Cheap Signals, Costly Proof: Award-Layer Triage for Cartel Enforcement

Darcio Genicolo-Martins[1,*], Paulo Furquim de Azevedo[1]

[1] *INSPER, São Paulo, Brazil*

Abstract

Can cheap data triage cartel investigations before costly forensics? Most antitrust enforcers see who won a public contract and at what price—not what every bidder offered. This thinner data is enough to flag which firms warrant costly forensic investigation. Our architecture sequences an award-layer screening stage before a costly bid-layer forensic stage, instantiated with a *frequent-loser* flag on São Paulo's BEC (2009–2019). The flag cuts the forensic stage's bid-microdata pool by 83% while still flagging two-thirds of adjudicated cartel cobidders. On the same data, it matches the seven-feature Imhof–Wallimann bid-distribution pipeline that requires bid microdata, adding non-redundant signal in combination. The discrimination is out-of-sample: built on 2009–2016 participation only, the screen prospectively flags adjudicated cobidders in 2017–2019. The screen triages; it does not adjudicate cartel membership. A simple separating-equilibrium argument motivates endogenous loser-side participation as the ranking primitive. Wherever award records are routinely available while bid microdata are forensic-recoverable, screening should be sequenced before forensics.

Keywords: screening under incomplete observability, cover bidding, cartel adjacency, award-layer enforcement, separating equilibrium, participation-based statistics

JEL No.: D44, D73, H57, K21, L41

*Corresponding author.

Email address: darciogm1@insper.edu.br (Darcio Genicolo-Martins)

This version: May 2026. We thank seminar participants at INSPER for comments on earlier drafts; all remaining errors are our own.

1. Introduction

Can cheap data triage cartel investigations before costly forensics? The empirical industrial-organization literature reads collusion off the distribution of submitted bids (Porter and Zona, 1993; Bajari and Ye, 2003; Imhof, 2019; Wallimann et al., 2023; Chassang and Ortner, 2022). Yet most antitrust enforcers see only who won a public contract and at what price—not what every bidder offered. Electronic procurement systems archive the award envelope at an analytical-warehouse interface that returns query results in minutes; per-bidder bid amounts, where retained at all, sit in operational logs reachable only after weeks of administrative back-and-forth (§3). Methods designed for the bid layer simply do not run on the layer that survives.

The enforcement-design literature (Becker, 1968; Stigler, 1970; Baker, 2003; Harrington, 2008) prescribes a response to this asymmetry: investigative resources should be sequenced. A cheap screening stage triages the costly forensic stage. The screening stage runs on the data the agency already routinely observes; the forensic stage inherits a smaller pool. We propose an enforcement architecture along these lines, in which an award-layer screening stage triages firms and environments before a costly bid-layer forensic stage interrogates them. We instantiate the architecture with a *frequent-loser* flag built from São Paulo's BEC contract-award records (2009–2019, 1,654,401 tender-items).

A simple separating-equilibrium framework with cover bidders and genuine entrants disciplines what the construct measures (Online Appendix A). Cover bidders in equilibrium satisfy $\Pr(\text{win} \mid C) = 0$: the wins = 0 filter is a separating choice, not a descriptive coincidence. Among always-loser firms, $\log(1 + \text{tenders\_count})$ ranks firms by their posterior probability of being type C, and endogenous sorting into cartel-affected items is part of what the screening object measures rather than a confound to clean. The binary median $+ 1.5 \times$ IQR rule coarsens this continuous primitive, and we call firms above the threshold *frequent losers*.

Three findings emerge from the empirical demonstration. *First*, the flag cuts the forensic stage's bid-microdata pool by 83% (2,000 of 11,676 firms) while still flagging 131 of 193 adjudicated cartel cobidders. The architecture buys substantial data-envelope econ-

omy without sacrificing recall against the externally-validated cartel-adjacent population (§10). *Second*, using only contract-award data, the flag matches the seven-feature Imhof–Wallimann bid-distribution pipeline that requires bid microdata, and adds non-redundant signal when the two are combined (+0.035 AUC, DeLong $p = 0.014$). The screen and the bid-distribution methods are informational complements, not substitutes (§10). *Third*, built on 2009–2016 participation only, the screen prospectively flags adjudicated cobidders in 2017–2019 at firm-level AUC 0.864 under temporal holdout. The discrimination is out-of-sample, not retrospective fitting (§6).

The cobidder population behaves like the modeled type. Cobidders deploy at higher intensity than FL non-cobidders (Cohen's $d = +1.00$ on unique winners faced), bid proximally to direct CADE defendants ($d = +0.46$), and concentrate participation in narrower markets ($d = +0.39$ on portfolio item-group HHI). At the bid level they bid plausibly close to winners with elevated within-firm cross-bid dispersion ($p < 10^{-3}$ in a multivariate logit holding participation intensity constant). The signature matches credible cover bidding—bids low enough to be plausible, dispersed enough to track rotation across cover roles—rather than the textbook deliberately-uncompetitive cover bid the early literature described (§6.2). Frequent-loser presence is also associated with a $+3.6\% - +7.7\%$ higher conditional log negotiated unit price across four estimators on the broad sample, with the positive sign concentrated in the largest-tender-value stratum and a sign reversal under overlap restriction; we report this descriptively (§7) and the paper does not rest on it.

The contribution sits at the intersection of four literatures. Bid-coordination tests (Bajari and Ye, 2003; Porter and Zona, 1993; Conley and Decarolis, 2016) require a researcher-supplied partition between colluders and a competitive fringe; the construct supplies a candidate *ex ante* firm-level flag computable from award records. Bid-distribution screens (Imhof, 2019; Wallimann et al., 2023; Huber and Imhof, 2019) require bid microdata at the analytical-warehouse layer; the construct operates on the operational layer and integrates with bid-distribution methods at the forensic stage where microdata are recoverable. Empirical cover-bidding work identifies cover bidders *ex post*, through leniency testimony or judicial discovery (Porter and Zona, 1999; Pesendorfer, 2000; Asker, 2010;

Kawai and Nakabayashi, 2022); we are not aware of a published prospective identification of cover-bidder candidates from public award-record data alone. The enforcement-design literature (Becker, 1968; Stigler, 1970; Harrington, 2008; Baker, 2003) prescribes how detection technology and information acquisition interact under costly observation; the screening-then-forensic sequencing this paper proposes is a particular structure of that information chain.

Four claims the paper does not make. The construct does not adjudicate cartel membership: cobidders are the validation object the data layer supports, and the AUC asymmetry against direct CADE defendants is the design's empirical signature of the loser-side scope, not a failure of the screening logic. The pricing imprint is not a causal estimate of cover bidding's effect on prices. The buyer-size gradient is scope information about where the screening signal varies across detection regimes, not identification of an institutional channel. The pregão–convite modal asymmetry is scope information about where the screening object discriminates better, not a positive test of the minimum-bidder-rule mechanism.

Roadmap. §2 situates the contribution against the bid-coordination, bid-distribution, cover-bidding, and enforcement-design literatures. §3 describes the BEC observability layers and the CADE adjudication portfolio. §§4–5 state the participation primitive, the binary rule, and the empirical strategy. §6 carries the discrimination evidence and the within-stratum theory-validation bridge. §7 reports the price imprint as descriptive corroboration and the segment-level decomposition of the sign reversal. §§8–9 develop the modal contrast, the detection-regime gradient, and the threshold and identification audits. §10 carries the architectural test against bid-distribution detectors trained on bid microdata, including the sequential-gatekeeper analysis and its temporal-holdout audit. §§11–12 catalog scope limitations and conclude.

2. Related Literature

The paper sits at the conjunction of four literatures. Enforcement-design work asks how investigative resources should be sequenced when monitoring is costly and the layer

at which adjudication happens is more expensive than the layer at which suspicion can be ranked. Bid-coordination tests detect collusion off bid-function exchangeability conditional on a researcher-supplied partition between suspected colluders and a competitive fringe. Bid-distribution screens detect collusion off within-tender moments of submitted bids. And cover-bidding studies identify the loser side of collusive arrangements retrospectively, after a cartel has been adjudicated. We position the contribution in the first tradition: an award-layer triage stage that sequences which firms and environments enter the bid-layer forensic interrogation. The other three literatures supply the empirical objects the architecture operates on—a partition the bid-coordination tests can be applied to, a forensic battery the bid-distribution screens implement, and a type the cover-bidding literature has characterized *ex post* that we identify *ex ante* on award-record data.

