The Spoofing Wedge

Aryan Ayyar*

June 14, 2026

"The door of the True is covered with a golden disk."
— *Isha Upanishad*, 15, trans. Max Muller (1879)

Abstract

Cancellations are actions; spoofing is a source of value. The bona fide value envelope, the largest payoff a displayed path can earn from execution-facing motives, separates them. A path is spoofing-like when this envelope is negative yet total value turns positive after induced response. With unrestricted hidden exposure, public data cannot identify this response wedge. Audit evidence requires a bona fide upper bound, a path-excluded response counterfactual, and signed exposure. Rational responders learning from displayed depth close the model: spoofing reduces display informativeness and creates an unpriced credibility externality. Classification is inconclusive when any leg cannot be signed and bounded.

*Manipal Academy of Higher Education. I thank Madhu Veeraraghavan, Pro Vice Chancellor of MAHE Bengaluru, for helpful comments and discussions. All errors are my own.

A credible surface can conceal the source of value beneath it. Every limit order a trader displays announces a willingness to trade, yet in modern order-driven markets the great majority of displayed orders are cancelled before they execute. Far from being pathological, cancellation is the ordinary grammar of liquidity supply. Traders revise and withdraw quotes to protect themselves against stale prices, to manage inventory and queue position, to complete hedges, and to avoid being picked off by better-informed counterparties (Kyle, 1985; Glosten and Milgrom, 1985; Biais, Hillion, and Spatt, 1995; Parlour, 1998; Foucault, 1999; Foucault, Kadan, and Kandel, 2005; Rosu, 2009). The institution that makes displayed depth informative also makes deception possible: a trader may post an order that is valuable because others believe it, while privately preferring not to trade. The two motives can generate identical public order paths. A screen that treats cancellation, short order lifetime, or a high order-to-trade ratio as the signature of manipulation is therefore a screen for false positives.

The question that separates the two cases is not what the trader did to the order but why the displayed path was worth choosing. Suppose the path is valuable because the trader might execute, might preserve the option to execute, might shed inventory or complete a hedge, might protect a resting order against a stale quote, or might improve queue position. Then the path earns its value on the trader's own execution frontier, and it is bona fide whether or not anyone else is watching. Suppose instead that the path loses money under every such motive and pays only because it moves other traders' beliefs or actions in a direction the trader is positioned to exploit. Then its value is response-facing, and the path is structurally spoofing-like. Cancellation is an action; spoofing is a source of value. The contribution of this paper is to make that distinction an object rather than a slogan, and to show what it takes to identify.

The object is the *bona fide value envelope*, the largest payoff an order path P in state S

can earn from the admissible class $\mathcal{B}$ of execution-facing motives,

$$\overline{V}^{B}(P;S) = \sup_{b \in \mathcal{B}} V_b^{B}(P;S).$$

A path exhibits the *spoofing wedge* when it is unprofitable on this frontier yet profitable once the response of others is added,

$$\overline{V}^{B}(P;S) < 0 \qquad \text{and} \qquad \overline{V}^{B}(P;S) + D(P;S) - K^{D}(P;S) > 0,$$

where $D(P;S)$ is the causal payoff from other traders' response to the displayed path, measured against a path-excluded response counterfactual, and $K^{D}(P;S)$ is the cost of producing that displacement. The first inequality rejects the maintained execution frontier; the second says that the path becomes worthwhile only after induced response is counted. The envelope gives these inequalities content. It collects the legitimate motives into one maintained frontier and maximizes over them, so that a negative envelope is a restriction, not a failure of imagination: the best available execution story still loses money.

The contribution is not the observation that spoofing can move other traders. That mechanism is central in existing legal, theoretical, and agent-based work on spoofing. The contribution here is an identification criterion. I ask what an analyst must observe to distinguish a cancelled displayed path that is valuable through own execution, inventory, hedging, queue, or stale-quote motives from an observationally similar path that is valuable only because it induces other traders to react. The answer is the spoofing wedge: a negative upper bound on bona fide value together with positive response-augmented value and exposure that monetizes the response.

Why should a displayed order move anyone at all? It moves the market because, in equilibrium, displayed depth is correlated with the genuine motives that produce it, and rational responders learn from it. Spoofing lives off that correlation. It imitates the public signal that bona fide liquidity emits while privately carrying negative own-execution value.

The trader extracts value from an informational asset that bona fide liquidity providers create, and degrades that asset for later display. The response value $D(P; S)$ is not an accounting residual: it is the payoff from a shared credibility resource. The social cost of spoofing is the depreciation of that resource, not only the single bad response it provokes.

This reframing dictates the shape of the argument. The first result is a boundary result. Over a model class that leaves hidden exposure and execution-frontier primitives unrestricted, no classifier measurable only with respect to coarse order-path statistics can be uniformly valid. The result is not a substitute for structure; it explains why the rest of the paper imposes structure on the execution frontier, the response counterfactual, and signed exposure. Proposition 1 states this coarse-statistic non-identification result, and Proposition 2 extends the point to the public book: even the complete public message history does not identify the wedge unless the analyst restricts or observes the trader's hidden economic exposure. These results locate the missing information in the trader's execution frontier and signed position rather than in the shape of the path.

Where to look is the envelope. Definitions 2 and 3 define the execution frontier and the admissible class, Proposition 3 establishes the envelope as a well-posed maximization, and Lemma 1 supplies the discipline that keeps the construction honest: because inventory, hedge, queue, and risk-limit states are bounded and are already maximized over, a negative envelope cannot be undone by some unmodeled but admissible motive inside those maintained sets. The envelope is a declared frontier, not an open-ended list of excuses, and the paper is explicit that enlarging $\mathcal{B}$ weakens every rejection while shrinking it strengthens every rejection. Each classification is therefore conditional on, and conservative relative to, a stated admissible class.

The equilibrium analysis makes the wedge a choice a trader makes, not an artifact the analyst imposes. Under the mixture experiment in Theorem 1, response value is parasitic on the credibility of bona fide displayed liquidity and collapses as manipulation becomes prevalent, so spoofing is self-limiting in its own success. Definition 6 and Theorem 2 define

credibility as signal informativeness and characterize the unpriced depreciation that manipulation imposes on the book. Theorem 3 characterizes a manipulation region as a threshold in signed exposure, and Theorem 4 gives an existence result in which public-statistic non-identification survives threshold optimal-stopping rules.

The same apparatus yields a test, but the test is an audit-data design rather than a public-message classifier. Public book data can motivate, calibrate, and test auxiliary predictions, but individual-path classification requires evidence that can bound the trader's execution frontier, the path-excluded response gap, and signed exposure. Because point identification of a counterfactual is rarely available, the test is deliberately one of partial identification. Theorem 5 measures how much induced response a path requires before it turns profitable, and Theorem 6 states the classification: reject the path as spoofing-like only when every admissible execution rationale is ruled out and displacement-augmented value remains strictly positive across the maintained parameter set. In practice the test has three legs: a negative bona fide envelope, an abnormal induced response relative to matched controls, and signed exposure that profits from that response. When any leg cannot be bounded with the right sign and margin, the test returns an inconclusive verdict rather than a manipulation label. The classification is economic, not legal: it identifies the payoff component a path requires, not the state of mind of the trader who chose it. The argument closes on market design. Proposition 6 and Theorem 7 show that hard caps on message traffic discipline the surface statistic while leaving the wedge intact below the cap, so effective regulation must raise the cost of deceptive response value rather than the cost of cancellation.

The paper draws on three literatures and contributes to each. The limit-order-book literature explains why aggressive cancellation is rational liquidity supply and supplies the execution-facing motives that populate the envelope (Parlour, 1998; Foucault, 1999; Foucault, Kadan, and Kandel, 2005; Rosu, 2009). The market-manipulation literature studies the price, settlement, and predatory channels through which strategic traders profit from distorted prices or forced trades, and locates spoofing within a broader account of manipulation

as the exploitation of others' responses (Allen and Gale, 1992; Kumar and Seppi, 1992; Brunnermeier and Pedersen, 2005; Goldstein and Guembel, 2008; Putnins, 2012). The spoofing and surveillance literature documents the conduct, models how spurious orders sway agents who learn from the book, and develops empirical detectors (Lee, Eom, and Park, 2013; Wang and Wellman, 2017, 2021; Cartea, Jaimungal, and Wang, 2020; Fox, Glosten, and Guan, 2022; Do and Putnins, 2023; Commodity Futures Trading Commission, 2013). Fox, Glosten, and Guan explain why quote-driven spoofing can mislead liquidity suppliers and why the harm is not cancellation per se. This paper takes the next identification step. It does not model misleading quotes as the novel mechanism; it asks how to classify the source of value when two order paths have the same public footprint but different hidden economic exposures. The marginal objects are the bona fide value envelope, the path-excluded response gap, and the signed-exposure test that links the response to the trader's payoff.

The remainder of the paper proceeds as follows. Section I sets out the institutional identification problem and the role of coarse surveillance screens. Section II develops the model and the causal response object. Section III states the public-path non-identification boundary. Section IV constructs the bona fide value envelope, and Section V defines the wedge and develops the credibility mechanism, its externality, and the equilibrium manipulation region. Section VI turns the theory into an audit-data partial-identification design. Section VII studies threshold rules and market design, Section VIII uses an NSE-style order book as an institutional illustration, and Section IX gives the simulation design and empirical implications. Section X concludes. Proofs not given in the text, together with the threshold-equilibrium construction, are collected in the Appendix.

I. The Identification Problem in Electronic Markets

An electronic limit order book is an identification problem before it is anything else. Participants submit, revise, and cancel orders continuously in response to movements in the

last traded price, their place in the queue, the toxicity of recent flow, and their inventory and hedging needs, and the same visible footprint that looks suspicious in isolation is the ordinary residue of competent liquidity supply. The problem sharpens in high-frequency markets, where dense message traffic, low-latency revision, and automated market making are not aberrations but the mechanics of modern liquidity provision (Hendershott, Jones, and Menkveld, 2011; Hasbrouck and Saar, 2013; Menkveld, 2013; O'Hara, 2015), and where automated order-book dynamics can interact with episodes of market stress (Kirilenko et al., 2017). The analyst who sees only the message stream sees the shadow of two very different motives and cannot, from the shadow alone, tell them apart.

Faced with this, exchange surveillance leans on coarse, message-based screens as a first filter: order-to-trade ratios, modification counts, cancellation rates, order lifetimes, quoting imbalance, and persistent-noise indicators. Such features have demonstrated predictive content in empirical spoofing work and are useful for triage and infrastructure discipline (Lee, Eom, and Park, 2013; Do and Putnins, 2023). They are not, however, structural tests. A cancelled order may have carried real execution-option value when it was posted, and an order that looks unremarkable under message statistics may have no execution-facing rationale and may pay only because it bends other traders' beliefs. The economic content lives in the source of the order's value, not in the silhouette it leaves in the message log, which is what ties the spoofing problem to the broader theory of trade-based, predatory, and price-feedback manipulation (Allen and Gale, 1992; Kumar and Seppi, 1992; Brunnermeier and Pedersen, 2005; Goldstein and Guembel, 2008; Putnins, 2012).

A high-message exchange of the kind operated in many emerging and developed markets, with order-to-trade monitoring, persistent-modification concerns, and algorithmic surveillance, is a natural laboratory for this gap, and the National Stock Exchange of India serves as the running example below (Securities and Exchange Board of India, 2020; National Stock Exchange of India, 2020, 2026). The point, however, is broader. A venue that tries to infer manipulation from the shape of the order path confronts the same identification problem,

and a structural classification must reach past the path to the economics that generated it.

II. Model

The model is built to make one distinction precise: the difference between value a trader earns from own execution and value the trader earns from the reactions of others. It has three kinds of participant. A strategic trader displays orders and holds a possibly hidden economic position; a population of responders observes the public book and acts on what it infers; and an analyst, standing outside both, observes some information set and tries to classify the trader's conduct. Three separations organize everything that follows. The first is between what the trader displays and what the trader earns: the displayed path is the visible stream of orders, revisions, and cancellations, whereas the full economic path also contains the trader's real positions, which may sit in the same book, on the opposite side, in a correlated instrument, or in a later monetizing trade. The second is between two channels of value: execution-facing value would exist even if the book were unobserved, while response-facing value exists only because others see the displayed path and react to it. The third is between information sets: the trader's private state is hidden, responders condition on the public book, and the analyst lies in between, commanding richer evidence the deeper an audit reaches. The remainder of this section formalizes these separations and, in particular, constructs the response channel against an explicit path-excluded counterfactual, so that value earned from others' reactions is never confused with value earned from ordinary public information.

