. Draw  $B$  bootstrap resamples of size  $n_k$  with replacement from  $\{(\hat{y}_t, y_t, r_t)\}_{t \in M_k}$ .
2. For each resample  $b = 1, \dots, B$ : recompute  $\tilde{\alpha}_k^{(b)} = (c_k^{(b)} + 1)/(n_k + 2)$ , the net return  $\tilde{r}_{\text{net},k}^{(b)}$ , and  $O^{(b)}(M_k) = (\tilde{\alpha}_k^{(b)})^{5/2} \times 2 \times \tilde{r}_{\text{net},k}^{(b)}$ .
3. The  $(1 - \alpha)$  confidence interval is the  $[\alpha/2, 1 - \alpha/2]$  quantiles of  $\{O^{(b)}\}_{b=1}^B$ .

In practice,  $B = 10,000$  provides stable quantile estimates. For the  $O$ -score (the coverage-weighted aggregate), the bootstrap resamples within each model independently and recomputes the weighted sum at each iteration, preserving coverage weights.

**Hypothesis testing.** To test whether a type’s  $O$ -score is significantly greater than zero (i.e., the model has genuine predictive value beyond

chance), we compute the bootstrap  $p$ -value:

$$p = \frac{1}{B} \sum_{b=1}^B \mathbf{1} \left[ O_{\text{score}}^{(b)}(\tau) \leq 0 \right]. \quad (18)$$

A type with  $p < 0.05$  provides evidence that its predictive accuracy and return-capture jointly exceed what would be expected under the null of no predictive power.

**Limitations of the bootstrap in this context.** The procedure as described bootstraps the prediction–outcome pairs *within* each model’s assigned subset, holding the cascade structure fixed. This underestimates the true uncertainty because it does not account for variability in the threshold selection itself—a different training sample would produce different thresholds  $\theta^*$ , different pass/fail sets, and different model assignments. A fuller assessment would embed the entire cascade fitting procedure within the bootstrap loop (re-running Youden’s  $J$  optimization on each resample), at substantially higher computational cost. We recommend this full-cascade bootstrap for final model validation, while the conditional bootstrap described above is suitable for rapid screening of type quality during development.

## 5 Completeness

**Theorem 5** (Partition Completeness). *For  $n$  moderating variables with dead-zone  $\delta > 0$ , the SICTC framework:*

- (a) *assigns every observation with non-NaN moderating variables to exactly one of  $3^n$  types,*
- (b) *within each type with  $|D_\tau| \geq n_{\min}$ , assigns every observation to exactly one of four models, and*
- (c) *produces a directional prediction  $\hat{y}_t \in \{-1, +1\}$  for every assigned observation with  $\alpha_{\text{eff}} \geq 0.5$ .*

*Proof.* Part (a) follows from Proposition 1. Part (b) follows from the binary cascade: every observation either passes or fails each threshold, and  $M_1 \cup M_2 \cup M_3 \cup M_4 = D_\tau$  by construction.

Part (c) follows from the flip mechanism (Definition 4): regardless of  $\alpha_{\text{raw}}$ , the classifier produces  $\alpha_{\text{eff}} = \max(\alpha_{\text{raw}}, 1 - \alpha_{\text{raw}}) \geq 0.5$ .  $\square$

## 6 Limitations and Extensions

**Sequential estimation bias.** The cascade introduces dependency between stages. Model 3’s Youden’s  $J$  is optimized on  $F = D_\tau \setminus P$ , a subset defined by the primary threshold  $\theta_1^*$ . The combined system’s variance is underestimated relative to simultaneous estimation. As discussed in Section 4.1, walk-forward validation provides the primary mitigation: out-of-sample evaluation absorbs the sequential dependency by penalizing thresholds whose apparent discriminative power derives from conditioning on overfitted upstream splits. This does not eliminate the bias analytically but reveals it empirically, which we argue is the appropriate strategy for a sequential architecture where analytical standard errors are unavailable.

**Boundary sensitivity.** Observations with  $|c_i| \approx \delta$  can switch types with small perturbations, a characteristic shared by all partition-based classifiers.

**Sparse types.** With  $3^n$  types and finite data, some types may lack sufficient observations. The minimum per type should satisfy  $n_{\min} \geq 4 \times n_{\text{model}}$ , where  $n_{\text{model}} \geq 15$  for stable threshold estimation, suggesting  $n_{\min} \geq 60$ .