Becker (1968); Stigler (1970) establish the optimal-deterrence baseline: when monitoring is imperfect, the sanction-detection trade-off has a structural solution that depends on the informational technology available to the enforcer. Baker (2003) treats the optimal sequencing of investigative stages under information asymmetry between agency and target: the screening rule that sorts candidates into investigation should economize on data the agency observes at low marginal cost. Harrington (2008) surveys the detection technologies available to antitrust authorities and the conditions under which each is informationally feasible. Two implications travel to our setting. First, when bid-level observability is absent at the operational layer, the appropriate screening object is whatever statistic the collusive arrangement generates that survives the data-coarsening step. Second, screening and forensic stages are sequenced, not parallel: the screening stage produces candidate environments for the forensic stage, which then deploys the more data-intensive bid-distribution moment battery (Chassang and Ortner, 2022). The architectural separation between an award-layer screening stage and a bid-layer forensic stage that the construct instantiates is the empirical implication of this sequencing logic. A growing empirical literature documents how procurement institutions shape what enforcement can observe (Coviello and Gagliarducci, 2017; Decarolis et al., 2020, 2025); we treat these institutional features as variation in the monitoring regime that the construct's signal is read against (§7.3).

Bajari and Ye (2003) establish the canonical framework for detecting collusion through exchangeability of bid functions and conditional independence of residuals, conditional on a researcher-supplied partition between suspected colluders and a competitive fringe. The papers that build on Bajari–Ye inherit the constraint: Porter and Zona (1993, 1999) detect bid rotation conditional on assumed cartel membership; Baldwin et al. (1997), Clark et al. (2021), and Schurter (2020) use post-prosecution records to identify colluders. The partition is constructed *ex post* or by assumption in every case; the construct supplies a candidate *ex ante* firm-level flag computable from award records that bid-coordination tests can take as input. Conley and Decarolis (2016) is the closest precedent in data parsimony: operating on Italian highway-construction auctions where investigators had identified a focal cartel but lacked a complete membership list, they infer bidder coalitions from co-bidding network structure. Three differences delimit the present contribution: their object is *group structure* (which firms belong to the same coalition given a focal investigation), while ours is *firm-level flags* on a screening basis where no investigation has been opened; their exercise operates on a post-investigation pool, ours on the *ex ante* universe of always-loser participants; their analysis recovers within-cartel structure on the assumption that a cartel exists, ours asks whether the participation footprint suggests a cartel-affected environment in the first place. The two methods are sequential, not substitutive.

Imhof (2019); Imhof et al. (2018); Huber and Imhof (2019) and Wallimann et al. (2023) train machine-learning classifiers on tender-level bid statistics—within-tender coefficient of variation, kurtosis, skewness, normalized spread, second-lowest distance, and related moments—reaching AUC above 0.85 against ground-truth European cartel cases; Abrantes-Metz et al. (2006) pioneer the variance screen and Chassang and Ortner (2022) provide moment-based theoretical foundations. We re-implement the seven-feature Imhof–Wallimann pipeline against our cobidder ground truth in §10; isolated single-moment benchmarks (the within-tender coefficient of variation alone) are strawmen of the literature's actual practice and we do not use them as the comparison object. The two layers stand in an asymmetric observability relation: bid-distribution screens require per-bidder bid amounts at the analytical-warehouse layer the screening pipeline runs on,

but in most enforcement environments those amounts are forensic-recoverable through administrative request rather than routinely operationally queryable (§3); the construct runs on the operational layer and integrates with bid-distribution screens at the forensic stage where microdata are recoverable.

Porter and Zona (1999); Pesendorfer (2000); Asker (2010) and Marshall and Marx (2012) document cover bidders in prosecuted cartels through case records, and Athey et al. (2011) compare cover-bidding incentives across auction formats theoretically; in every empirical case the cover bidders are identified *ex post*, through leniency testimony, post-conviction records, or judicial discovery, after the cartel has been adjudicated. We are not aware of a published prospective identification of cover-bidder candidates from public award-record data alone. The bid-level signature implied by the type varies across this literature: Porter and Zona (1993); Bajari and Ye (2003) read the constraint $b_C > b^*$ loosely, with cover bidders placing *visibly* uncompetitive bids that manufacture the appearance of competition; Marshall and Marx (2012) and Asker (2010) read it tighter, with cover bids remaining plausible enough to disguise coordination and cross-bid dispersion arising from rotation through cover roles. The bid-level extension of the within-stratum bridge in §6.2 discriminates between the two readings and is consistent with the second; the formal distinction is recorded in Online Appendix A.

3. Institutional Setting and the Observability Architecture

BEC carries three institutional features that the screening interpretation rests on. Two procurement modalities expose different observability regimes. A layer asymmetry separates the award records audit courts can query in minutes from the bid microdata they can recover only through administrative request. And a CADE adjudication portfolio supplies the cartel-adjacency labels we read the screen against, with a process-to-ruling interval long enough to make the validation in §6 partly prospective. None of this is Brazil-specific—analogous regimes in EU and U.S. enforcement environments expose a similar award/bid asymmetry.

Brazilian public procurement during 2009–2019 is governed by Lei 8.666/93 and Lei 10.520/2002, and two modalities dominate BEC. *Convite* (Lei 8.666/93, Art. 22)

is a sealed-bid invitation procedure with a statutory three-bidder quorum requirement and a low value cap; the only object preserved at the analytical-warehouse layer is the final award record. *Pregão* (Lei 10.520/2002) is a reverse electronic auction in which bidders submit progressively lower offers in real time through a public interface, with no minimum-bidder requirement; the real-time interaction footprint is observable to other bidders during the auction itself, and a richer set of bidder-by-bidder behaviors enters the operational record. The modalities differ in two dimensions simultaneously—bidding rule (sealed/quorum vs. real-time/no-quorum) and observability regime (award-only vs. real-time-public)—and we treat the modal contrast as variation in the regime within which the screening signal is recovered, not as the mechanism the construct exploits. The 2021 reform (Lei 14.133/2021) postdates our sample and consolidates pregão-style auctions as the institutional default.

BEC exposes two distinct observability layers, and the screening problem the construct addresses is defined on the asymmetry between them. The *operational layer* that BEC's analytical warehouse exports carries winner identity, participant identity, item code, negotiated price, procuring unit, and procurement- procedure metadata; audit-court analysts and competition-authority staff query this layer directly with results returned in minutes. The *forensic-recoverable layer* carries per-bidder bid amounts, bid timestamps, and bid-revision sequences; these records exist in the BEC operational system but are not exported to the analytical warehouse and are obtained, when at all, through case-specific administrative requests with weeks of latency, a cost flagged in OECD reviews of Brazilian procurement infrastructure (OECD, 2019).[2] The two-layer asymmetry is structural rather than parochial. The U.S. federal procurement system (FPDS-NG) reports contract awards but does not preserve bid amounts at the contracting-officer-accessible layer; bid microdata, when retained, sit in agency-level solicitation files that the FTC and DOJ Antitrust Division query through case-specific subpoenas rather than routine analytical access. The European Union's TED publishes the award-record layer mandatorily across

[2]The bid-distribution comparison in §10 and the bid-level extension of the within-stratum bridge in §6.2 draw on the forensic-recoverable layer through a LANCES export covering the entire sample window; see the disclosure footnote in §5.1.

member states but preserves bid microdata heterogeneously: a recent review of 12% procurement systems (Sánchez Graells, 2019) reports that fewer than half make per-bidder bid amounts retrievable at the analytical- warehouse layer. The U.K. Contracts Finder system preserves awards; bid-level archival depends on contracting-authority discretion. In each system per-bidder bid amounts are forensic-recoverable through subpoena, FOI request, or administrative procedure; what is not universally available is routine operational access. A screening object that operates on the operational layer is, by construction, deployable wherever that layer carries the participation primitive.

The Conselho Administrativo de Defesa Econômica (CADE) is the federal competition authority that adjudicates bid-rigging cases under Lei 12.529/2011, with leniency provisions (Art. 86) generating the disclosure trail the cartel-detection literature uses as ground truth. CADE's procurement-cartel portfolio relevant to our sample comprises 12 adjudicated cases covering 65 firm-defendants (47 active in BEC during the sample window), with final decisions spanning 2015–2025 and process-to-ruling intervals of 4–16 years. The interval is the relevant benchmark for the screening-stage interpretation: an award-layer screen that flags candidate environments at t accelerates the path to forensic investigation relative to the post-adjudication identification on which the prior literature relies. The construction of the validation populations and the prospective-versus-contemporaneous benchmark distinction are reported in §6; the administrative-audit institutions surrounding BEC (TCE-SP, CGE) supervise procurement legality and maintain debarred-firm registries that we treat as part of the broader monitoring environment within which the screening signal is read.

4. Data and Frequent-Losers Definition

4.1. BEC Platform

The data come from the Bolsa Eletrônica de Compras (BEC), São Paulo's centralized electronic procurement platform, operated by the State Treasury since 2002. São Paulo state has 46.1 million residents (21.6% of Brazil's population) and a 2023 GDP of R$3.45 trillion (31.5% of Brazil's GDP; IBGE *Sistema de Contas Regionais*, 2025). Treated

as an independent economy it would sit alongside Switzerland or Argentina, and state-level procurement spending is large enough in absolute terms to make the deployment value of an award-layer screening stage a meaningful policy question. BEC serves 1,308 purchasing units (PBUs)—state agencies, public universities, hospitals, schools—and clears two procurement modalities during our sample: *convite* (sealed-bid invitation procedure with a statutory three-bidder requirement, Lei 8.666/93) and *pregão* (electronic reverse auction with no minimum-bidder requirement, Lei 10.520/2002). The 2009–2019 window comprises 4,533,754 contract-record items across 22 calendar semesters, generated from 39,961,357 individual bids placed by 41,444 distinct firms (Table B.1). The contract-record layer preserves participants, winners, negotiated prices, reference prices, and procurement metadata; the bid-by-bid log is retained in the operational system but exported only on administrative request, the layer asymmetry characterized in §3. Pregão expands from a minority share toward near-universal coverage by 2019, consistent with the 2021 consolidation of pregão-style auctions as the institutional default; convite remains a meaningful share in the small-cap range and in early years, providing the within-platform variation that the modal contrast in §7.1 and §8.2 reads.