A. States, paths, and feasibility

Time is discrete, $t = 0, \ldots, T$, and discounted by $\beta \in (0, 1)$. The public limit-order-book state is

$$s_t = (p_t, d_t, Q_t, z_t),$$

where p_t is a price or best-quote state, d_t is displayed depth or imbalance, Q_t is queue state, and z_t collects public order-flow information. The strategic trader carries a private state

$$x_t = (q_t, h_t, \theta_t, \lambda_t, \ell_t),$$

in which q_t is inventory, h_t is hedge exposure, θ_t is a private trading motive or urgency state, λ_t is private execution-risk information, and ℓ_t is a surveillance or detection state. The full state is $S_t = (s_t, x_t) \in \mathcal{S}$.

At each date the trader chooses a displayed order vector

$$o_t = (\text{side}_t, \text{price}_t, \text{size}_t),$$

a cancellation or modification decision c_t, and a real economic trading decision y_t. The decision y_t collects executions in the same book, opposite-side trades, later monetizing trades, and positions in correlated instruments. Writing the displayed and full economic paths as

$$P^d = \big((o_t, c_t)\big)_{t=0}^{T} \qquad \text{and} \qquad P = \big((o_t, c_t, y_t)\big)_{t=0}^{T},$$

the separation between them is the crux of the model: a displayed order may be costly to have filled, yet the trader may still profit through y_t after others revise their quotes, withdraw liquidity, submit marketable orders, or reprice a correlated instrument. To keep the two roles of position distinct, y_t always denotes the real economic trading action, and the scalar Y is reserved for the signed monetization exposure that the y-path carries with respect to the

response gap; empirical data vectors are denoted separately so that Y never stands for an observable.

A path is feasible when it respects the mechanics of the book: an unposted order cannot be cancelled, an execution cannot exceed displayed size, and queue position evolves under the venue's priority rule. Let $\mathcal{P}(S)$ denote the set of feasible full economic paths from state S.

B. Responders and the causal response channel

Other market participants observe public order-book states and choose responses $r_t \in \mathcal{R}$, among them quote improvement, quote retreat, cancellation, marketable-order submission, queue migration, and cross-instrument repricing. Let $\vartheta_t \in \{H, L\}$ index a short-horizon value, order-flow-pressure, or adverse-selection state. Responders hold the posterior belief

$$\mu_t = \Pr(\vartheta_t = H \mid \mathcal{O}^t),$$

formed on the public order-book history $\mathcal{O}^t$, which includes displayed depth and the strategic trader's displayed path when visible, and a representative responder solves

$$r_t \in \arg\max_{r \in \mathcal{R}} U_R(r, \mu_t, Q_t, p_t).$$

Displayed depth and order paths update beliefs through the Bayesian operator

$$\mu_{t+1} = \Psi(\mu_t, d_t, Q_t, z_t, P_t^d),$$

and the induced equilibrium response path is written $R(P; S) = \{r_t(P; S)\}_{t=0}^{T}$.

To isolate the causal response to the suspect displayed path, define a path-excluded information intervention. Let $\mathcal{I}^t$ be the public information set observed by responders, including ordinary public order-flow information and, when visible, the suspect displayed path

P^d. Let $\mathcal{I}^{-P,t}$ be the information set obtained by removing the suspect displayed messages from $\mathcal{I}^t$ while holding fixed all non-suspect public information, including marketwide order flow, public trades, news, and other displayed depth. Equivalently, $\mathcal{I}^{-P,t}$ is a masking or placebo intervention in which the suspect path is not available as an informative signal. The path-excluded response is

$$R^0(P; S) = R(\mathcal{I}^{-P}; S),$$

and the observed response is

$$R(P; S) = R(\mathcal{I}; S).$$

The induced-response payoff is

$$D(P; S) = \Pi(P, R(P; S); S) - \Pi(P, R^0(P; S); S), \tag{1}$$

where $\Pi(P, R; S)$ is the trader's payoff from path P under response path R. Thus $D(P; S)$ measures the payoff effect of allowing the suspect displayed path to enter responders' information sets, holding fixed ordinary public information. Responders need not be irrational or mechanical: they update optimally from the information they observe. The counterfactual changes the information experiment, not the rationality of the responders.

The pieces of the model have familiar antecedents. The adverse-selection component is in the spirit of information-based microstructure (Kyle, 1985; Glosten and Milgrom, 1985); the order-placement component follows the limit-order-market literature in which execution probability, waiting costs, and queue conditions are endogenous state variables rather than background frictions (Parlour, 1998; Foucault, 1999; Foucault, Kadan, and Kandel, 2005; Rosu, 2009); and the response channel formalizes the insight of agent-based spoofing models that displayed orders move traders who learn from the book (Wang and Wellman, 2017, 2021). What is new is the decomposition of payoff into an execution-facing part and the response-facing part (1).

C. A tractable responder block

For the equilibrium results, specialize the representative responder's problem as follows. The responder observes a signal m derived from displayed depth and forms posterior $\mu = \Pr(\vartheta = H \mid m)$. The response action r is a scalar summary of quote improvement, liquidity retreat, marketable-order submission, or cross-instrument repricing in the direction favored by state H. Let

$$U_R(r, \mu, Q, p) = r\left[a_L(Q,p) + (a_H(Q,p) - a_L(Q,p))\mu\right] - \frac{\kappa_R(Q,p)}{2} r^2,$$

where $a_H(Q,p) > a_L(Q,p)$ and $\kappa_R(Q,p) > 0$. The unique optimal response is

$$r^*(\mu, Q, p) = \frac{a_L(Q,p) + (a_H(Q,p) - a_L(Q,p))\mu}{\kappa_R(Q,p)},$$

which is strictly increasing in μ. For notational economy, suppress Q and p when they are fixed over the event window and write $r^*(\mu)$.

A trader with signed monetization exposure Y earns response payoff

$$D(m; Y, \rho) = Y\left[r^*(\mu(m;\rho)) - r^*(\mu^0)\right],$$

where $\mu(m;\rho)$ is the posterior when the suspect displayed signal is observed and μ^0 is the posterior under the path-excluded information set. In this special case, the reduced-form response-gain function used below is not primitive; it is induced by responder optimization:

$$G(\mu) = r^*(\mu) - r^*(\mu^0).$$

The signed exposure Y is observed by the trader and, in an audit design, by the analyst. It is not observed by responders in real time. If responders observed Y, their optimal response would condition directly on the possibility of deceptive monetization.

D. Maintained assumptions

ASSUMPTION 1 (Regular state and path spaces): The state space $\mathcal{S}$, the private inventory, hedge, and risk-limit sets, and the motive set $\mathcal{B}$ are compact. For every S, the feasible path set $\mathcal{P}(S)$ is nonempty, and the payoff primitives introduced below are bounded and continuous in the relevant state, path, and motive arguments.

ASSUMPTION 2 (Coarse observability): The coarse statistic

$$X(P) = \big(\text{size}(P), \text{lifetime}(P), \text{cancelRate}(P), \text{modifyCount}(P), \text{OTR}(P), \text{distance}(P)\big)$$

is measurable with respect to displayed order messages and public order-path attributes, such as size, lifetime, cancellation rate, modification count, order-to-trade ratio, and distance from the touch. It does not include the trader's hidden inventory, hedge book, cross-instrument exposure, beneficial ownership, private execution-risk information, risk-limit state, or real economic trading path y_t. The maintained model class permits more than one feasible hidden-state completion of the same displayed statistic.

ASSUMPTION 3 (Execution-facing admissible class): The admissible bona fide class $\mathcal{B}$ contains only motives whose value comes from own execution, execution optionality, inventory reduction, hedge reduction, stale-quote protection, adverse-selection avoidance, queue value, or risk-limit compliance. Any payoff component that requires changing other agents' beliefs or actions without own execution-facing value enters D, not $\mathcal{B}$.

ASSUMPTION 4 (Endogenous response and credibility): There exist public states and displayed paths at which displayed depth is informative, because bona fide liquidity sometimes carries positive execution-frontier value. A change in displayed depth or imbalance can therefore move posterior beliefs μ_t and the response r_t, and the resulting causal response value $D(P; S)$ is bounded and may be positive even when the displayed path has negative

own-execution value.

Assumption 5 (Outside-option normalization): The value of inaction over the event window is normalized to zero, so a path is chosen only when its total continuation value is weakly positive relative to that outside option.

Given a response rule, the trader solves the Bellman problem

$$V(S) = \max_{(o,c,y)} \Big\{ \pi^B(S,o,c,y,R^0) + D(S,o,c,y) \\ - K^B(S,o,c,y) - K^D(S,o,c,y) + \beta\, \mathbb{E}[V(S') \mid S,o,c,y,R] \Big\}, \tag{2}$$

in which the execution-facing return π^B and its cost K^B are evaluated under the counterfactual response R^0, while continuation is taken under the actual response R. The decomposition in (2) is the formal home of the paper's thesis: the term π^B records what the path is worth on the trader's own frontier, and the term D records what it is worth through the reactions of others.

III. Public-Path Non-Identification

The first result formalizes a boundary on surveillance that reads only the shape of the order path. The claim is not that path statistics are uninformative in any particular sample. It is that, over a model class with unrestricted hidden exposure and execution-frontier primitives, a classifier based only on those statistics is not uniformly valid. To state it, let P denote an order path and let

$$X(P) = \big(\text{size}(P), \text{lifetime}(P), \text{cancelRate}(P), \text{modifyCount}(P), \text{OTR}(P), \text{distance}(P)\big)$$

be the coarse path-statistic vector available to a message-based screen. A surveillance rule that uses only $X(P)$ is measurable with respect to

$$\mathcal{G}_X = \sigma(X(P)),$$

and the question is whether any such rule can recover the structural classification.

Definition 1 (Pattern classifier): A pattern classifier is a measurable map

$$\phi_X : \text{range}(X) \to \{B, M\},$$

where B denotes bona fide and M denotes manipulation-like. Equivalently, the induced classifier $\phi_X(P) = \phi_X(X(P))$ is $\mathcal{G}_X$-measurable.

Proposition 1 (Coarse-statistic non-identification over unrestricted hidden exposure): *Under Assumptions 1, 2, 3, and 4, suppose the market contains both execution-facing cancellation motives and response-facing manipulation motives. Let $Z \in \{B, M\}$ denote the structural source-of-value classification. Then there exist two structural economies, with probability laws $\mathbb{P}_B$ and $\mathbb{P}_M$, such that*

$$\mathcal{L}_{\mathbb{P}_B}(X) = \mathcal{L}_{\mathbb{P}_M}(X),$$

but

$$\mathbb{P}_B(Z = B) = 1 \quad \textit{and} \quad \mathbb{P}_M(Z = M) = 1.$$

Thus structural spoofing is not uniformly identified over the two economies by any $\mathcal{G}_X$-measurable classifier. In particular, for any pattern classifier ϕ_X,

$$\max\left\{\mathbb{P}_B(\phi_X(X) \neq Z), \mathbb{P}_M(\phi_X(X) \neq Z)\right\} \geq \frac{1}{2},$$

under the common distribution of X. Hence no classifier measurable with respect to X is uniformly consistent over the two-economy class.

Proof. Let ν be any probability distribution over feasible coarse statistics x with support on displayed paths that involve order entry, modification, cancellation, and nontrivial order-to-trade ratios. By many-to-one observability, each x in the support can be generated by more than one feasible full economic path and private state.

In the first economy, $\mathbb{P}_B$, assign to each x a feasible path-state pair $(P^B(x), S^B(x))$ generated by stale-quote protection, queue-position value, inventory-risk reduction, hedge-risk reduction, or execution-risk management. Choose the bounded execution-frontier primitives so that

$$\overline{V}^B(P^B(x); S^B(x)) \geq 0$$

for ν-almost every x. Then $Z = B$ almost surely under $\mathbb{P}_B$.