**Extension to continuous output.** SICTC currently produces a direction and quality score, not calibrated probabilities. A natural extension maps  $O(M_k) \rightarrow P(Y = 1)$  via a monotone link function, recovering probability estimates while retaining the cascade and flip architecture.

## 7 Domain Generalization

The framework developed in Sections 2–6 is domain-agnostic. The three architectural components—ternary partition, cascading threshold, and sign invariance—require only a binary

outcome, standardized predictor features, moderating variables, and an outcome magnitude for quality scoring.

## 7.1 The SICTC Interface

Any application requires: (i) a binary outcome  $Y \in \{-1, +1\}$ ; (ii) standardized predictors  $\mathbf{x} \in \mathbb{R}^p$ ; (iii) moderating variables  $(c_1, \dots, c_n)$  with dead-zone  $\delta$ . Optionally, for  $O$ -value computation: outcome magnitudes  $w_t$ , a local scale  $\sigma_t$ , and an impact function  $\phi(w, \sigma)$ . When magnitudes are unavailable, set  $\phi = 1$ , reducing the  $O$ -value to a pure Bayesian accuracy score.

## 7.2 Domain Instantiations

**Clinical diagnostics.**  $Y =$  disease present/absent. Moderators: disease stage, comorbidity burden, age category. The contrarian phenomenon appears in clinical prediction: elevated white blood cell count indicates infection in immunocompetent patients but may be suppressed in immunocompromised patients, inverting the sign. SICTC handles this automatically if immunocompromised status is a moderating variable.

**Manufacturing quality control.**  $Y =$  defect/no-defect. Moderators: operating mode (startup/steady-state/shutdown), material batch quality, ambient conditions. Regime structure is explicit: the relationship between sensor readings and defect probability changes fundamentally between startup transients and steady-state operation.

**Ecological monitoring.**  $Y =$  threshold exceedance (e.g., harmful algal bloom). Moderators: season, water body type, land-use category. Nutrient indicators that predict blooms in shallow lakes may be contrarian in deep stratified lakes.

## 7.3 When to Use SICTC

SICTC is appropriate when: (1) the outcome is binary or meaningfully binarizable; (2) regime

structure is known or strongly suspected; (3) predictor signs may invert across regimes; (4) data is moderate-sized and tabular ( $10^2$ – $10^3$  observations,  $p < 100$ ); and (5) interpretability is valued.

# 8 Implementation

## 8.1 Financial Prediction Instantiation

The framework was developed on monthly financial data:  $n = 3$  moderating variables (CAPE valuation change, price change, earnings change);  $\delta = 0.005$ ; five  $z$ -score features per predictor (6M, 12M, 24M, velocity, acceleration); 60-month rolling lookback; primary  $p < 0.05$ , secondary  $p < 0.40$ ; Laplace prior with pseudocount 1;  $O$ -value exponent  $5/2$ ; impact function: raw  $|r_t|$  (volatility-normalized variant recommended for cross-regime comparisons). The implementation uses Numba JIT compilation for the inner Youden  $J$  search loop and multiprocessing for parallel per-type analysis.

## 8.2 General Guidance

**Moderator selection.** Choose  $n$  moderating variables that capture the domain’s regime structure. Fewer moderators ( $n = 1$  or  $2$ ) yield fewer types with more observations per type; more ( $n \geq 4$ ) yield finer regimes but risk sparsity.

**Dead-zone calibration.** Set  $\delta$  to the smallest practically meaningful change. A useful heuristic:  $\delta$  such that approximately 15–25% of observations fall in the zero state for each moderator.

**Impact function.** When outcome magnitudes are heteroskedastic across regimes, use  $\phi(w, \sigma) = w/\sigma$ . When homoskedastic, use  $\phi(w) = w$ . When unavailable, use  $\phi = 1$ .

## Acknowledgments

The SICTC framework was developed as part of the HYACINTH integrated market prediction system (Zoltar Predicts, 2024–2026).

## References

- B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1993.
- R. Fluss, D. Faraggi, and B. Reiser. Estimation of the Youden index and its associated cut-off point. *Biometrical Journal*, 47(4):458–472, 2005.
- A. F. Hayes. *Introduction to Mediation, Moderation, and Conditional Process Analysis*. Guilford Press, 3rd edition, 2022.
- J. L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926, 1956.
- P. S. Laplace. *Essai philosophique sur les probabilités*. Courcier, Paris, 1814.
- R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.
- W. J. Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.