4.2. Frequent-Losers Definition

The framework's wins = 0 filter is a separating choice under standard expected-profit competition: with per-bid cost c, mean order value $\bar{V}$, and gross margin m, a firm that bids without ever winning crosses break-even on n losses at $n^* = \bar{V}m/c - 2$ (Online Appendix A). At BEC parameters ($\bar{V} \approx$ R\$86,000, $m = 0.10$, documented bid-preparation cost $c \in [\text{R\$200}, \text{R\$500}]$), $n^* \in [15, 41]$. The empirical FL threshold of 14 tender-items lies in this interval; the 90$^{\text{th}}$-percentile FL participates in 50+ zero-win tenders and accumulates direct cost above R\$10,000. Sustained zero-win participation past the break-even is consistent with cover bidding, capacity signaling, or eligibility maintenance; we organize the empirical work around the cover-bidding reading without claiming it as the unique interpretation (§11.2).

We define frequent losers as the intersection of two conditions: (i) *always-losers*, firms with a zero win rate across all 2009–2019 BEC tenders (16,843 firms, 41% of all BEC

participants), and (ii) participation counts above median + 1.5×IQR of the always-loser participation distribution.[3] The threshold is fixed *ex ante* on the participation distribution and the framework's break-even, before any examination of CADE labels, and yields 14 or more tenders with zero wins (2,735 frequent-loser firms). The treatment variable losers_{igt} equals one if tender-item i in item group g at time t has at least one frequent-loser participant. The economic construct is broader than the binary rule: Proposition 2 identifies $\log(1 + \text{tenders\_count})$ as a sufficient ranking statistic for cover-bidder type given award-record data, and the binary rule is the information-coarsening of this statistic under the framework's separating equilibrium; the horse race in §8.1 confirms that the continuous statistic strictly dominates the binary (DeLong et al. 1988, $p < 2e-05$). Out-of-sample generalization of the threshold itself is reassuring: estimating the cutoff on 2009–2016 alone gives a value of 7 with firm-level AUC 0.767 ([0.734, 0.800]); evaluating the 2009–2016-frozen score against 2017–2019 items, the binary indicator drops to 0.565 while the continuous primitive retains 0.770 (Table 2; full audit in §9.3).

Table 1: Loser-Side Concentration as Concept, FL as Operational Rule

Measure	Editorial role	Threshold	Full-sample AUC	Strict firm AUC	Strict item AUC
Continuous log(1 + tc)	Latent construct	—	0.939	0.750	0.770
Binary FL: median + 1.5 IQR	Operational rule	13.5	0.924	0.767	0.565
Alternative: Q3 + 1.5 IQR	Tukey alternative	18.5	0.834	—	—
Alternative: median + 1.0 IQR	Looser rule	10.0	0.902	—	—

Notes: The strict-train threshold estimated only on 2009–2016 is 7.0, far below the full-sample paper-rule threshold of 13.5. The framework identifies $\log(1 + \text{tenders\_count})$ as the participation primitive; the binary FL rule is its information-coarsening rather than a theoretically unique boundary.

4.3. Sample and Descriptive Statistics

The participation distribution among always-losers is heavily right-skewed (Figures B.1–B.2). The regression sample retains convite and pregão tenders with a recorded winner

[3] We use median + 1.5× IQR rather than the Tukey rule (Q3 + 1.5× IQR) because the participation distribution is heavily right-skewed; the adopted rule yields a stricter threshold. The Tukey alternative flags 1,981 firms with AUC 0.834, against 0.924 under the adopted rule. Robustness to the multiplier choice is in §9.1.

Table 2: Strict Temporal Holdout with Threshold Frozen on 2009–2016

Scope	Score	Threshold	AUC	95% CI	Positives / N
Firm-level (AL)	Binary FL	7.0	0.767	[0.734, 0.800]	193/21,819
Firm-level (AL)	$\log(1+\text{tc})$	—	0.750	[0.706, 0.795]	193/21,819
Item-level 17–19	Binary FL	7.0	0.565	[0.564, 0.566]	5,299/1,009,729
Item-level 17–19	$\log(1+\text{tc})$	—	0.770	[0.764, 0.776]	5,299/1,009,729

Notes: Train-window threshold (2009–2016 AL): 7.0. Full-sample reference: 13.5. Binary FL threshold estimated only on 2009–2016; item-level rows apply the frozen threshold to 2017–2019 items.

and item types with at least one frequent-loser participation during the sample window; this conditioning ensures that controls come from item markets where the screening statistic can plausibly fire, and the unrestricted sample returns a coefficient within 0.5 percentage points of the baseline (§9). The regression sample contains 1,654,401 tender-items across 18,783 item types defined at the finest product-code level; 15,101 (82%) of those item types have both frequent-loser-present and frequent-loser-absent tenders within sample, contributing to within-item identification under Equation (1). Frequent-loser firms participate in 4.8% of tender-items. The raw unconditional price ratio is roughly elevenfold (R$140,673 vs. R$12,465; Table 3), reflecting the sorting of frequent losers toward larger contracts; under within-item, within-year, within-procuring-unit fixed effects, the 2.43-log-point raw gap compresses to roughly 0.05 log points (Table B.2). The selection that the fixed-effects structure absorbs is large; the residual within-item association is small in log points but systematically positive across estimators (§7.1). The CADE adjudication portfolio that supplies the validation labels is described in §6.1.

5. Empirical Strategy

5.1. The Empirical Object

The screen we evaluate operates on four fields and four fields only: participant identity, winner identity, item code, and negotiated price. Our primary empirical object is how well the participation-based statistic computed from those fields discriminates

Table 3: Descriptive Statistics: Tenders With vs. Without Frequent Losers

	With FL			Without FL		
	Mean	SD	N	Mean	SD	N
Negotiated price	140,673.37	3,148,893.84	79,452	12,465.30	484,167.26	1,574,949
Log negotiated price	4.92	3.87	79,452	2.49	2.59	1,574,949
Number of firms	9.06	6.35	79,456	4.90	3.17	1,574,991
Log number of firms	1.99	0.67	79,456	1.39	0.65	1,574,991
Number of bids	18.56	24.55	79,456	9.58	13.19	1,574,991
Non-FL firms	7.82	6.01	79,456	4.90	3.17	1,574,991
FL count per tender	1.25	0.75	79,456	0.00	0.00	1,574,991

Notes: Sample restricted to item types with ≥1 FL tender. FL = frequent losers defined by median + 1.5×IQR threshold.

cartel-adjacency labels.[4] The discrimination object is what the framework's sufficient ranking statistic predicts will dominate when only the operational layer is queryable; the validation evidence in §6 carries it. The conditional log-price association reported in §7 is a secondary object: descriptive corroboration that the screening signal also leaves a price imprint on the items it concentrates on, not a treatment-effect target. Under the separating equilibrium of Online Appendix A, frequent-loser participation is generated by the cartel's deployment problem, and endogenous sorting into items where the cartel deploys is constitutive of the screening object; the relevant estimator integrates over the distribution of items into which frequent losers actually enter.

We report two empirical objects with the discrimination object as the primary defense and the conditional association as descriptive corroboration. The discrimination object is the AUC of the binary frequent-loser indicator and the underlying continuous statistic $\log(1+\text{tenders\_count})$ against three cartel-adjacency labels: cobidders inside the

[4] *Disclosure on data access.* The bid-distribution comparison in §10 and the bid-level extension of the within-stratum bridge in §6.2 require per-bidder bid amounts. Those amounts are not exported to BEC's analytical warehouse and were obtained, for the purposes of this paper, through a LANCES administrative export from the BEC operational system covering the entire 2009–2019 window. The screening object is built around the operational layer because that is the layer audit courts and competition authorities query routinely; the forensic comparisons use the forensic-recoverable layer because that is where the bid-distribution moments and the bid-level signatures are computable. The architectural claim is that the screening stage acts as a Stage-1 gatekeeper for the forensic stage, not that bid microdata are universally absent.

always-loser stratum on a strict pre-2020 contemporaneous benchmark with participation-stratified permutation null (§6.3); cobidders under temporal holdout, train 2009–2016 and test 2017–2019 (§6.4); and the same target read against the Imhof (2019)-style bid-distribution pipeline trained on the bid layer (§10). The conditional association β is the within-cell conditional mean difference in log negotiated price between tender-items that include at least one frequent-loser participant and otherwise comparable items in the same product code, year, and procuring unit; we estimate it on 1,654,401 tender-items and bracket it with four estimators (OLS, cross-fit, CEM, IPW), with the overlap-restricted counterpart β^{ov} reported in parallel (§7.2).