In the second economy, $\mathbb{P}_M$, assign the same distribution ν to coarse statistics, but generate each x from a feasible path-state pair $(P^M(x), S^M(x))$ with the same displayed statistics and negative own-execution value:

$$\overline{V}^B(P^M(x); S^M(x)) < 0.$$

Let the displayed path induce belief updating or action displacement by other traders–quote improvement, liquidity retreat, cancellation, market-order submission, queue migration, or cross-instrument repricing–that the trader monetizes through y_t. Choose response primitives so that

$$D(P^M(x); S^M(x)) - K^D(P^M(x); S^M(x)) > -\overline{V}^B(P^M(x); S^M(x))$$

for ν-almost every x. Then $Z = M$ almost surely under $\mathbb{P}_M$.

By construction, the marginal law of X is ν in both economies, so $\mathcal{L}_{\mathbb{P}_B}(X) = \mathcal{L}_{\mathbb{P}_M}(X)$. Any $\mathcal{G}_X$-measurable classifier is a function only of X. Its accuracy under the two economies

is therefore

$$\int \mathbf{1}\{\phi_X(x) = B\}\, d\nu(x) + \int \mathbf{1}\{\phi_X(x) = M\}\, d\nu(x) = 1.$$

Thus

$$\max\left\{\mathbb{P}_B(\phi_X(X) \neq Z), \mathbb{P}_M(\phi_X(X) \neq Z)\right\} \geq \frac{1}{2}.$$

The missing information is not a better cutoff inside X; it is the structural source of continuation value. ■

The proposition is intentionally a model-class statement. It does not say that order-path statistics are useless in a particular sample. It says that their identifying content depends on restrictions outside the statistic itself. The rest of the paper supplies such restrictions through a declared execution-facing frontier, a response counterfactual, and exposure evidence.

Corollary 1 (No sufficient cancellation statistic): *No statistic built only from order size, order lifetime, cancellation frequency, modification count, order-to-trade ratio, or distance from the touch is a sufficient statistic for structural spoofing whenever execution-facing cancellation and response-facing manipulation can generate the same statistic.*

The same logic extends, with more force, to the richest public information an analyst could hope to observe. One might suspect that the screen fails only because it discards detail, and that the full public message stream would suffice. It does not. Let

$$H^{\text{book}}(P) = \left((s_t, P_t^d, \text{public trades}_t, \text{public responses}_t)\right)_{t=0}^{T}$$

denote the public book history generated by the displayed path, public trades, and publicly visible responses. Let

$$\mathcal{G}_{\text{book}} = \sigma(H^{\text{book}}(P)).$$

This sigma-field is richer than $\mathcal{G}_X$. It may contain the complete displayed message stream, best quotes, depth, public trades, and responses by other traders. It still excludes the trader's

hidden inventory, hedge book, beneficial ownership, cross-instrument exposure, risk-limit state, private execution-risk information, and real economic trading y_t.

PROPOSITION 2 (Public-book non-identification without exposure restrictions): *Fix a feasible public book history h. Suppose the maintained admissible set does not rule out either of the following two full economic completions of h:*

(i) a bona fide completion (P^B, S^B) with the displayed history h and

$$\overline{V}^B(P^B; S^B) \geq 0;$$

(ii) a response-facing completion (P^M, S^M) with the same displayed history h, the same public response path, and hidden economic exposure satisfying

$$\overline{V}^B(P^M; S^M) < 0 \quad \textit{and} \quad \overline{V}^B(P^M; S^M) + D(P^M; S^M) - K^D(P^M; S^M) > 0.$$

Then structural spoofing is not identified by $\mathcal{G}_{\text{book}}$ alone. In particular, any $\mathcal{G}_{\text{book}}$-measurable classifier assigns the same label to the two completions and therefore cannot be correct on both.

Proof. Both completions induce the same public book history h, so every $\mathcal{G}_{\text{book}}$-measurable classifier must take the same value on them. The first completion is bona fide because its own-execution envelope is nonnegative. The second completion is structurally spoofing-like because own-execution value is negative but displacement-augmented value is positive. Thus the structural label differs while the public-book observation is identical. A public-book classifier cannot distinguish the two. ■

Proposition 2 is deliberately conditional. It is not an absolute impossibility theorem. It says that public-book data alone are not a sufficient statistic when hidden economic exposure is unrestricted. The response-facing completion must still be feasible and must still satisfy

the wedge inequalities; the missing object is the exposure sign and magnitude needed to rule that completion out. Account-level audit trails, beneficial-ownership data, cross-instrument positions, hedge records, or maintained restrictions on feasible exposure can shrink the set of completions and turn the public-book history into useful structural evidence.

Table 1 summarizes the identification ladder. Let $\mathcal{G}_A$ denote an audit-and-exposure information set that augments $\mathcal{G}_{\text{book}}$ with account-level exposure records, beneficial ownership, cross-instrument positions, hedge records, and risk-limit information. Structural classification lives at the top of this ladder, where exposure evidence is combined with the response-gap design.

Information set	What it contains	Structural role
$\mathcal{G}_X$	Coarse path statistics: size, lifetime, cancellation frequency, modification count, OTR, and distance from the touch.	Useful for triage and message discipline; does not by itself identify the source of continuation value.
$\mathcal{G}_{\text{book}}$	Full public book history: displayed messages, public trades, best quotes, depth, and visible responses.	Identifies public response moments; still misses hidden exposure and the trader's own execution frontier.
$\mathcal{G}_A$	Public book plus audit-trail and exposure records: account ownership, inventory, hedges, correlated positions, and risk limits.	Can bound bona fide value and signed monetization exposure Y.
$\mathcal{G}_S$	Audit-and-exposure evidence plus matched counterfactual responses, markouts, and response-gap estimates.	Supports structural classification: no execution rationale, abnormal response, profitable exposure.

Table 1: Information-set ladder. Moving from path shape to structural classification requires hidden exposure evidence and a response-gap design. Public-book data alone are not the final identifying object when hidden economic exposure is unrestricted.

The result is not a claim that pattern screens are useless. Pattern screens can be useful for triage, message discipline, or prioritizing investigation. Proposition 1 says that the source of continuation value is not identified by cancellation-space observables alone, even distributionally. Proposition 2 extends the point: even a full public book history does not

identify the wedge without restrictions on hidden economic exposure.

Both results point to the same missing object. What the observables fail to reveal is not a sharper feature of the path but the trader's own execution frontier: the set of payoffs a path could earn from fills, queue position, inventory, hedges, stale-quote exposure, and risk limits, before any other trader reacts. Identification must therefore stop asking what the path looks like and start asking what the path could be worth on that frontier. Section IV constructs the envelope that answers this question and converts non-identification into a usable decision boundary.

IV. The Bona Fide Value Envelope

Non-identification leaves a precise gap: the source of continuation value cannot be read off the path. This section fills the gap with one object. The idea is to evaluate a displayed path against the most favorable execution-facing story available to the trader and ask whether even that story makes the path worth holding. If it does, the path is rationalizable without anyone else reacting, and there is nothing to explain. If it does not, the path's value must come from outside the declared execution frontier; in the maintained decomposition, that remaining channel is the response of other traders. The envelope makes "the most favorable execution-facing story" a well-defined supremum rather than a rhetorical concession, which is what allows a negative envelope to function as a restriction rather than a relabeling.

Definition 2 (Execution frontier): The trader's execution frontier is the set of payoffs generated by the trader's own fills, fill probabilities, inventory exposure, hedge exposure, queue position, stale-quote exposure, adverse-selection exposure, and risk-limit constraints. These payoffs may depend on public and private state, but they do not require other traders to update from the suspect displayed path.

Definition 3 (Admissible bona fide class): The class $\mathcal{B}$ contains policies or motives b whose

payoff is generated only by the execution frontier. For each $b \in \mathcal{B}$, define

$$V_b^B(P;S) = \mathbb{E}\left[\sum_{t=0}^{T} \beta^t \left(\mathrm{CF}_t^b + \mathrm{IR}_t^b + \mathrm{HR}_t^b + \mathrm{QO}_t^b + \mathrm{AS}_t^b + \mathrm{RL}_t^b - \mathrm{MC}_t^b\right) \middle| P, S, R^0(P;S)\right].$$

Here $R^0(P;S)$ is the path-excluded response from Section II. The term CF_t^b is cash-flow value from fills, IR_t^b is inventory-risk reduction, HR_t^b is hedge-risk reduction, QO_t^b is queue-option value, AS_t^b is adverse-selection or stale-quote protection, RL_t^b is risk-limit value, and MC_t^b is message, monitoring, and execution cost. A motive is not admissible if its positive value requires induced belief distortion, induced liquidity retreat, quote improvement by others, market-order submission by others, detector gaming, or cross-market monetization of a false displayed signal.

The admissibility restriction is therefore this: if $b \in \mathcal{B}$, then $V_b^B(P;S)$ may depend only on own fills, own risk, own hedge, own queue position, and own execution option value. It may not depend positively on other traders being induced to cancel, retreat, improve quotes, submit market orders, migrate queues, or revise beliefs.

Define the bona fide value envelope:

$$\overline{V}^B(P;S) = \sup_{b \in \mathcal{B}} V_b^B(P;S).$$

The envelope is a maintained admissible frontier, not a maximization over every imaginable story. Enlarging $\mathcal{B}$ weakens the rejection of bona fide rationalizability; shrinking $\mathcal{B}$ strengthens it. The paper therefore makes a conditional claim: relative to a declared, bounded, execution-facing class, a negative envelope means the path cannot be rationalized by own-execution motives inside that class. It does not say that unconstrained private explanations are impossible.

Definition 4 (Bona fide rationalizability): A path P is bona fide rationalizable in state S if $\overline{V}^B(P;S) \geq 0$. It is not bona fide rationalizable relative to $\mathcal{B}$ if $\overline{V}^B(P;S) < 0$.

PROPOSITION 3 (Bona fide envelope): *If $\mathcal{B}$ is compact and $V_b^B(P;S)$ is continuous in b, then the supremum is attained. A path is not bona fide rationalizable relative to $\mathcal{B}$ if and only if*

$$\max_{b\in\mathcal{B}} V_b^B(P;S) < 0.$$

Proof. Compactness and continuity imply the maximum exists. If the maximum is non-negative, the maximizing motive rationalizes the path. If the maximum is negative, every admissible execution-facing motive gives negative value. ■

LEMMA 1 (Private-state discipline): *Let private inventory, hedge, execution-risk, queue, and risk-limit states lie in bounded admissible sets. If $\overline{V}^B(P;S) < 0$, then no admissible bounded private-state explanation in $\mathcal{B}$ rationalizes P. Any rationalization of P must therefore use either a payoff outside the execution frontier or a state outside the maintained admissible set.*

Proof. The envelope maximizes over all admissible motives and bounded private states included in $\mathcal{B}$. A negative maximum means every such execution-frontier explanation has negative value. A remaining positive rationalization must add a payoff not in $\mathcal{B}$ or change the maintained state space. ■

This discipline is what keeps the theory from collapsing into either of its degenerate forms. It does not assert that no private explanation could ever exist; it asserts that the standard private explanations, inventory, hedge, queue, stale-quote, and risk-limit motives, inhabit bounded admissible sets that the envelope has already searched. When the envelope remains negative, the path lies outside the bona fide frontier, and whatever residual value it carries must come from somewhere the frontier does not reach. Any payoff that depends on inducing other traders to cancel, retreat, improve, submit marketable orders, migrate queues, or revise beliefs, without own execution value, is by definition assigned to the response term $D(P;S)$. The next section studies precisely the region in which that response value is not incidental but necessary for the path to be worth choosing.

V. The Manipulation Wedge

With the envelope in hand, spoofing can be defined as an economic event rather than a behavioral one. The defining feature is not that the trader cancels, nor that cancellation is fast, but that the chosen path is unprofitable on the trader's own frontier and is rescued by the reaction of others. This section gives that event a name, derives the response value created by credible displayed liquidity, and traces the consequences of that dependence: a dynamic externality on the informativeness of the book, and an equilibrium region in which traders with sufficient exposure choose to spoof.