5.2. Reduced-Form Specification

The reduced-form specification is

$$y_{igt} = \beta \cdot \text{losers}_{igt} + \mathbf{x}'_{igt}\boldsymbol{\delta} + \alpha_g + \lambda_t + \gamma_k + \varepsilon_{igt} \tag{1}$$

where y_{igt} is the outcome for tender-item i in item group g at time t and procuring unit k; losers_{igt} indicates the presence of at least one frequent-loser participant; $\mathbf{x}_{igt}$ includes a convite indicator and specification-specific controls; α_g, λ_t, and γ_k are item, year, and procuring-unit fixed effects; ε_{igt} is the disturbance, clustered at the item level. The headline outcome is $\log p_{\text{negotiated}}$; auxiliary outcomes (number of bidders, number of bids, non-frequent-loser bidder count) are reported in §7.1 and the appendix. We estimate Equation (1) under four specifications: a general within-item-and-year baseline, a within-procuring-unit variant that adds γ_k, and two modal subsamples (pregão-only, convite-only) that decompose the headline along the observability-regime contrast of §3. Three complementary estimators bracket the OLS baseline. Cross-fitting defines the frequent-loser list on odd years and estimates on even years (and reverses), removing within-sample classification noise. CEM and IPW adjust for selection on observables (year, procedure, item group, procuring-unit-size quartile). All four estimators bracket the broad-sample object β rather than a treatment-effect target; the conditional range they jointly bound is $+3.6\% - +7.7\%$ (§7.1). A leave-one-out instrumental-variable specification

using the supply of frequent losers in other procuring units in the same product market and year as instrument returns a magnitude larger than OLS, the direction predicted when the binary indicator coarsens the underlying continuous loss-intensity primitive (§8.1); we do not interpret the IV estimate as a causal magnitude. Two design-based strategies return null: a sharp regression discontinuity at the convite/pregão statutory value caps is gated by the absence of bunching (McCrary-style density tests around the R\$80,000 cap return a left/right ratio of 0.94 and a log-density discontinuity of -0.063; Table B.5), and a difference-in-differences design using Decreto 9.412/2018's 2018 raise of the convite cap fails parallel trends at the available pre-treatment window. We report both failures openly; neither constrains the screening interpretation, which does not require a causal identification of β.

The role of the four identification audits in §9.2—Cinelli–Oster sensitivity bounds, strict-overlap matching, leakage decomposition, and the sub-threshold-always-loser placebo—is to calibrate the descriptive scope of β, not to instantiate causal identification: each audit answers a distinct question about the screening object rather than a question about the credibility of a causal estimate. The validation in §6 carries the discrimination evidence against three cartel-adjacency targets and the within-stratum theory-validation bridge. The within-item price association in §7 reports the broad-sample β, its overlap-restricted counterpart β^{ov}, and the segment-level decomposition that locates the sign reversal in the largest-tender-value stratum. The architectural test in §10 reads the screening statistic against bid-distribution detectors trained on bid microdata, including the sequential-gatekeeper analysis and its temporal-holdout audit.

6. Validation: Cartel Adjacency Under Coarsened Observability

The award-layer envelope (§3) does not contain the information required for firm-level cartel *membership* identification, and the screening framing does not require it: the participation primitive the framework identifies is a loser-side exposure measure, not a winner-side label. The validation object is therefore cartel *adjacency*, operationalized as the cobidder population—always-loser firms that participated alongside CADE direct defendants. We test the construct against this population on two benchmarks (a strict pre-

2020 contemporaneous one and a more permissive prospective one), bridge the population to the modeled cover-bidder type along firm- and bid-level dimensions, and document the design's empirical signature against direct defendants.

6.1. Validation populations

The validation reads against five nested populations (Table 4): from the broadest universe of 41,444 BEC participants inward through the 16,843-firm always-loser stratum to the 2,735 flagged frequent losers; the CADE adjudication record splits into 47 direct defendants—typically frequent winners and outside the construct's target—and 193 cobidders inside the always-loser stratum, which are the cartel-adjacency labels the screening statistic is designed to recover. Strong AUC against the cobidder label corroborates that the screening statistic concentrates on cartel-relevant environments; it does not adjudicate membership.

Table 4: Validation populations and object discipline

Population	Size	Definition and role
All BEC participants	41,444	Universe of firms bidding in BEC, 2009–2019.
Always-losers	16,843	Firms with zero wins; stratum within which the screening statistic operates.
Frequent losers	2,735	Always-losers above the median $+1.5\times$IQR participation threshold.
CADE direct defendants in BEC	47	Firms named in CADE rulings; typically frequent winners; outside target.
Cobidders in the always-loser stratum	193	Always-losers that participated alongside CADE direct defendants; the cartel-adjacency label the screening statistic is designed to recover.

6.2. Within-stratum profile of cobidders

A natural objection is that cobidders are just always-loser firms with a proximity label glued on, and that any discrimination AUC against them reflects participation intensity rather than cover-bidder type. We address this directly: Table 5 compares cobidders against three reference populations along four firm-level operational predictions of the

cover-bidder type formalized in Online Appendix A. The within-stratum comparison—cobidders versus FL non-cobidders, both above the FL threshold—moves in the direction the framework predicts on four of five measured dimensions. Cobidders deploy at higher intensity (136.5 vs 76.7 mean total participations, Cohen's $d = +0.67$, $p < 10^{-15}$; sharper still on unique winners faced, $d = +1.00$, $p < 10^{-22}$), are deployed proximally to the cartel (1.5% of their loss-bids face a direct CADE defendant against 0.2% for FL non-cobidders, $d = +0.46$, $p < 10^{-70}$), and specialize in narrower market slices (portfolio item-group HHI 0.380 vs 0.288, $d = +0.39$, $p < 10^{-13}$; touching 7.6 vs 9.5 distinct item-groups, $d = -0.32$, $p < 10^{-6}$). The one wrong-direction dimension is the share of own loss-bids in repeat-pair clusters of count ≥ 5 ($d = -0.38$); we report it without leaning on it.

Table 5: Cobidders vs. Reference Populations: Operational Predictions of the Cover-Bidder Type

	Cobidders ($N = 191$)	FL non-cobidders	Always-loser, non-FL	Other winners	Cohen's d vs. FL	Wilcoxon p
Mean portfolio item-group HHI	0.380	0.288	0.719	0.294	+0.39	< 0.001
Mean share of loss-bids facing a direct CADE defendant	0.015	0.002	0.002	0.003	+0.46	< 0.001
Mean share of loss-bids in winner-pairs of count ≥ 5	0.216	0.334	0.026	0.412	−0.38	< 0.001
Mean total participations	136.471	76.652	12.119	1603.320	+0.67	< 0.001

Notes: Each row reports a firm-level mean for four populations: cobidders (193 always-losers participating alongside CADE direct defendants), FL non-cobidders (2,735–193 frequent losers without recorded direct-defendant adjacency), always-losers below the FL threshold, and other winners (firms with at least one win, excluding direct CADE defendants). The two rightmost columns report Cohen's d and a Wilcoxon rank-sum test of cobidders against FL non-cobidders, the relevant within-stratum comparison. The four operational measures map directly to the cover-bidder type in Online Appendix A: total participation (deployment intensity), share of own loss-bids facing a direct CADE defendant (deployment proximity), share of loss-bids in winner-pairs of count ≥ 5 (repeat co-bidding density), and portfolio item-group HHI (specialization in the cartel's market).
Source: `scripts/60_theory_validation_bridge.R`.

The bridge extends to the bid level. We read per-bid *Valor Unitário Proposta* and *Flag Vencedor* from the forensic-recoverable LANCES export, and for each firm with at least five usable losing bids in convite or pregão we compute the per-bid ratio (bid − winning bid)/winning bid within each (OC, item) pair, winsorize at $[-0.99, 10]$, and aggregate at the firm level (182 cobidders, 2,369 FL non-cobidders). Reference-price ratios are not aggregated, because *Valor Unitário Referência* is populated for only 22% of convite and 32% of pregão losing bids and most firms end up without coverage. Three new dimensions move in a consistent direction. The median gap to the winning bid is 0.582 for cobidders against 0.809 for FL non-cobidders ($d = -0.281$, $p < 10^{-6}$); the same ordering holds at the 75th percentile ($d = -0.156$, $p = 0.007$); and the within-firm standard deviation of the gap is 1.207 vs. 1.099 ($d = +0.147$, $p = 0.05$). A multivariate

logit on cobidder status (within the FL stratum) holding $\log(1+\text{tenders\_count})$ constant returns a median-gap coefficient of -0.708 ($z = -4.14$, $p < 10^{-4}$) and a within-firm dispersion coefficient of $+0.442$ ($z = +3.62$, $p < 10^{-3}$). Both bid-level signs survive controlling for participation intensity, so the bid-level signature is not a multicollinearity artefact of the firm-level deployment-intensity result.

Table 6: Bid-Level Within-Stratum Profile of Cobidders vs. FL Non-Cobidders

	Cobidder	FL non-cob.	Cohen's d	Wilcoxon p
Panel A. Distance from the winning bid (per-bid (bid − winner)/winner, restricted to losing bids, winsorized at $[-0.99, 10]$ before per-firm aggregation).				
Median (bid − winner) / winner	0.582	0.809	−0.28	< 0.001
P75 (bid − winner) / winner	1.300	1.498	−0.16	0.007
Mean (bid − winner) / winner	0.957	1.102	−0.17	0.014
Panel B. Within-firm cross-bid dispersion of (bid − winner)/winner.				
Within-firm SD of (bid − winner) / winner	1.207	1.099	+0.15	0.050
Panel C. Multivariate profile: logit of cobidder on per-firm bid-level moments within the FL stratum, holding $\log(1+\text{tenders\_count})$ constant. Estimates show whether bid-level metrics discriminate cobidders beyond their higher participation count.				
$\log(1+\text{tenders\_count})$	coef. +0.651	SE 0.094	—	< 0.001
Median (bid − winner) / winner (winsorized)	coef. −0.708	SE 0.171	—	< 0.001
Within-firm SD of (bid − winner) / winner (winsorized)	coef. +0.442	SE 0.122	—	< 0.001

Notes: Per-bid *Valor Unitário Proposta* and per-(OC, item) winning bid (identified by *Flag Vencedor*) read from `bid_level_full_v14.parquet`; restricted to losing bids in CONVITE or PREGÃO ELETRÔNICO. Per-bid (bid − winner)/winner ratio winsorized at $[-0.99, 10]$ before per-firm aggregation; firms retained with at least 5 usable losing bids. Reference-price ratios are reported in the online appendix only: BEC's *Valor Unitário Referência* is populated for ∼22 % of CONVITE losing bids and ∼32 % of PREGÃO losing bids, leaving most cobidders without firm-level coverage. The cover-bidder framework in Online Appendix A predicts that cobidders, as the modeled type, place bids that maintain credible separation from the winner with cross-bid dispersion above what genuine losers display. The signs reported here *do not* match the textbook "deliberately uncompetitive" cover bid: cobidders bid systematically *closer* to winners than FL non-cobidders, while their within-firm dispersion is mildly elevated. The pattern is consistent with a refined credible-cover-bidding interpretation under which cover bids must be plausible to disguise coordination, not with an indiscriminately-high-bid version of the type. Cohen's d uses pooled standard deviations; Wilcoxon p-values from two-sided rank-sum tests; logit standard errors are conventional MLE.
Source: `scripts/62_theory_bridge_bidlevel.R`.

The signs run against the textbook reading of the cover-bidder type. A deliberately uncompetitive cover bid would put cobidders *further* from the winner, not closer. What we see lines up instead with the credible-cover-bidding reading of Marshall and Marx (2012) and Asker (2010), in which cover bids must remain plausible enough to disguise coordination and the firm cycles through cover roles across tenders. We adopt that reading explicitly (Online Appendix A records the formal R1/R2 distinction). Within what the data can measure, cobidders match the modeled type along seven of nine dimensions, and the discrimination evidence in the rest of this section reads against a validation object whose firm-level *and* bid-level profile match the modeled type rather than a label of convenience.

6.3. Conservative benchmark

The primary benchmark restricts the ground truth to 4 CADE cases adjudicated before 2020: 30 firm-defendants active in BEC, with 210 always-losers co-bidding with at least one of them, of which 108 are frequent losers. We construct the counterfactual baseline by stratified permutation: for each of 1,000 iterations we permute cobidder/non-cobidder labels within strata defined by participation-volume quartile, recompute the share of frequent losers among the relabeled cobidders, and accumulate the empirical

distribution under the null of random matching conditional on participation volume. The participation-stratified co-bidding rate observed in the data is 3.95% against a permutation baseline of 1.24% ($p < 0.001$, 1,000 iterations), an excess ratio of 3.2×; the ROC AUC against the contemporaneous ground truth is 0.748 (95% DeLong CI [0.713, 0.783]). This is the primary benchmark on three grounds: the construct uses no information from cases adjudicated post-sample; cartel existence and membership were established by enforcement before the screening statistic was built; and the permutation baseline strips out the mechanical co-bidding rate that participation volume would produce by itself, isolating the residual concentration the screening statistic is meant to identify.

6.4. Prospective benchmark

A more permissive benchmark uses CADE's full procurement-cartel portfolio (12 cases, 65 firm-defendants, of which 47 are active in BEC). Among 2,735 frequent losers, 193 (7.1%) co-participate with at least one direct defendant; the participation-stratified excess ratio is 3.5× ($p < 0.001$). Firm-level AUC is 0.924 in-sample (95% CI [0.921, 0.926]) and 0.864 under temporal holdout (train 2009–2016, test 2017–2019; CI [0.858, 0.870]). Figure 1 shows the rolling-origin temporal-holdout ROC curves separating cleanly from the diagonal across all six test years; year-by-year AUCs are in Online Appendix B. The benchmark treats post-sample adjudications as informative labels for cartel involvement during the BEC window, a reasonable but non-trivial assumption that the conservative benchmark avoids.

Two further diagnostics tighten the discrimination evidence (Tables 7–8). The participation-stratified permutation null behind the conservative benchmark places the empirical excess ratio of 3.2× in the upper tail of the null distribution at $p < 0.001$. The leakage decomposition separates the in-sample item-level discrimination AUC of 0.939 into a structural component recovered when cobidder firms are held out of score construction within cross-validation folds (AUC $\approx$ 0.86–0.89) and a mechanical-tautology component ($\approx$ 0.10–0.13); the structural component is what the screening interpretation rests on. A within-firm exercise on the same portfolio: 3 of 7 always-loser direct defendants (43%) cross the frequent-loser threshold against a population baseline of 16%; these are the

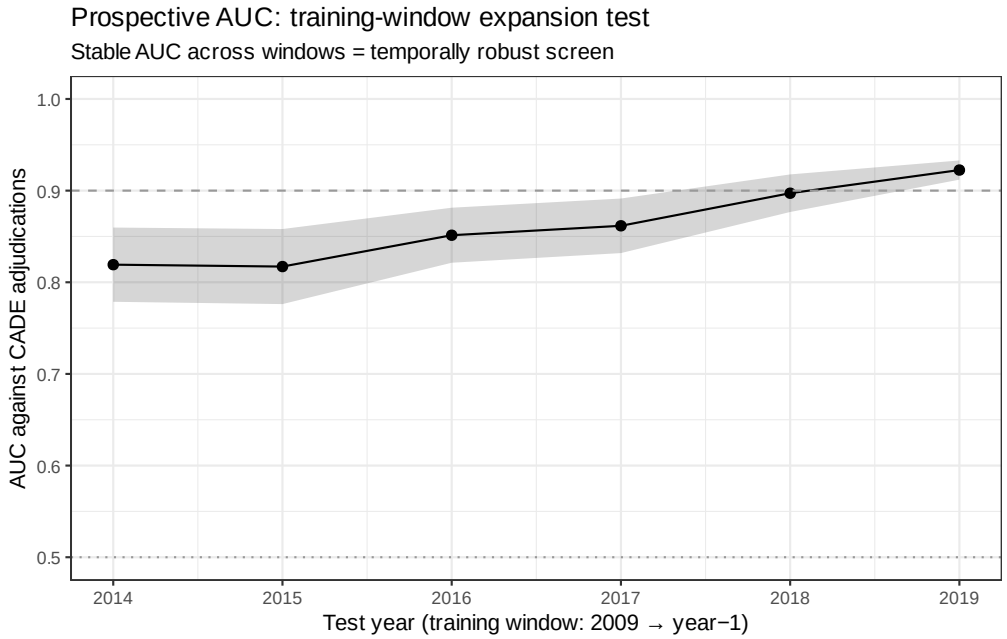

Figure 1: ROC curves under rolling-origin temporal holdout; training window is 2009 through test year-1, the test year provides the out-of-sample cobidder ground truth.

exception within a direct-defendant population that is otherwise frequent-winner-heavy.

The AUC asymmetry between cobidder discrimination (0.924 in-sample, 0.864 under temporal holdout) and direct-defendant discrimination (0.49 across the broader BEC firm universe) is not a failure of validation but the design's empirical signature: the screen recovers loser-side participation footprints, and direct defendants are by construction the winner side of the same arrangements. Dropping the 31,447 CADE-involved tender-items and re-estimating the within-PBU baseline yields $\hat{\beta} = 0.062$ (N = 1,622,954), virtually unchanged from 6.4%, so the broad-sample screening signal does not depend on within-sample CADE adjudications for its content. The four pieces of evidence point in one direction: a 3.2× contemporaneous excess, a 3.5× prospective excess, cobidder AUCs of 0.748 and 0.864, and the within-firm direct-defendant exception. Inside the always-loser stratum, frequent losers concentrate near adjudicated cartel members at rates well above what participation-stratified random matching would produce. That is what the validation establishes. It does not adjudicate firm-level membership, and the screening framing does not require it to.