Definition 5 (Manipulation wedge): A path P has a manipulation wedge in state S if

$$\overline{V}^B(P;S) < 0 \quad \text{and} \quad \overline{V}^B(P;S) + D(P;S) - K^D(P;S) > 0.$$

Proposition 4 (Manipulation wedge): *If P has a manipulation wedge, then the path is loss-making under every admissible execution-facing motive but profitable after induced-response value is included. Hence P is structurally spoofing-like.*

Proof. The first inequality rules out bona fide rationalizability. The second inequality states that total value is positive after displacement value net of displacement cost is included. Thus the path is rational only through induced-response value. ■

Figure 1 gives the basic economic decomposition. The displayed path is first evaluated against the most favorable admissible own-execution rationale. If that envelope is negative, the path crosses into the spoofing-like region only when induced-response value more than offsets both the own-execution loss and displacement cost.

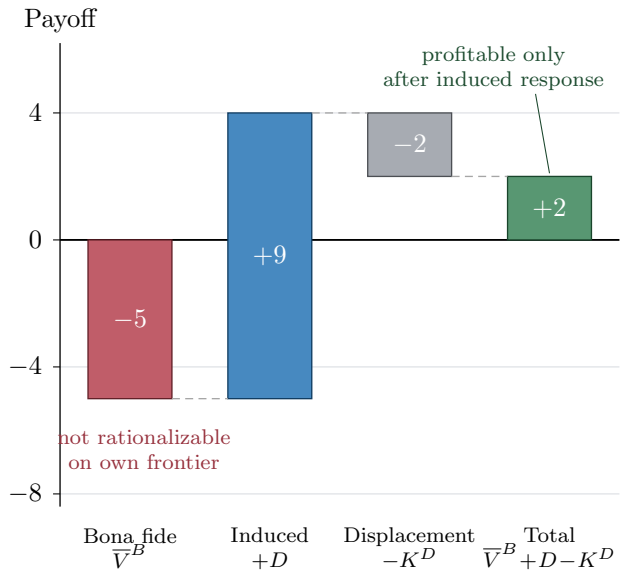

Figure 1: The spoofing wedge. The figure decomposes a path with negative own-execution value, $\overline{V}^B < 0$, positive induced-response value D, displacement cost K^D, and positive total value. The numbers ($\overline{V}^B = -5$, $D = +9$, $K^D = 2$, total $= +2$) are the running illustration of Section VIII.

Bona fide liquidity moves beliefs too; the claim is not that legitimate orders go unnoticed. The distinction is one of dependence. A bona fide path retains nonnegative value when the response term D is stripped away, because its worth rests on the trader's own execution and risk reduction. A spoofing-like path loses that rationalization when the response term is removed. The wedge is the set of paths for which induced response is load-bearing.

Lemma 2 (Credibility necessity): *If $R(P;S) = R^0(P;S)$, then $D(P;S) = 0$. If in addition $\overline{V}^B(P;S) < 0$ and $K^D(P;S) \geq 0$, then P cannot be profitable relative to the normalized outside option.*

Proof. By the definition of the causal response term,

$$D(P;S) = \Pi(P, R(P;S); S) - \Pi(P, R^0(P;S); S).$$

If $R(P;S) = R^0(P;S)$, then $D(P;S) = 0$. Total value is therefore

$$\overline{V}^B(P;S) - K^D(P;S) < 0$$

whenever $\overline{V}^B(P;S) < 0$ and $K^D(P;S) \geq 0$. With the outside option normalized to zero, the path is not profitable. The path can be rational only if it changes other traders' response. ■

THEOREM 1 (Responder learning and parasitic response value): *Let $\mathcal{M} = [\underline{m}, \overline{m}] \subset \mathbb{R}_+$ be a compact displayed-depth interval. Let $\vartheta \in \{H, L\}$ be a short-horizon value, demand, or adverse-selection state. Bona fide displayed depth has conditional densities f_H and f_L on $\mathcal{M}$. Suppose f_H, f_L, and g are continuously differentiable and bounded away from zero on $\mathcal{M}$. Let*

$$\ell(m) = \frac{f_H(m)}{f_L(m)}$$

have strict monotone likelihood ratio on $\mathcal{M}$: there is $\kappa_\ell > 0$ such that $\ell'(m) \geq \kappa_\ell$ for all $m \in \mathcal{M}$. A fraction $\rho \in [0,1]$ of displayed depth is manipulation-driven and drawn from density $g(m)$, independent of ϑ. The observed conditional density is

$$h_\vartheta(m;\rho) = (1-\rho) f_\vartheta(m) + \rho g(m).$$

Responders form posterior

$$\mu(m;\rho) = \frac{\pi h_H(m;\rho)}{\pi h_H(m;\rho) + (1-\pi) h_L(m;\rho)}$$

and choose the optimal response $r^*(\mu)$ from the quadratic payoff in Section II. A trader with signed exposure $Y > 0$ earns response payoff

$$D(m;Y,\rho) = Y\left[r^*(\mu(m;\rho)) - r^*(\mu^0)\right],$$

where μ^0 is the posterior under the path-excluded information set. Then there is $\bar{\rho} > 0$ such

that, for all $m \in \mathcal{M}$,

$$\frac{\partial D(m; Y, \rho)}{\partial m} > 0 \quad \textit{for all } \rho < \bar{\rho}.$$

If manipulation-driven depth becomes fully prevalent and uninformative, then

$$\lim_{\rho \to 1} \mu(m; \rho) = \pi \quad \textit{and} \quad \lim_{\rho \to 1} \frac{\partial D(m; Y, \rho)}{\partial m} = 0.$$

Hence, in this mixture experiment, manipulation-driven display has response value because bona fide display makes depth informative.

Proof. When $\rho = 0$, the posterior odds are

$$\frac{\mu(m; 0)}{1 - \mu(m; 0)} = \frac{\pi}{1 - \pi} \ell(m).$$

Since $\ell'(m) \geq \kappa_\ell > 0$ and the densities are bounded away from zero, $\partial \mu(m; 0)/\partial m$ is strictly positive on $\mathcal{M}$. The derivative $\partial \mu(m; \rho)/\partial m$ is continuous in (m, ρ) on the compact set $\mathcal{M} \times [0, 1]$. Therefore strict positivity at $\rho = 0$, uniformly on $\mathcal{M}$, implies that there is $\bar{\rho} > 0$ such that $\partial \mu(m; \rho)/\partial m > 0$ for all $m \in \mathcal{M}$ and all $\rho < \bar{\rho}$. Since $Y > 0$ and $\partial r^*/\partial \mu = (a_H - a_L)/\kappa_R > 0$, the response payoff is also strictly increasing in displayed depth on this region.

When $\rho \to 1$, both conditional densities converge uniformly on $\mathcal{M}$ to the same uninformative density $g(m)$. Thus $h_H(m; \rho)/h_L(m; \rho) \to 1$, and Bayes' rule gives $\mu(m; \rho) \to \pi$. The posterior no longer changes with displayed depth, so $\partial \mu(m; \rho)/\partial m \to 0$. Since $\partial r^*/\partial \mu$ is finite on the posterior range, the marginal response payoff converges to zero. ■

Figure 2 plots the same mechanism. Manipulation is most valuable when displayed depth remains credible because bona fide depth still carries information. As manipulation becomes prevalent, the marginal response value of displayed depth collapses.

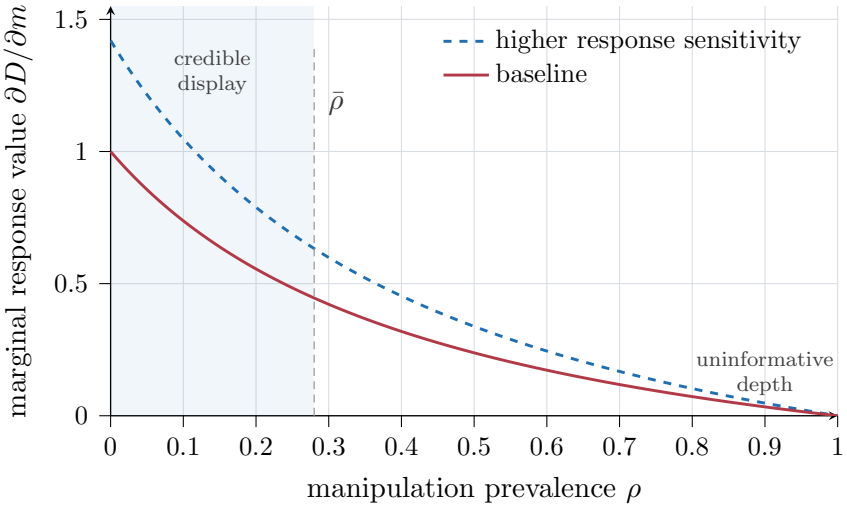

Figure 2: Parasitic credibility. The figure plots the marginal response value of displayed depth, $\partial D/\partial m$, in the mixture experiment. At low manipulation prevalence ρ, displayed depth remains informative; as manipulation-driven display becomes prevalent, the marginal response value falls toward zero. Greater response sensitivity (dashed) raises the curve but does not change its collapse as $\rho \to 1$.

Definition 6 (Display credibility): Display credibility is the informativeness of displayed depth about the latent short-horizon state. In the mixture experiment of Theorem 1, define

$$C(\rho) = I(\vartheta; M_\rho),$$

where M_ρ is the displayed-depth signal with conditional density $h_\vartheta(m; \rho)$. Higher C means that displayed depth carries more information for responders. Because manipulation-driven display is drawn from a state-independent density, increasing manipulation prevalence garbles the displayed-depth experiment and weakly lowers C.

For dynamic comparative statics, it is useful to approximate the law of motion by using L_t^B for bona fide displayed-liquidity supply and L_t^M for aggregate manipulation-driven displayed depth:

$$C_{t+1} = (1 - \delta)C_t + \gamma L_t^B - \chi L_t^M,$$

which is a local linearization of the learning-based credibility stock around a steady state.

The externality result below uses the structural sign restriction behind this approximation: manipulation raises the prevalence of state-independent display and therefore weakly lowers the informativeness of displayed depth.

THEOREM 2 (Credibility externality from signal garbling): *Let manipulation intensity* $m \in [0, \overline{m}]$ *raise the next-period prevalence of state-independent displayed-depth noise, so that* $\rho^+(m)$ *is increasing in* m*. Let display credibility be* $C(\rho) = I(\vartheta; M_\rho)$*, with* $C_\rho(\rho) \leq 0$*. The private payoff from manipulation is*

$$W^M(m; C) = D(m; S, C) - A(m) - K^D(m).$$

Here $D(m; S, C)$ *is the response payoff from the same responder block, with signed exposure collected in* S *and the displayed-depth experiment summarized by credibility* C*. The social objective is*

$$SW(m; C) = W^M(m; C) + \Gamma(C(\rho^+(m))),$$

where $\Gamma_C > 0$*. Assume all functions are continuously differentiable,* $\rho_m^+(m) \geq 0$*, and* $W^M(\cdot; C)$ *and* $SW(\cdot; C)$ *are concave in* m*. If the private optimum is interior, then it satisfies*

$$D_m(m; S, C) = A_m(m) + K_m^D(m).$$

If the social optimum is interior, then it satisfies

$$D_m(m; S, C) = A_m(m) + K_m^D(m) - \Gamma_C(C(\rho^+(m)))\, C_\rho(\rho^+(m))\, \rho_m^+(m).$$

Since $C_\rho \leq 0$ *and* $\rho_m^+ \geq 0$*, the extra social term is nonnegative. Private manipulation is weakly above the social level, strictly so when credibility is valuable and manipulation strictly garbles the signal.*

Proof. The private first-order condition is obtained by differentiating $W^M(m; C)$ with respect

to m. The social objective adds the continuation value of credibility. Since

$$\frac{\partial}{\partial m}\Gamma(C(\rho^+(m))) = \Gamma_C(C(\rho^+(m)))C_\rho(\rho^+(m))\rho_m^+(m),$$

the social first-order condition is

$$D_m(m;S,C) - A_m(m) - K_m^D(m) + \Gamma_C(C(\rho^+(m)))C_\rho(\rho^+(m))\rho_m^+(m) = 0,$$

which gives the stated expression. The additional term is positive whenever manipulation reduces credibility and credibility has positive continuation value. At an interior private optimum m^M, the private marginal payoff is zero. The social marginal payoff is

$$\frac{\partial SW(m^M;C)}{\partial m} = \Gamma_C(C(\rho^+(m^M)))C_\rho(\rho^+(m^M))\rho_m^+(m^M) \leq 0.$$

Concavity of $SW(\cdot;C)$ implies that a social optimum cannot lie to the right of m^M. If the inequality is strict and the social optimum is interior, it lies strictly to the left. ■

The credibility externality is the missing term in the private problem. The private trader extracts response value from an informational asset created by bona fide displayed liquidity, but does not internalize the loss of future order-book informativeness.