Table 7: CADE Validation: Co-Participation and Permutation Test

Panel A: Co-participation with CADE Cartelists		
CADE procurement cartel cases (unique processes)	12	
firm-defendants across all cases	65	
final adjudication dates	2015–2025	
cases adjudicated after 2019 (prospective)	8	
CADE firm-defendants matched to BEC via CNPJ	47	
	FL Firms	Non-FL Firms
Co-bids with CADE	193	2,290
Does not co-bid	2,542	36,419
Total	2,735	38,709
Co-participation rate	7.1%	5.9%
χ^2 statistic	5.7	
p-value	0.017	
Odds ratio	1.21	
Panel B: Permutation Test (1,000 iterations)		
Observed FL-CADE rate	0.0706	
Permutation mean rate	0.0201	
Permutation SD	0.0026	
p-value (rate $\geq$ observed)	0.0000	

Notes: Panel A: contingency table of firm co-participation with CADE cartel-defendants. Eight of the 12 cases were adjudicated after the BEC sample ends (post-2019), so the validation is partially prospective: it asks whether 2009–2019 FL flags anticipate firms later confirmed by enforcement. Panel B: 1,000 random draws of 2,735 firms from the always-loser pool, stratified by participation quartile, computing co-participation rate with CADE firm-defendants.

Table 8: Leakage Audit: Decomposition of the In-Sample Item-Level AUC (0.995) into a Structural Component (out-of-fold and temporal-holdout AUC $\approx$ 0.86–0.89) and a Pure-Leakage Component (0.10–0.13). The structural component is what the screening interpretation rests on.

Audit specification	What it tests	AUC	95% CI
Original (in-sample, full pool)	Reference: log(1+max tc) predicting any-cobidder participation.	0.995	[0.995, 0.995]
Audit 1: Different label, same score (scope)			
Same score $\rightarrow$ any-direct-CADE label	Whether the screen also flags items with direct CADE defendants (N+=54,039 items).	0.506	[0.505, 0.507]
Audit 2: 5-fold cross-validation at the cobidder-firm level (tautology)			
Pooled out-of-fold	Hold out $\sim$ 1/5 of cobidders per fold; recompute tenders_count without their participation; predict items containing held-out cobidders using the recomputed score. Pool out-of-fold predictions across folds.	0.891	[0.887, 0.894]
Audit 3: Temporal holdout (generalization)			
Train 2009–2016 $\rightarrow$ test 2017–2019, any-cobidder	Compute tenders_count using only 2009–2016 items; predict items in 2017–2019.	0.864	[0.859, 0.870]
Train 2009–2016 $\rightarrow$ test 2017–2019, any-direct-CADE	Same temporal split, against direct-CADE label.	0.511	[0.510, 0.513]

Notes: The in-sample AUC of 0.995 is partly tautological by construction. Audit 2 isolates the structural component by holding cobidder firms out of score construction within each fold; Audit 3 isolates it temporally by computing the score on 2009–2016 participation only. AUC retains ≥ 0.85 under both audits; the 0.10–0.13 drop is the pure-leakage component. Audits against direct-CADE labels return random AUC, consistent with the scope asymmetry between loser-side cobidders and winner-side direct defendants documented in §6.

7. Pricing Imprint as Descriptive Corroboration

Do environments the screen flags also carry a price imprint? This section reports the conditional association between frequent-loser presence and log negotiated unit price as descriptive corroboration of the discrimination evidence in §6. The paper does not rest on this section. The screening contribution is carried by the discrimination evidence in §6 and the architectural test in §10; the price imprint is a third piece of evidence on whether the construct also tracks an outcome the collusion literature would expect cover bidding to leave on prices. We report it briefly across three steps: the broad-sample conditional association β (§7.1); the overlap-restricted β^{ov} and the segment-level decomposition of the resulting sign reversal (§7.2); and the heterogeneity of the signal across detection regimes (§7.3). The screening-value formalization that motivates why the broad-sample β remains an economic object under coarsened observability is in Online Appendix A (Proposition 4); we do not lean on it in the body of the paper.

7.1. Broad-Sample Conditional Association (β)

Frequent-loser presence associates with a higher log negotiated unit price in all four specifications of Equation (1): 6.8% within item and year, 6.4% adding procuring-unit fixed effects, and 9.3% in pregão against 3.8% in convite (Table 9). Cross-fit estimation yields 3.6%, CEM +0.077, IPW 0.055, and a leave-one-out IV returns 0.194 ($p < 0.05$, first-stage $F = 396$). The four estimators jointly bound the conditional range $+3.6\% - +7.7\%$.

The pregão–convite asymmetry falsifies a quorum-filler reading. Lei 8.666/93 Art. 22 § 3 requires a three-bidder minimum in convite; if the construct's signal were mechanically driven by firms entering to satisfy that quorum, the imprint would be sharper in convite. The data show the opposite (+9.59% in pregão vs. +3.92% in convite; Table B.4).

β is a conditional association across items where frequent losers deploy. Whether it is also a treatment effect on prices is a separate question that the available identification cannot settle. §7.2 shows that overlap restriction reverses the sign and decomposes the reversal across tender-value segments.

Table 9: Negotiated Prices (log): Effect of Frequent Losers

	(1) General	(2) General	(3) Pregão	(4) Convite
FL presence	0.0677***	0.0636***	0.0933***	0.0382**
	(0.0230)	(0.0215)	(0.0255)	(0.0186)
Convite	−0.0090**	0.0095***		
	(0.0037)	(0.0035)		
Observations	1,654,401	1,654,401	546,549	1,107,852
R^2	0.879	0.886	0.882	0.893
Item FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
PBU FE	NO	YES	YES	YES

Notes: Standard errors clustered at the item level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (3) and (4) restrict the sample to pregão and convite procedures.

7.2. Overlap Restriction, Sign Reversal, and Segment-Level Decomposition

Restricting comparisons to cells where frequent-loser-present and -absent items genuinely overlap on observables and reweighting to the average treatment effect on the treated produces a reversed coefficient: overlap-cell ATT −9.72%, propensity-score-trimmed ATT −30.67% (Online Appendix B, Table C.5). The reversal is real; the cell-dropping decomposition reported in the same appendix table (Table C.6) shows it is a reweighting result rather than a dropped-counterfactual result: only 1.06% of treated items lack a within-cell counterfactual and are strictly dropped, the remaining 99% participate in both estimates, and what changes is which untreated counterfactuals each estimator up-weights. The paper does not rest on either sign. The segment-level decomposition that follows gives the sharper structure. Within tender-value quintiles, the broad-sample coefficient is already negative in Q1–Q3 (−0.065, −0.057, −0.040, $p < 10^{-15}$ in each) and positive in Q4 (+0.046, $p = 0.012$). Both the overlap restriction and the ATT reweighting leave each within-quintile estimate essentially unchanged: the Q4 positive survives all three specifications at +0.041 ($p = 0.045$ under ATT weights), and the Q1–Q3 negatives survive at quantitatively similar magnitudes. The aggregate $\widehat{\beta} = +0.064$ is therefore the population-weighted average of three negative within-quintile

coefficients and one positive Q4 coefficient, with Q4's 40,878 treated items dominating the volume-weighted mean; $\widehat{\beta}^{\text{ov}}$ is the same average under cell-size weights that down-weight Q4's volume (Table 10, Panel A). Trimming the most heavily ATT-weighted cells *strengthens* the negative estimate—from −0.097 at no trim to −0.118 at the top decile and −0.133 at the median—so the negative is structural in the bulk of the overlap sample, not extreme leverage on cells with thin counterfactuals (Panel C). Inside the overlap sample, two further wedges echo the segment reading: $\widehat{\beta}^{\text{ov}}$ is statistically null among items containing direct CADE defendants (−0.061, $p = 0.45$, against −0.097 in non-direct-CADE items, $p < 0.001$), and positive in the highest tender-value quartile (+0.041, $p = 0.045$).