For small manipulation intensity around $m = 0$, the credibility component of deadweight loss has the first-order approximation

$$DWL(m) = \Gamma_C(C(\rho^+(0)))\,|C_\rho(\rho^+(0))|\,\rho_m^+(0)\,m + o(m).$$

If the private and social objectives are twice differentiable and locally concave, then the private-social manipulation gap is approximately

$$m^M - m^S \approx \frac{\Gamma_C(C(\rho^+(m^M)))\,|C_\rho(\rho^+(m^M))|\,\rho_m^+(m^M)}{-SW_{mm}(m^M;C)}.$$

THEOREM 3 (Equilibrium manipulation region): *Let the spoofing-depth action set be a nonempty compact interval* $\mathcal{M} = [\underline{m}, \overline{m}] \subset (0, \infty)$*, with inaction available separately at value zero. A trader with signed monetization exposure* Y *chooses displayed depth* $m \in \mathcal{M}$*. Suppose own-execution value from the displayed path is negative and total payoff is*

$$W^M(m; Y, \rho) = YG(\mu(m; \rho)) - \Phi(m),$$

where $G(\mu) = r^*(\mu) - r^*(\mu^0)$ *is the response gain generated by responder optimization and* $\Phi(m) = A(m) + c(m) + K^D(m)$ *is the total cost of the displayed path. This is the exposure-explicit version of the private payoff* $W^M(m; C)$*, with credibility determined by the manipulation prevalence* ρ*. Assume* Φ *is continuous and strictly positive on* $\mathcal{M}$*, and* $G(\mu(m; \rho))$ *is continuous and strictly positive on* $\mathcal{M} \times [0, 1]$*. Define the exposure threshold*

$$\bar{Y}(\rho) = \min_{m \in \mathcal{M}} \frac{\Phi(m)}{G(\mu(m; \rho))}.$$

Then the trader strictly prefers some spoofing-like displayed path to inaction if and only if $Y > \bar{Y}(\rho)$*; at equality the trader is indifferent. If* $Y \sim F$ *has a continuous distribution on compact support, an equilibrium manipulation prevalence* ρ^* *exists and satisfies*

$$\rho^* = 1 - F(\bar{Y}(\rho^*)).$$

If $G(\mu(m; \rho))$ *is decreasing in* ρ*, then credibility erosion weakly raises* $\bar{Y}(\rho)$*. Pointwise increases in accidental execution cost, message cost, or detection cost weakly raise* $\bar{Y}(\rho)$*. Pointwise increases in response sensitivity weakly lower* $\bar{Y}(\rho)$*.*

Proof. For fixed ρ, the trader chooses a spoofing-like displayed depth if and only if there exists $m \in \mathcal{M}$ such that $YG(\mu(m; \rho)) > \Phi(m)$. This is equivalent to

$$Y > \min_{m \in \mathcal{M}} \frac{\Phi(m)}{G(\mu(m; \rho))} = \bar{Y}(\rho).$$

The minimum is attained because the ratio is continuous on compact $\mathcal{M}$, and $\bar{Y}(\rho)$ is continuous by the maximum theorem. Aggregate manipulation prevalence is therefore the continuous map $T(\rho) = 1 - F(\bar{Y}(\rho))$ from $[0,1]$ into itself. Let $H(\rho) = T(\rho) - \rho$. Since $H(0) = T(0) \geq 0$ and $H(1) = T(1) - 1 \leq 0$, the intermediate value theorem gives a fixed point $\rho^* = T(\rho^*)$.

If $G(\mu(m;\rho))$ decreases in ρ, then the ratio defining $\bar{Y}(\rho)$ rises pointwise, so the threshold weakly rises. Increasing any component of $\Phi(m)$ also raises the ratio pointwise. Increasing response sensitivity raises the denominator pointwise and therefore weakly lowers the threshold. ■

The fixed point in Theorem 3 has a transparent reading, shown in Figure 3. Each trader spoofs only when the trader's signed exposure clears the threshold $\bar{Y}(\rho)$, so the prevalence of manipulation that the population best-responds to is $T(\rho) = 1 - F(\bar{Y}(\rho))$, and equilibrium prevalence solves $\rho^* = T(\rho^*)$. The theorem only requires existence. In the single-crossing case drawn in the figure, credibility erosion raises the threshold as ρ rises, T slopes down, and the equilibrium is unique. The comparative statics act through this curve: anything that raises the cost of displacement or dampens the response of others shifts T inward and lowers equilibrium prevalence.

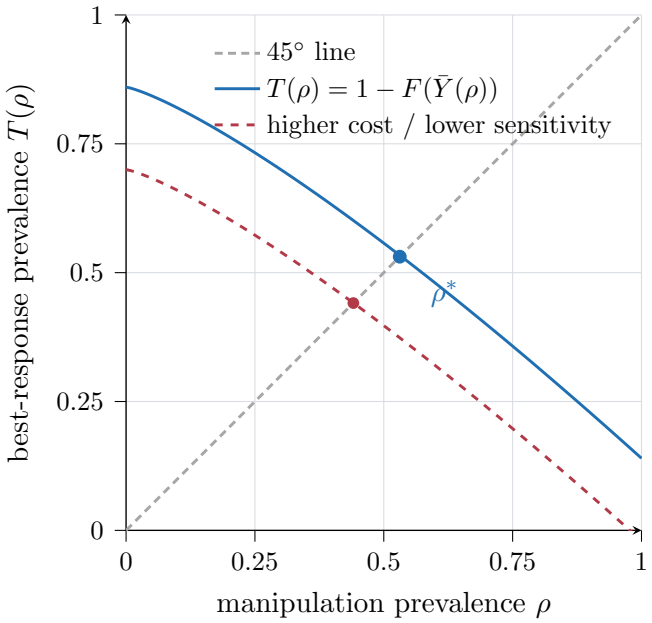

Figure 3: Equilibrium manipulation prevalence as a fixed point. In the plotted single-crossing case, prevalence ρ^* solves $\rho = T(\rho) = 1 - F(\bar{Y}(\rho))$ at the intersection with the 45° line. A higher displacement cost or a lower response sensitivity shifts T inward (dashed) and reduces equilibrium prevalence.

The manipulation-region threshold raises a final question: can the observational equivalence of Proposition 1 survive once cancellation is generated by optimizing behavior rather than chosen by hand? It can. A bona fide and a manipulation-driven threshold system can be matched to produce the same law of displayed statistics while keeping their payoff decompositions distinct. Appendix A states the matching lemma (Lemma 4) and the finite-support construction in full; the main-text consequence is the following.

THEOREM 4 (Equilibrium non-identification): *There exist threshold-policy primitives satisfying the optimal-stopping rules in Appendix A. Under those primitives, bona fide cancellation and manipulation-driven cancellation generate the same distribution of displayed order-path statistics:*

$$\mathcal{L}(X(P^B)) = \mathcal{L}(X(P^M)),$$

while

$$\overline{V}^{B}(P^{B};S) \geq 0 \quad \textit{and} \quad \overline{V}^{B}(P^{M};S) < 0 < \overline{V}^{B}(P^{M};S) + D(P^{M};S) - K^{D}(P^{M};S).$$

Thus optimizing threshold behavior does not, by itself, make public path statistics identify the source of value.

Proof. By Lemma 4 in Appendix A, choose bona fide and manipulation-driven threshold systems with a common displayed-statistic law. In the bona fide economy, cancellation is triggered by stale-quote exposure, inventory pressure, hedge pressure, or execution-risk information, with $\overline{V}^{B}(P^{B};S) \geq 0$. In the manipulation-driven economy, cancellation occurs after induced-response value has been extracted but before accidental execution or detection risk dominates, with $\overline{V}^{B}(P^{M};S) < 0 < \overline{V}^{B}(P^{M};S) + D(P^{M};S) - K^{D}(P^{M};S)$. The same law of X arises from the two payoff decompositions, so the observational equivalence survives threshold-based equilibrium behavior. ■

Taken together, these results make spoofing a choice rather than a label. It is the decision to display a path that the trader's own frontier cannot justify, in a market credible enough that others' reactions can justify it for the trader. The next section asks what an analyst must observe, beyond the path itself, to tell that this is what has happened.

VI. Structural Detection as an Audit-Data Design

The detection design is a partial-identification procedure, not a verdict. It does not claim to recover intent or to point-identify the wedge in every case. It asks a single counterfactual question, whether the observed path can be rationalized without the induced response of other traders, and it answers only when the relevant signs and margins can be bounded. When they cannot, the procedure returns an inconclusive case rather than a manipulation

label. Everything below is an audit-data design for classifying the source of continuation value; the legal question of intent is separate and is not addressed by the test.

By Proposition 1 and Proposition 2, pattern tests based on the coarse sigma-field $\mathcal{G}_X = \sigma(X(P))$, and even public-book histories without exposure restrictions, do not identify the two sources of value over the maintained model class. Table 1 shows the resulting information-set ladder. Structural detection operates above the public book and requires richer information:

$$\mathcal{G}_S = \sigma(\mathcal{G}_A, P, \mathcal{Z}(P;S), \text{matched controls}, \text{response}).$$

Here $\mathcal{Z}(P;S)$, defined below, collects event-window observables used to bound bona fide value and estimate response value. The key object is the response gap

$$\Delta R(P;S) = R(P;S) - R^0(P;S),$$

the behavior of other traders after observing the displayed path, relative to a path-excluded response counterfactual.

Table 2 states the data requirement explicitly. Public message data can triage paths and estimate auxiliary response moments, but they do not by themselves sign the wedge. Individual-path classification requires audit evidence that bounds the execution frontier, constructs a path-excluded response benchmark, and verifies signed exposure.

Object	Minimum evidence	Classification role
Bona fide value envelope	Fills, queue position, inventory, hedge book, stale-quote exposure, risk-limit state, and markouts.	Upper-bounds execution-facing value; the first leg requires the bound to be negative.
Path-excluded response gap	Public-book response variables, matched pre-event states, common-shock controls, and no-anticipation checks.	Estimates $R(P;S) - R^0(P;S)$; the second leg requires a positive lower bound on induced response.
Signed monetization exposure	Account-level positions, beneficial ownership, correlated-instrument exposure, hedge records, and monetizing trades.	Tests whether the trader profits from the estimated response; the third leg requires alignment with the response gap.
Public message screens	OTR, cancellation rate, lifetime, modification count, displayed size, distance from the touch, and persistent-noise measures.	Triage and calibration only; without the three objects above they do not classify source of value.

Table 2: Audit-data requirements for structural classification. Public message screens organize the search set, but the spoofing-wedge classification needs evidence for the envelope, the path-excluded response gap, and signed monetization exposure.

Definition 7 (Event window and response observables): For a displayed path P^d, define an event window

$$\mathcal{T}(P) = [t_0 - L_0, t_1 + L_1],$$

where t_0 is the first displayed message in the suspect path and t_1 is the cancellation or terminal message. Let

$$\mathcal{Z}(P;S) = \big(\text{fills},\ \text{queue changes},\ \text{inventory changes},\ \text{hedge changes},\ \text{markouts},\ \Delta\text{depth},\ \Delta\text{spread},\ \Delta\text{imbalance},\ \Delta\text{mktflow},\ \Delta\text{cross-instrument response}\big)$$

collect structural observables in $\mathcal{T}(P)$. The own-fill, queue, inventory, hedge, and markout components bound bona fide execution value. The response components estimate $\Delta R(P;S)$ and therefore $D(P;S)$.