Table 10: Where the Sign Reversal Lives: Segment-Level $\widehat{\beta}$ Estimates

Segment	Level	Broad $\widehat{\beta}$	Overlap unwt. $\widehat{\beta}$	Overlap ATT $\widehat{\beta}^{ov}$	N (broad)
Item group	12	+0.265	+0.256	−0.072	95,490
Item group	13	+0.255	+0.245	−0.129	99,364
Item group	29	+0.029	+0.022	−0.051	82,334
Tender-value Q	1	−0.065	−0.064	−0.053	434,419
Tender-value Q	2	−0.057	−0.057	−0.048	476,570
Tender-value Q	3	−0.040	−0.042	−0.037	391,299
Tender-value Q	4	+0.046	+0.046	+0.041	352,106
Year cohort	2009-2016	+0.062	+0.047	−0.081	1,381,033
Year cohort	2017-2019	+0.064	+0.033	−0.105	273,368
Panel B. ATT-weight concentration and trim sensitivity.					
HHI of cell weights	—	—	—	0.0006	8,625 cells
Top 1% cells weight share	—	—	—	0.152	—
Top 10% cells weight share	—	—	—	0.546	—
Panel C. Overlap ATT $\widehat{\beta}^{ov}$ after trimming most heavily weighted cells.					
no trim	—	—	—	−0.097	1,517,868 ($N_{\text{trt}} = 78{,}613$)
drop top 1%	—	—	—	−0.107	1,440,988 ($N_{\text{trt}} = 66{,}844$)
drop top 5%	—	—	—	−0.113	1,192,587 ($N_{\text{trt}} = 48{,}407$)
drop top 10%	—	—	—	−0.118	973,582 ($N_{\text{trt}} = 36{,}677$)
drop top 25%	—	—	—	−0.115	550,801 ($N_{\text{trt}} = 18{,}438$)
drop top 50%	—	—	—	−0.133	240,307 ($N_{\text{trt}} = 7{,}468$)

Notes: Panel A re-estimates the headline regression of Panel A of Table C.6 within segments. "Broad" uses all items; "Overlap unwt." restricts to overlap cells without ATT weights; "Overlap ATT" adds the cell-size-based ATT weights that produce $\widehat{\beta}^{ov}$ in Table C.6. Panel B reports the Herfindahl of cell-level ATT weights and the share of total weight concentrated in the top 1 % / 10 % of cells. Panel C re-estimates $\widehat{\beta}^{ov}$ after dropping the most heavily weighted cells; if the negative reversal is mechanically driven by a small fraction of cells with thin untreated counterfactuals, $\widehat{\beta}^{ov}$ should attenuate sharply as we trim.
Source: `scripts/61_sign_reversal_segment_decomp.R`.

What the segment decomposition supports is narrower than a treatment-effect identifi-

cation. In Q4—the segment where deployment value is most plausible *a priori* on contract size—the imprint is positive and robust to design choice (+0.046 broad, +0.041 under ATT). In Q1–Q3 the imprint is negative, consistent with several non-exclusive readings (thin-market price formation, selection on item characteristics, composition heterogeneity in the firm pool) that the available identification cannot disentangle. The paper does not rest on either sign. The discrimination evidence in §6 and the architectural test in §10 carry the contribution; the broad-sample imprint is bracketed by Cinelli and Hazlett (2020)'s $RV_{q=1} = 17.5\%$ and Oster (2019)'s $\hat{\delta} = 261.6$ against selection on observables already absorbed by the fixed effects. Figure 2 traces the coefficient across the four anchor specifications. The conditions under which β remains an economic object under coarsened observability are formalized in Online Appendix A (Proposition 4); the body of the paper does not lean on that formalization.

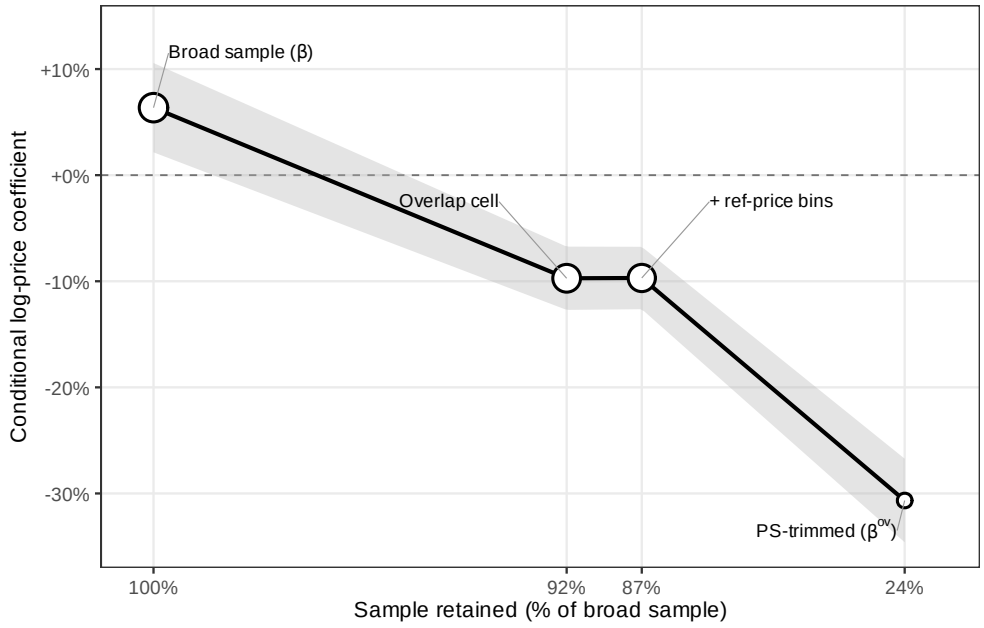

Figure 2: Conditional log-price coefficient across the four anchor specifications of §7.2 (broad-sample, overlap-restricted, ATT-weighted, PS-trimmed; full coefficients in Online Appendix B, Table C.5), with point size encoding sample retention and shaded band the 95% CI; the gap between β and β^{ov} decomposes into a positive Q4 component and three negative Q1–Q3 components, all robust to design choice (Table 10).

7.3. Heterogeneity Across Detection Regimes

Where the screening signal varies tells us something the level does not. Splitting the sample into quartiles of procuring-unit size produces a sharp extreme-quartile contrast: +21.4% at the smallest-buyer quartile (Q1) and +1.7% at the largest (Q4), a 12.6× gradient preserved across alternative size measures (annual contract volume, cumulative item count, headcount of distinct purchasers). The intermediate quartiles are imprecisely estimated and not statistically distinguishable from each other or from Q4 (Table B.3).

The framework predicts the contrast direction through the detection-cost comparative static $\partial m^*/\partial \theta_k < 0$: cover-bidder deployment is more aggressive where the principal cost of detection is lower. Buyer size proxies for several correlated institutional features—procurement-officer tenure (Coviello and Gagliarducci, 2017), internal-audit infrastructure, item composition, and discretionary procedure use (Decarolis et al., 2025, 2020)—none separable with the variation we have. We read the extreme-quartile contrast as heterogeneity in the screening object across the monitoring environment, in the framework's predicted direction, with the magnitude concentrated at the extremes.

8. Heterogeneity of the Screening Signal

Two pieces of within-evidence discipline the framing. First, the continuous loss-intensity statistic strictly dominates the binary rule in discrimination, in line with Proposition 2's identification of $\log(1 + \text{tenders\_count})$ as the sufficient ranking statistic (§8.1). Second, the modal contrast between pregão and convite reads as variation in the observability regime within which the screen is recovered (§8.2). Two finer framework predictions—a cell-heterogeneity grid and a discrete first-time-FL imprint—are reported in §11.2 and Online Appendix C; the data do not adjudicate them under disciplined estimation, and the screening interpretation does not require them.

8.1. Loss Intensity Dominates the Threshold Cutoff

The binary rule is the information-coarsening of the participation primitive identified by the framework, not a primitive in itself. On a harmonized same-sample basis, the continuous log(1+tenders_count) score discriminates the 193 adjudicated CADE-cobidder

firms inside the always-loser stratum at AUC 0.939 (95% CI [0.932, 0.946]) versus AUC 0.911 ([0.898, 0.925]) for the binary FL14 indicator (DeLong et al. 1988 test, $Z = -4.30$, $p = 2e-05$): the $\sim$2.8-point AUC gain on the same firm panel is the dose-response signature the framework predicts when the binary coarsens a continuous primitive. Standalone price coefficients align (Table 11, columns 1–2): FL14 binary +0.0653 ($p < 0.001$), continuous log(1 + max tc) +0.0188 ($p < 0.001$). When both are included additively (column 3), the FL14 coefficient flips sign (-0.0746, $p < 0.05$) while the continuous coefficient nearly doubles (+0.0349, $p < 0.001$): once the continuous primitive is in the specification, the binary indicator captures only tail residuals, and reading the binary in isolation overstates the cover-bidder type information by absorbing the variation the continuous primitive carries. The dose-response is consistent with the screening interpretation and inconsistent with a pure-confound reading of the broad-sample association.

Table 11: Horse Race: FL14 Binary vs Continuous Loss Intensity (Same Sample)

Specification	Coef on FL14	Coef on log(1+ max tc)	Convite	N
(1) FL14 only	+0.0653***	–	+0.171***	1,653,658
	(0.0216)		(0.013)	
(2) Continuous only	–	+0.0188***	+0.169***	1,653,658
		(0.0055)	(0.013)	
(3) Both, additive	−0.0746*	+0.0349***	+0.169***	1,653,658
	(0.0383)	(0.0108)	(0.013)	

Firm-level AUC against 193 *CADE-cobidder labels (always-loser pool,* $N = 16{,}843$*)*	
FL14 binary	AUC = 0.911 [0.898, 0.925]
log(1+tenders_count)	AUC = 0.939 [0.932, 0.946]
DeLong test (continuous vs FL14)	$Z = -4.30$, $p = 2e-05$

Notes: Outcome is $\log p_{\text{negotiated}}$ within the v13 panel of 1,653,658 items. Item-level treatment is **1**[any FL14 participant] in column (1) and the log of the maximum loss-intensity score among always-loser participants in column (2). Column (3) includes both. The FL14 coefficient flips sign when continuous intensity is added: the binary discretization at 14 tenders is the information-coarsening of the underlying continuous signal, not an economically primitive threshold. SEs clustered at item level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We use the binary FL14 rule as the discrete cut-off and the continuous statistic as the discriminating instrument: the headline range +3.6% – + 7.7% in §7.1 is computed on the binary, and the AUC headline 0.924 (95% CI [0.921, 0.926]) reported in §6.4 is binary firm-level discrimination.