DEFINITION 8 (Matched path-excluded counterfactual): For a suspect path P, let $\mathcal{C}(P)$ be a matched control set of paths P_j satisfying:

$$X(P_j) \approx X(P), \qquad S_j^- \approx S^-,$$

where S^- is the pre-event public and private-proxy state. Controls must pass pre-trend filters for depth, spread, imbalance, marketable flow, and cross-instrument returns. Let weights $w_j(P) \geq 0$ sum to one. The empirical path-excluded response is

$$\widehat{R}^0(P; S) = \sum_{P_j \in \mathcal{C}(P)} w_j(P) R(P_j; S_j),$$

and the estimated response gap is

$$\widehat{\Delta R}(P; S) = R(P; S) - \widehat{R}^0(P; S).$$

The controls need not be known bona fide paths. They are path-excluded comparisons: paths drawn from similar pre-event states whose response window is not contaminated by the suspect displayed path being evaluated.

Figure 4 illustrates the empirical object. The structural question is not whether the displayed path has a high cancellation rate; it is whether the post-event response exceeds the response implied by matched path-excluded controls.

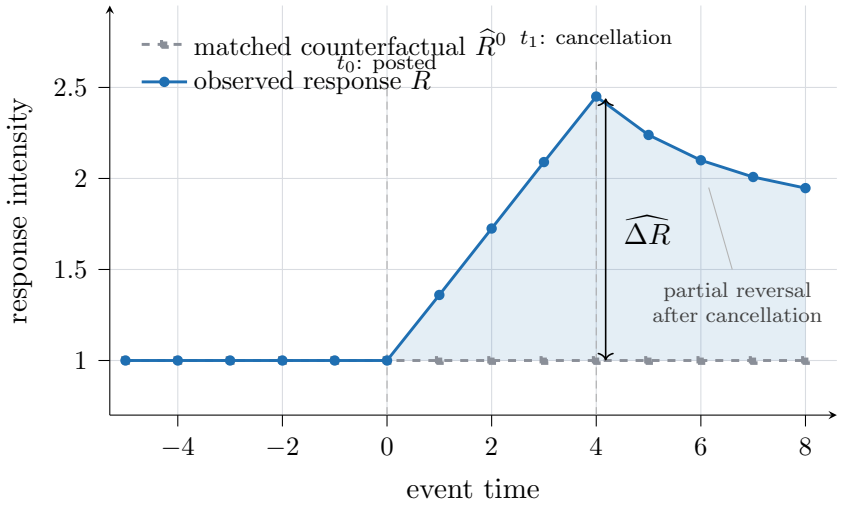

Figure 4: Response-gap event study. Structural detection compares the observed response R after a displayed path to a matched path-excluded counterfactual $\widehat{R}^0$ built from similar pre-event states, pre-trends, and displayed-path statistics. The shaded region is the estimated response gap $\widehat{\Delta R}$. A path that creates a temporary belief distortion rather than durable information produces a response that partially reverses after the cancellation at t_1.

ASSUMPTION 6 (Counterfactual response design): For each suspect path P, the matched controls satisfy:

(i) *Overlap*: $\mathcal{C}(P)$ is nonempty in a neighborhood of the pre-event state S^- and displayed statistic $X(P)$.

(ii) *No anticipation*: response variables have no abnormal pre-trend over $[t_0 - L_0, t_0)$.

(iii) *Conditional counterfactual*: conditional on S^-, $X(P)$, and pre-trends, the matched-control response equals the response that would have occurred under the path-excluded information intervention for P^d.

(iv) *Cross-instrument discipline*: correlated instruments and marketwide order-flow controls absorb common shocks unrelated to the suspect displayed path.

(v) *Stable response map*: the payoff map from response moments to induced-response value is locally Lipschitz.

DEFINITION 9 (Response elasticity estimand): Let $Z_R(P)$ be a scalar response moment, such as liquidity retreat, quote improvement, marketable-flow response, or cross-instrument repricing. Let $Z_D(P)$ be the displayed-depth shock created by P^d. The local response elasticity is

$$\mathcal{E}_R(P) = \frac{\widehat{\Delta Z_R}(P)}{Z_D(P)}.$$

High $\mathcal{E}_R(P)$, combined with negative own-execution value, is evidence that the path's value comes from induced response rather than own execution.

LEMMA 3 (Response-gap identification): *Under Assumption 6,*

$$\widehat{\Delta R}(P;S) = \Delta R(P;S) + u_R(P),$$

where $u_R(P)$ is the matching and sampling error. If the induced-response payoff map is locally Lipschitz with constant L_D, then

$$|\widehat{D}(P;S) - D(P;S)| \le L_D \|u_R(P)\|.$$

Proof. By the conditional-counterfactual condition, the weighted matched-control response estimates $R^0(P;S)$ up to matching and sampling error. Subtracting it from the observed response gives $\Delta R(P;S) + u_R(P)$. The bound for D follows from the local Lipschitz property of the payoff map from response moments to induced-response value. ■

The empirical object is partial identification, not universal point identification. Let $\mathcal{U}_R(P)$ be the admissible uncertainty set generated by matching error, sampling error, common-shock adjustment, and response-map calibration. A conservative response-gap interval is

$$\Delta R(P;S) \in \left[\widehat{\Delta R}(P;S) - \bar{u}_R(P),\ \widehat{\Delta R}(P;S) + \bar{u}_R(P)\right],$$

where $\bar{u}_R(P) = \sup_{u \in \mathcal{U}_R(P)} \|u\|$ after projecting response moments into the payoff-relevant

direction. If $Y(P;S)$ is known only to lie in an audit-feasible set $\mathcal{Y}(P)$, the signed exposure leg is bounded by

$$\inf_{Y\in\mathcal{Y}(P),\, u\in\mathcal{U}_R(P)} \big\langle Y, \widehat{\Delta R}(P;S) + u \big\rangle.$$

Only a positive lower bound signs monetization. If the response gap, signed monetization exposure, or bona fide envelope cannot be bounded with sign and margin, the structural test returns an inconclusive case. It does not classify the path by default.

Definition 10 (Response-dependence threshold): Scale the response channel by $\alpha \in [0,1]$:

$$R_\alpha(P;S) = R^0(P;S) + \alpha\big[R(P;S) - R^0(P;S)\big].$$

The corresponding payoff is

$$\Pi_\alpha(P;S) = \overline{V}^B(P;S) + \alpha D(P;S) - K^D(P;S).$$

The derivative $\partial\Pi_\alpha(P;S)/\partial\alpha = D(P;S)$ measures revealed response dependence. It is a counterfactual payoff dependence, not a claim to observe subjective intent.

Thus the word "revealed" refers to payoff dependence under a counterfactual response channel. A trader may have many subjective reasons for acting. The structural question is narrower: can the observed path be rationalized without the induced response of other market participants?

Theorem 5 (Revealed response dependence): *Suppose* $K^D(P;S) \geq 0$,

$$\overline{V}^B(P;S) < 0 \quad \textit{and} \quad \Pi_1(P;S) > 0.$$

Then $D(P;S) > 0$ and there is a unique response-dependence threshold

$$\alpha^*(P;S) = \frac{K^D(P;S) - \overline{V}^B(P;S)}{D(P;S)} \in (0,1)$$

such that

$$\Pi_\alpha(P;S) < 0 \quad \textit{for} \quad \alpha < \alpha^*(P;S),$$

and

$$\Pi_\alpha(P;S) > 0 \quad \textit{for} \quad \alpha > \alpha^*(P;S).$$

Lower α^* means stronger dependence on induced response: less response intensity is needed to make the path profitable.

Proof. The condition $\Pi_1(P;S) > 0$ gives

$$D(P;S) > K^D(P;S) - \overline{V}^B(P;S).$$

Since $K^D(P;S) \geq 0$ and $\overline{V}^B(P;S) < 0$, the right-hand side is strictly positive, so $D(P;S) > 0$. The payoff $\Pi_\alpha(P;S)$ is affine and strictly increasing in α. Solving $\Pi_\alpha(P;S) = 0$ gives the stated $\alpha^*(P;S)$. The inequalities follow from strict monotonicity. ■

Definition 11 (Partial-identification bounds): Let Θ_B be the admissible bona fide parameter set. The sharp upper bound on bona fide value is

$$\overline{V}^B(P;S) = \sup_{\psi_B \in \Theta_B} V^B(P;S,\psi_B),$$

which is the empirical counterpart of the envelope $\overline{V}^B(P;S)$ of Section IV: maximizing over the admissible class $\mathcal{B}$ corresponds to maximizing over its parameterization Θ_B. The two coincide when Θ_B indexes exactly the motives in $\mathcal{B}$. Let Θ collect admissible parameters for bona fide value, induced response, and displacement cost. The lower bound on displacement-

augmented value is

$$\underline{W}(P;S) = \inf_{\psi\in\Theta} \left[V^B(P;S,\psi_B) + D(P;S,\psi_D) - K^D(P;S,\psi_K)\right].$$

Equivalently, the response component can be written as an interval

$$D(P;S) \in [\underline{D}(P;S), \overline{D}(P;S)]$$

induced by the response-gap interval and the audit-feasible exposure set. The lower endpoint $\underline{D}(P;S)$, not the point estimate $\widehat{D}(P;S)$, is what enters a conservative classification.

Definition 12 (Three-leg structural test): The empirical detection design has three legs:

$$\overline{V}^B(P;S) < 0,$$

$$\underline{D}(P;S) > 0,$$

and

$$\inf_{Y\in\mathcal{Y}(P),\, u\in\mathcal{U}_R(P)} \left\langle Y, \widehat{\Delta R}(P;S) + u \right\rangle > 0,$$

where $\mathcal{Y}(P)$ is the trader's audit-feasible signed-exposure set across the monetized instruments, $\widehat{\Delta R}(P;S)$ is the estimated response gap in those same instruments, $\mathcal{U}_R(P)$ is the response-gap uncertainty set, and $\langle\cdot,\cdot\rangle$ is the Euclidean inner product. The third leg therefore requires that every admissible exposure-and-response completion preserve profitable alignment. In words: no execution rationale, positive lower-bound response value, profitable signed exposure.

Theorem 6 (Margin classification): *A path is structurally classified as spoofing-like, relative to the maintained model, if*

$$\overline{V}^B(P;S) < 0$$

and

$$\underline{W}(P;S) > 0.$$

The classification is conservative relative to the maintained parameter sets: every admissible bona fide rationalization is rejected, while displacement-augmented rationality remains strictly positive throughout the admissible set. If both bounds are estimated with uniform error at most ε, classification is robust whenever

$$\overline{V}^{B}(P;S) < -\eta \quad \textit{and} \quad \underline{W}(P;S) > \eta$$

for $\eta > 2\varepsilon$.

Proof. If $\overline{V}^{B}(P;S) < 0$, then even the most favorable admissible own-execution explanation gives negative value. Thus the path is outside the bona fide frontier. If $\underline{W}(P;S) > 0$, then even the least favorable displacement-augmented valuation rationalizes the path. Uniform error at most ε, including response-gap error from Lemma 3, cannot reverse either inequality when the margin is larger than 2ε. ■

This is an economic classification, not a legal conclusion. It does not claim to observe subjective intent. It identifies the payoff component necessary to rationalize the observed path. The spoofing-specific agent-based literature shows that displayed orders can be valuable through the beliefs of order-book learners (Wang and Wellman, 2017, 2021); the structural test here asks whether that induced-response channel is necessary to rationalize the path. Recent Level-3 surveillance work also moves toward manipulation-gain calculations from order-book features rather than pure message counts (Fabre and Challet, 2025). The structural object here is the economic version of that move:

Can the path be rationalized without induced response?

Operationally, Definition 12 asks for negative bona fide value, abnormal induced response,

and monetization. The first leg rejects the own-execution frontier. The second leg estimates the causal response gap. The third leg verifies that the trader had an economic exposure that benefits from that gap.

VII. Threshold Rules and Market Design

Regulators reach for hard caps and threshold rules because they are transparent and enforceable, but the identification result implies that such rules act on the wrong object. They discipline the surface statistic while leaving untouched the wedge that defines the conduct. The argument here connects to the broader market-design debate over whether continuous limit order books manufacture avoidable speed races and fragile priority contests (Budish, Cramton, and Shim, 2015), and to evidence that pre-trade transparency and order-disclosure rules shift spoofing incentives (Lee, Eom, and Park, 2013). In the present model, design matters because every lever enters the wedge,

$$\mathcal{W}(P; S, \zeta) = \overline{V}^{B}(P; S, \zeta) + D(P; S, \zeta) - K^{D}(P; S, \zeta),$$

where ζ is a vector of market-design parameters: tick size, cancellation friction, order-book transparency, queue-priority rule, accidental execution risk, and surveillance intensity. A path is in the manipulation region when $\overline{V}^{B}(P; S, \zeta) < 0$ and $\mathcal{W}(P; S, \zeta) > 0$. For a manipulation-like path, write $\mathcal{W}^{M}(P; S, \zeta)$ for this same wedge evaluated at that path.

Proposition 5 (Monotone design levers): *Fix P and S. If a design parameter ζ_j affects only one component of the wedge, then:*

$$\frac{\partial \mathcal{W}}{\partial \zeta_j} > 0 \quad \textit{if} \quad \frac{\partial D}{\partial \zeta_j} > 0,$$

and

$$\frac{\partial \mathcal{W}}{\partial \zeta_j} < 0 \quad \textit{if} \quad \frac{\partial K^D}{\partial \zeta_j} > 0 \quad \textit{or} \quad \frac{\partial \overline{V}^B}{\partial \zeta_j} < 0.$$

Thus manipulation incentives rise with induced-response sensitivity and fall with cancellation friction, detection cost, accidental execution risk, or trader skepticism that lowers D.

Proof. Differentiate $\mathcal{W} = \overline{V}^B + D - K^D$ with respect to ζ_j. If only one term changes, the stated signs follow directly. ■

PROPOSITION 6 (Threshold evasion and bunching): *Let $x = X(P)$ be a one-dimensional threshold statistic, such as OTR or a persistent-noise score, and suppose a rule applies penalty*

$$\kappa(x) = \begin{cases} 0, & x \leq \bar{x}, \\ K, & x > \bar{x}. \end{cases}$$

The trader solves

$$\max_x \{D(x) - A(x) - c(x) - \kappa(x)\}.$$

If $D(x) - A(x) - c(x)$ is continuous and positive for some $x < \bar{x}$, then a manipulation-motivated trader can strictly prefer a below-threshold spoofing path. If the unconstrained optimum lies above $\bar{x}$, the threshold rule generates bunching just below $\bar{x}$ rather than eliminating manipulation.

Proof. The discontinuous penalty creates a notch. Moving from just above to just below $\bar{x}$ preserves almost all continuous displacement value while avoiding the discrete penalty. If displacement value remains positive below the threshold, the trader has a profitable manipulation-like path that does not violate the hard cap. Thus the rule changes the chosen statistic but does not identify or eliminate the source of value. ■

COROLLARY 2 (Below-threshold displacement): *If $\overline{V}^B(x) + D(x) - K^D(x)$ is positive for some $x < \bar{x}$, then a hard threshold can leave profitable displacement paths below the enforce-*

ment cutoff:

$$\overline{V}^{B}(x) + D(x) - K^{D}(x) > 0 \quad \textit{for some } x < \bar{x}.$$

PROPOSITION 7 (Tick-size and transparency sign conditions): *Let Δ denote tick size and τ denote displayed-book transparency. Then the sign of their effect on manipulation incentives is determined by*

$$\frac{\partial \mathcal{W}}{\partial \Delta} = \frac{\partial \overline{V}^{B}}{\partial \Delta} + \frac{\partial D}{\partial \Delta} - \frac{\partial K^{D}}{\partial \Delta}$$

and

$$\frac{\partial \mathcal{W}}{\partial \tau} = \frac{\partial \overline{V}^{B}}{\partial \tau} + \frac{\partial D}{\partial \tau} - \frac{\partial K^{D}}{\partial \tau}.$$

Thus larger ticks or greater transparency increase spoofing-like incentives only when their induced-response effect exceeds their effect on bona fide execution value and displacement cost. They decrease spoofing-like incentives when they primarily raise accidental execution risk, improve structural surveillance, or reduce the credibility of displayed depth.

Proof. Both identities follow by differentiating the wedge with respect to the design variable. The comparative-static sign is positive exactly when the induced-response channel dominates the other channels, and negative when execution-cost or detection-cost channels dominate. ■

THEOREM 7 (Targeted response regulation): *Consider two policies. A uniform cancellation tax T_C reduces both bona fide value and manipulation value:*

$$\frac{\partial \overline{V}^{B}}{\partial T_C} < 0, \qquad \frac{\partial \mathcal{W}^{M}}{\partial T_C} < 0.$$

A targeted response-based surveillance rule T_R raises displacement cost in states with high response elasticity,

$$\frac{\partial K^{D}}{\partial T_R} > 0 \quad \textit{when} \quad \left|\frac{\partial R}{\partial P}\right| \textit{ is high,}$$

and leaves execution-facing value unchanged:

$$\frac{\partial \overline{V}^{B}}{\partial T_R} = 0$$

for bona fide paths. If bona fide liquidity has positive welfare value and response-elastic manipulation has positive social cost, then T_R weakly dominates T_C whenever it achieves the same reduction in manipulation with a smaller reduction in bona fide liquidity supply.

Proof. A uniform cancellation tax acts on the message action itself, so it reduces the value of both stale-quote protection, inventory control, and queue management, and manipulation-like paths. A response-based rule acts on the source of manipulation value by raising K^D when the response elasticity of other traders is high. If the two policies achieve the same reduction in manipulation but T_R imposes a smaller loss on bona fide liquidity, then the welfare comparison follows directly from the maintained signs. ■

These propositions separate two policy objectives. Caps can reduce message traffic, crude noise, and infrastructure burden. They are not, by themselves, source-of-value tests. Structural market design should therefore distinguish message discipline from manipulation identification. This distinction is also consistent with welfare-oriented legal analysis of spoofing regulation, which asks how misleading displayed orders affect market quality and allocative efficiency rather than treating cancellation as the primitive harm (Fox, Glosten, and Guan, 2022).

VIII. NSE-Style Illustration

This section is an institutional illustration, not an individual-path empirical classification. Without confidential audit-trail exposure, beneficial ownership, and account-level inventory or hedge records, the structural test cannot be run as a full individual-path classifier. The

NSE-style environment is useful because its order-to-trade and persistent-noise screens make the distinction between message discipline and source-of-value classification concrete. The exchange observes order entry, modification, cancellation, and execution, together with algorithm identifiers and member-level surveillance states, and it polices conduct in part through order-to-trade ratios and persistent-noise measures (Securities and Exchange Board of India, 2020; National Stock Exchange of India, 2020, 2026). These screens are sensible triage devices, and nothing in what follows argues against their use. The model explains, however, why they cannot be the last word: they live in the coarse information set $\mathcal{G}_X$, and the source of value lives above it.

Consider an NSE-style state

$$S_t^{\mathrm{NSE}} = (p_t, \mathrm{LTP}_t, d_t, Q_t, \omega_t, \nu_t, q_t, h_t, \lambda_t, \ell_t),$$

where LTP_t is last traded price, Q_t is queue state, ω_t is toxicity, ν_t is the message-intensity or order-to-trade state, q_t is inventory, h_t is hedge exposure, λ_t is execution-risk information, and ℓ_t is surveillance state.

The OTR and persistent-noise observables are components of $X(P)$. Algo identifiers and member-level surveillance states improve triage and repeated-conduct analysis, but they still do not reveal $\overline{V}^B(P;S)$, signed monetization exposure Y, or the causal response gap $R(P;S) - R^0(P;S)$ by themselves. This is the practical version of Proposition 2: public message data become structural evidence only after they are combined with exposure restrictions, audit-trail information, or credible bounds on bona fide execution value. The same data can nevertheless calibrate the first-stage simulation in Section IX: OTR distributions, persistent-noise states, response elasticities, and threshold bunching are public-facing inputs even when individual-path labels remain unclassified.

Proposition 8 (NSE coarse-screen non-identification): *In an NSE-like order book, there exist paths P and P' with identical order-to-trade ratio, cancellation rate, modification count,*

lifetime, displayed size, and distance from the touch, where P is bona fide and P' is structurally spoofing-like.

Proof. This is the NSE specialization of Proposition 1. Let P be a high-message path generated by stale-quote protection, queue management, hedge-risk reduction, execution-risk management, or adverse-selection avoidance. Let P' have the same coarse message statistics but no beneficial execution-facing state. Its value is positive only through induced response. Coarse screens see the same $X(P)$; value decomposition separates the paths. ■

Corollary 3 (NSE threshold bunching): *If OTR or persistent-noise rules impose a hard threshold and displacement value remains positive below that threshold, then manipulation-like paths can bunch below the cutoff. Therefore threshold compliance does not imply absence of a manipulation wedge.*

Proof. This is the threshold-evasion result applied to an NSE-style OTR or persistent-noise statistic. The penalty notch changes the displayed statistic. It does not by itself change the path's source of value. ■

A numerical illustration makes the point. Let

$$X(P) = X(P') = (1000, 40 \text{ ms}, 0.98, 20, 45, 3 \text{ ticks}),$$

with components interpreted as size, lifetime, cancellation rate, modification count, OTR, and distance from the touch. For P, suppose execution option value is 2, avoided stale-quote loss is 5, queue/risk-management value is 1, and message cost is 1, so $\overline{V}^B(P; S) = 7 > 0$. For P', suppose $\overline{V}^B(P'; S') = -5$, induced-response value is 9, and displacement-related cost is 2, so total value is $2 > 0$. The paths are identical to the OTR screen and different under the structural test.

The policy implication is layered surveillance: use OTR, persistent-noise, and modification screens as first-stage triage, use public-book event studies to calibrate response elastici-

ties and threshold bunching, and then estimate the bona fide value envelope, path-excluded response gap, and signed exposure for identification. The right second-stage question is not whether OTR is high. It is whether the path can be rationalized without the causal response of other traders.

IX. Simulation Design and Empirical Implications

The theory does not ask the analyst to observe intent; it asks the analyst to observe response and exposure. A structural empirical design estimates the bona fide envelope $\widehat{\overline{V}^B}(P;S)$ from fill probabilities, queue state, markouts, inventory and hedge records, and stale-quote exposure, and it estimates induced response from the behavior of others, cancellations, quote improvement or retreat, liquidity depletion, imbalance decay, short-horizon price response, queue migration, and cross-instrument repricing. Existing surveillance work already studies many of these moments, emphasizing order-book imbalance, quoting activity, abnormal cancellations, posting distance, Level-3 features, and manipulation-gain calculations rather than raw message counts (Lee, Eom, and Park, 2013; Do and Putnins, 2023; Fabre and Challet, 2025). What the framework adds is an interpretation: it tells the analyst which moments bound the frontier, which estimate the response gap, which require audit exposure, and when the evidence is insufficient for classification.

A. Empirical implications

The central contrast is between high-message paths with positive estimated bona fide value and high-message paths whose only apparent profit comes from induced response. In an NSE-style market the model predicts that coarse order-to-trade and persistent-noise screens contain both. Public data can motivate and calibrate these predictions, but individual-path classification requires the audit-data legs in Table 2. The model yields the following testable

implications.

1. *Coarse screens mix types.* High order-to-trade ratios, high modification counts, and short lifetimes are populated by both bona fide and spoofing-like paths, so conditioning on a high screen value leaves the structural label uncertain,

$$\Pr(M \mid X \text{ high}) \in (0, 1).$$

2. *Spoofing-like paths show worse own-execution value.* Conditional on fill risk, a structurally spoofing-like displayed order has lower estimated bona fide value than a genuine cancellation,

$$\widehat{\overline{V}^B}(P^M; S) < \widehat{\overline{V}^B}(P^B; S).$$

3. *Spoofing-like paths provoke stronger responses.* After a suspect displayed order, the response gap appears as abnormal movement in

$$\Delta\text{depth}, \quad \Delta\text{spread}, \quad \Delta\text{imbalance}, \quad \Delta\text{mktflow}, \quad \Delta\text{queue migration},$$

relative to matched path-excluded controls.

4. *Public-data evidence is informative but incomplete.* Public message data can reject some simple stories, for example when matched public controls absorb the response gap or when there is no reversal. But without exposure records or maintained exposure restrictions, public data cannot determine whether the trader monetized the response,

$$\inf_{Y \in \mathcal{Y}(P)} \left\langle Y, \widehat{\Delta R}(P; S) \right\rangle > 0.$$

5. *Cancellation is timed differently.* A bona fide order is cancelled when the trader's own fill risk rises; a spoofing-like order is cancelled once induced response has been harvested

but before accidental execution becomes likely, so its cancellation time approximates the harvest-maximizing instant,

$$\tau^M \approx \arg\max_t \left[D_t - A_t - K_t^D \right].$$

6. *The response partially reverses.* If the displayed path creates a temporary belief distortion rather than durable information, the response weakens or reverses after the cancellation message, absent news. Writing p_t for the response price, t_0 for the first displayed message, and t_1 for the cancellation, spoofing-like paths satisfy

$$\operatorname{sign}(\Delta p_{t_0,t_1}) = -\operatorname{sign}(\Delta p_{t_1,t_1+h})$$

more often than bona fide paths with similar $X(P)$ and pre-event states, and the reversal is strongest when audit exposure can be signed in the payoff-relevant direction,

$$\inf_{Y \in \mathcal{Y}(P),\, u \in \mathcal{U}_R(P)} \left\langle Y, \widehat{\Delta R}(P;S) + u \right\rangle > 0.$$

7. *Thresholds create bunching.* Where order-to-trade or persistent-noise rules impose discrete penalties, spoofing-like behavior bunches just below the cutoff, most heavily in states with high response elasticity.

8. *Spoofing value falls as credibility erodes.* If participants learn that displayed depth is unreliable, the marginal response value of depth declines, $\partial D/\partial depth \downarrow$, and spoofing becomes less profitable. This is the empirical content of Theorem 1.

The design is falsifiable, and it is meant to be. The spoofing-wedge interpretation is defeated if matched controls absorb the response gap, if bounding the envelope fails to deliver negative own-execution value, if the trader holds no signed exposure that monetizes the response, or if cancellation is not followed by the predicted decay or reversal. In any of

these cases the model does not force a manipulation verdict; it favors a bona fide reading or returns an inconclusive structural test, in keeping with the partial-identification discipline of Section VI.

B. Simulation design

A simulation can establish the logic before confidential audit-trail data are touched. It is not a substitute for audit-data classification. It is a controlled bridge from the public-data problem to the structural design: the econometrician knows the latent type, the public classifier does not, and the audit classifier is evaluated against the known data-generating process.

The simulation has four components.

1. *Environment.* Generate a limit-order book with a latent short-horizon state ϑ_t, queue position, spread, depth, stale-quote risk, toxicity, and cross-instrument price pressure. Bona fide traders receive inventory, hedge, queue, and stale-quote shocks. Manipulation-capable traders receive signed exposure Y_t in the same or correlated instruments.

2. *Path generation.* Bona fide paths choose display and cancellation policies to maximize execution-facing value. Spoofing-like paths choose visually similar display and cancellation policies to maximize

$$\overline{V}^B(P;S) + D(P;S) - K^D(P;S),$$

with negative own-execution value and positive response-facing value. The design should deliberately generate pairs with the same $X(P)$, including cancellation rate, lifetime, displayed size, modification count, and OTR.

3. *Responder learning.* Responders observe displayed depth and public order flow, update beliefs about ϑ_t, and choose quote retreat, quote improvement, marketable flow, or

queue migration. Manipulation intensity garbles displayed-depth informativeness, so simulated display credibility falls when spoofing-like paths become common.

4. *Observed data.* Record three information sets: public message data $\mathcal{G}_X$, public-book data $\mathcal{G}_{\text{book}}$, and audit data $\mathcal{G}_A$ containing inventory, hedge, beneficial-ownership, cross-instrument exposure, and monetizing trades.

The classifiers should be compared on the same simulated paths.

1. *Coarse public classifier*: uses OTR, cancellation rate, modification count, lifetime, displayed size, and distance from the touch.

2. *Public response classifier*: adds event-study response moments and matched public controls but no exposure records.

3. *Audit structural classifier*: uses the bona fide envelope, path-excluded response interval, and signed exposure alignment from Definition 12.

The evaluation reports false positives, false negatives, inconclusive rates, response-gap coverage, and robustness to common shocks. The central comparison is not whether public data are useless. It is whether public classifiers improve triage while failing to uniformly sign the source of value, and whether the audit structural classifier succeeds only when all three legs are observed or bounded with margin. Such an exercise is the controlled counterpart of Propositions 1 and 2, and the natural first implementation of the detection design.

X. Conclusion

Holding cancellation and source of value apart changes the object that surveillance and theory should pursue. The question is whether the displayed path can be rationalized on the trader's own execution frontier or only through its effect on others. Pattern tests do not answer that question without restrictions on hidden exposure. Over the unrestricted

hidden-exposure class, order-path statistics and the complete public book leave the source of value unidentified. The bona fide value envelope answers a different question: what could the path be worth before anyone reacts? The wedge marks the path as spoofing-like when that value is negative yet total value turns positive once induced response is counted.

The equilibrium gives the wedge its economic meaning. In the responder-learning model, spoofing pays because bona fide liquidity keeps displayed depth informative. That makes it self-limiting in its own prevalence: as manipulation spreads, the credibility it feeds on erodes, and response value disappears. While credibility holds, a trader with sufficient signed exposure can rationally display a path that loses money under every execution-facing motive. The harm therefore exceeds the single response a deceptive display induces. Bona fide liquidity builds the credibility of the book as a common resource; spoofing draws that resource down.

Two conclusions follow for practice. Detection must combine exposure evidence with a response counterfactual and must return inconclusive when signs and margins cannot be bounded, rather than convict on the silhouette of the path. The NSE-style illustration treats order-to-trade and persistent-noise screens as triage, not structural tests. Market design must reach the source of value rather than the message: hard caps discipline traffic while leaving the wedge intact below the threshold, whereas rules that raise the cost of deceptive response value reach the conduct itself. The economic classification developed here is deliberately silent on intent; it identifies the payoff a path requires, and leaves to law the question of the mind behind it.

The bona fide envelope supplies the conservative frontier. The path-excluded counterfactual supplies the response gap. Signed exposure supplies monetization.

Appendix A. Threshold-Equilibrium Construction

This appendix supplies the matching lemma and the finite-support construction behind Theorem 4. Its purpose is to show that the observational equivalence of Proposition 1 is not merely an artifact of hand-picked static paths. It can also arise when both bona fide and manipulation-driven cancellation follow optimal stopping against threshold indices.

LEMMA 4 (Threshold-statistic matching): *Let bona fide cancellation follow*

$$\tau^B = \inf\{t : \Lambda_t^B \geq 0\}, \qquad \Lambda_t^B = \lambda_t A_t - Q_t - \mathrm{IR}_t - \mathrm{HR}_t,$$

and let manipulation-driven cancellation follow

$$\tau^M = \inf\{t : \Lambda_t^M \geq 0\}, \qquad \Lambda_t^M = A_t + c_t + K_t^D - D_t.$$

If the joint law of displayed size, modification count, distance from the touch, and stopping time is the same under the two threshold systems, then

$$\mathcal{L}(X(P^B)) = \mathcal{L}(X(P^M)).$$

On any finite support for X, such matching can be implemented by choosing threshold indices Λ^B and Λ^M to induce the same distribution of first-passage times and displayed-order attributes while keeping their payoff components distinct.

Proof. The statistic $X(P)$ is a measurable function of displayed order attributes and stopping behavior: size, lifetime, cancellation rate, modification count, OTR, and distance from the touch. If the joint law of the displayed attributes and stopping time is identical under the two threshold systems, then the induced law of any measurable function of those variables is identical. On finite support, assign each displayed statistic x its target probability and

choose threshold-index paths that first cross at the stopping time and displayed attributes associated with x. The indices Λ^B and Λ^M can have the same crossing law while being built from different payoff components. ■

The bona fide index $\Lambda^B_t = \lambda_t A_t - Q_t - \mathrm{IR}_t - \mathrm{HR}_t$ crosses zero when accidental-execution exposure $\lambda_t A_t$ overtakes the queue, inventory, and hedge value of keeping the order posted; cancellation is then execution-facing and the path satisfies $\overline{V}^B(P^B; S) \geq 0$. The manipulation index $\Lambda^M_t = A_t + c_t + K^D_t - D_t$ crosses zero when the accumulated accidental-execution, message, and detection cost overtakes the induced-response value already extracted; cancellation is then timed to harvest D before exposure dominates, and the path satisfies $\overline{V}^B(P^M; S) < 0 < \overline{V}^B(P^M; S) + D(P^M; S) - K^D(P^M; S)$. Because the two indices are distinct functions of distinct payoff primitives, their crossing laws can be aligned on any finite support without aligning the underlying decompositions. Aligning the crossing law aligns the joint law of displayed attributes and stopping time, which by the lemma aligns $\mathcal{L}(X(P))$. Theorem 4 then follows as an existence construction: the same distribution of displayed statistics can be generated by a nonnegative and a negative bona fide envelope.

References

Allen, Franklin, and Douglas Gale, 1992, Stock-price manipulation, *Review of Financial Studies* 5, 503–529.

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance* 50, 1655–1689.

Brunnermeier, Markus K., and Lasse Heje Pedersen, 2005, Predatory trading, *Journal of Finance* 60, 1825–1863.

Budish, Eric, Peter Cramton, and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics* 130, 1547–1621.

Cartea, Álvaro, Sebastian Jaimungal, and Yixuan Wang, 2020, Spoofing and price manipulation in order-driven markets, *Applied Mathematical Finance* 27, 67–98.

Commodity Futures Trading Commission, 2013, Antidisruptive practices authority, *Federal Register* 78, 31890.

Do, Bao Linh, and Talis J. Putnins, 2023, Detecting layering and spoofing in markets, Working paper, University of Technology Sydney.

Fabre, Timothée, and Damien Challet, 2025, Learning the spoofability of limit order books with interpretable probabilistic neural networks, Working paper, arXiv:2504.15908.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets* 2, 99–134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, *Review of Financial Studies* 18, 1171–1217.

Fox, Merritt B., Lawrence R. Glosten, and Sue S. Guan, 2022, Spoofing and its regulation, *Columbia Business Law Review* 2021, 1244–1340.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Goldstein, Itay, and Alexander Guembel, 2008, Manipulation and the allocational role of prices, *Review of Economic Studies* 75, 133–164.

Hasbrouck, Joel, and Gideon Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16, 646–679.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1–33.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2017, The flash crash: High-frequency trading in an electronic market, *Journal of Finance* 72, 967–998.

Kumar, Praveen, and Duane J. Seppi, 1992, Futures manipulation with cash settlement, *Journal of Finance* 47, 1485–1502.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lee, Eun Jung, Kyong Shik Eom, and Kyung Suh Park, 2013, Microstructure-based manipulation: Strategic behavior and performance of spoofing traders, *Journal of Financial Markets* 16, 227–252.

Menkveld, Albert J., 2013, High frequency trading and the new market makers, *Journal of Financial Markets* 16, 712–740.

National Stock Exchange of India, 2020, High order to trade ratio (OTR), Circular Ref. No. NSE/SURV/45016.

National Stock Exchange of India, 2026, Persistent noise creators (PNC), Surveillance guidance.

O'Hara, Maureen, 2015, High frequency market microstructure, *Journal of Financial Economics* 116, 257–270.

Parlour, Christine A., 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789–816.

Putnins, Talis J., 2012, Market manipulation: A survey, *Journal of Economic Surveys* 26, 952–967.

Rosu, Ioanid, 2009, A dynamic model of the limit order book, *Review of Financial Studies* 22, 4601–4641.

Securities and Exchange Board of India, 2020, Guidelines for order-to-trade ratio (OTR) for algorithmic trading, Circular SEBI/HO/MRD1/DSAP/CIR/P/2020/107.

Wang, Xintong, and Michael P. Wellman, 2017, Spoofing the limit order book: An agent-based model, in *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems*, 651–659.

Wang, Xintong, and Michael P. Wellman, 2021, Spoofing the limit order book: A strategic agent-based analysis, *Games* 12, 46.