8.2. Modal Contrast as Observability-Regime Variation

The screening signal is sharper in pregão than in convite along two independent dimensions. On the price-imprint side, the within-modal estimates of §7.1 place the binary specification at +9.59% ($p < 10^{-3}$) inside pregão versus +3.92% ($p = 0.037$) inside convite, a ratio of 2.45× on the larger pregão sample. On the discrimination side, the modal-by-modal AUC against the 193 adjudicated CADE-cobidder firms is 0.952 in pregão primary auctions versus 0.816 in convite primary auctions, with a bootstrap difference of −0.136 ($p \approx 0$). The two dimensions point in the same direction; the convite-modal AUC carries a sample-size caveat (6 adjudicated cobidders inside the convite stratum against several dozen on the pregão side) and we treat it as a directional indicator, with the much larger 1,105,852-tender-item price-imprint evidence carrying the statistical weight.

The convite minimum-bidder rule (Lei 8.666/93, Art. 22 § 3) is the obvious institutional suspect: a quorum-filler reading would predict a sharper screening signal where the rule binds, that is, in convite. The data reverse this prediction. We do not read the reversal as a positive test of any alternative institutional theory, but the contrast is consistent with a structural feature of the two modalities that the cover-bidding literature has long recognized. Cover bidding in a sealed-bid environment is a single *passive* act: the cover bidder submits an envelope and the participation footprint is one observation per item. Cover bidding in a real-time electronic auction is a sequence of *active* acts: the cover bidder must respond to other bids, maintain relevance without winning, and replicate plausible timing patterns over the auction's lifespan. Each revision is an additional observation against which the deployment pattern is read. Pregão therefore does not just expose a richer interaction footprint than convite—it forces the cover bidder to leave more of one. This is the information-theoretic counterpart of the dynamic-vs-sealed-bid distinction in Marshall and Marx (2012); Asker (2010); Athey et al. (2011): a sealed envelope hides whatever the auction does not call out of it; a real-time auction calls more out. Two compatible mechanisms operate alongside this—bidder-pool composition that survives convite's quorum requirement may itself be more homogeneous than the pregão pool, and the observability regime shapes what the analyst can read off the

data envelope—and the design separates none of the three from each other. What the design *does* establish is the asymmetry's direction and its alignment with how the cover-bidder type behaves under each modality. We therefore report the modal asymmetry as *scope information* for the screening object—the construct discriminates better in pregão environments—and not as a positive test of the minimum-bidder-rule mechanism, which would require institutional variation the BEC setting does not deliver.

The implication for portability is concrete. The post-sample reform (Lei 14.133/2021) consolidates pregão-style auctions as the institutional default, which is the regime in which the screening signal is sharpest. The construct travels best to jurisdictions whose observability regime resembles pregão—real-time electronic auctions with public bidder-by-bidder interaction—and the regulatory direction is favorable rather than adversarial.

9. Robustness

9.1. Threshold and Specification Sensitivity

The headline price gap and its detection-regime gradient survive reasonable variations in the design choices that define the screen. The IQR multiplier governs the bright-line cutoff that separates frequent losers from the broader always-loser stratum. Under five multipliers from 1.0× to 3.0× (Table D.1), the price coefficient is positive and significant in every row, declining monotonically from +0.079 at 1.0× (3,442 firms) to +0.050 at 3.0× (1,456 firms). The decline is a feature, not a defect: as the cutoff tightens, the binary indicator filters fewer firms with high loss-intensity, and §8.1 establishes that the underlying primitive is the continuous $\log(1 + \text{tenders\_count})$ statistic. Standard-error clustering does not change the point estimate; under item-level clustering (baseline, SE 0.0215), procuring-unit-level clustering (SE 0.0144, tighter), and two-way Cameron et al. (2011) clustering (SE 0.0240, looser), the coefficient is significant at $p < 0.01$ (Table D.2). The continuous score discriminates the same cobidder population at AUC 0.939 (95% CI $[0.932,\ 0.946]$) versus 0.911 ($[0.898,\ 0.925]$) for FL14 (DeLong $Z = -4.30$, $p = 2e-05$); standalone price coefficients of +0.0653 ($p < 0.001$, FL14 binary) and +0.0188 ($p < 0.001$, continuous) point in the same direction (Table 11). The headline price gap is neither a threshold artifact nor a binary-discretization artifact.

9.2. Identification Audits

Four audits calibrate the descriptive scope of β under specific identification concerns. The Cinelli and Hazlett (2020) robustness value $RV_{q=1} = 17.5\%$ implies an unobserved confounder would need to explain at least 17.5% of the residual variation in both frequent-loser presence and log price to drive the coefficient to zero; Oster (2019) bounds agree directionally with $\hat{\delta} = 261.6$ on the full sample (Table C.1). The pregão subsample yields $\hat{\delta} = 634.00$; the convite subsample carries a smaller margin ($\hat{\delta} = 43.62$), consistent with §8.2's finding that the construct's signal-to-noise is highest in pregão environments.

The strict-overlap matching exercise asks a different identification question. Restricting comparisons to cells in which frequent-loser items and non-frequent-loser items genuinely overlap on observables returns $\widehat{\beta}^{\text{ov}} = -9.72\%$ ($N = 1{,}517{,}868$); propensity-score-trimmed matching returns -30.67%. The contrast with the broad-sample $\widehat{\beta} = +0.064$ is decomposed in §7.2: the unweighted overlap regression returns $+0.044$ (only 1.06% of treated items are strictly dropped), the segment-level decomposition locates the broad-sample positive in tender-value Q4 ($+0.046$ robust to all three specifications) and the negatives in Q1–Q3 (-0.065 to -0.040, also robust), and trim sensitivity *strengthens* the negative ATT as the most heavily weighted cells are removed (from -0.097 to -0.118 at the top decile). Both signs describe the same population from different design angles; we report the four specifications jointly so the reader can locate the construct's descriptive claim within the appropriate scope.

The leakage audit (Table 8) addresses tautological leakage from cobidders being by construction elevated-participation always-losers. Holding the score fixed and swapping the label to direct CADE defendants returns 0.506, random discrimination that confirms the construct's structural scope (cover bidders, not ringleaders). Withholding cobidder firms from score construction within five cross-validation folds and re-evaluating on items containing held-out cobidders returns AUC 0.891 (95% CI [0.887, 0.894]); under temporal holdout (train 2009–2016, test 2017–2019) AUC is 0.864 (95% CI [0.858, 0.870]). The 0.10–0.13 drop from the in-sample reference is the pure-leakage component, and we adopt the temporal-holdout AUC of 0.864 as the conservative discriminating reference. A placebo specification using the supply of sub-threshold always-losers as instrument confirms the

construct's price imprint draws on the above-threshold cover-bidder population, not on generic always-loser supply (Online Appendix C, Table C.2).

9.3. Temporal Holdout: Generalization of the Screening Statistic

In-sample evaluation confounds classification with prediction: items in the late panel contribute both to score construction and to the labels the score is evaluated against. Constructing the score on 2009–2016 participation only and evaluating on items in 2017–2019 yields a firm-level AUC of 0.864 (95% CI [0.858, 0.870]) against the 193 adjudicated CADE-cobidder labels—the generalization reference for the screening statistic throughout the paper. Operational metrics (Table E.1) show in-sample precision overstates by approximately 50% at top-500 (0.132 in-sample versus 0.070 holdout) because roughly 47% of the top-500 in-sample ranking comes from 2017–2019 participation, after CADE adjudications were already underway for some cartels. We disclose the gap explicitly and report headline metrics on the holdout column. At top-1,000 the screen flags adjudicated cobidder firms with recall 34% under temporal holdout, and the architectural extension of the temporal-holdout audit to the sequential-gatekeeper rule appears in Table 14 (§10).

10. Screening vs. Forensic Stages: An Architectural Test

The screen we have built so far runs on award records. The claim of this section is that doing so is not a concession to data scarcity—it is the right division of labor between the screening stage and the forensic stage that follows.

Under incomplete operational observability of the bid layer (§3), enforcement architecture splits naturally into two stages: an award-layer screening stage that produces candidate flags, and a bid-layer forensic stage that adjudicates suspected coordination on the bid distribution. We test that division along three dimensions. The two stages discriminate the same cartel-adjacency labels at comparable AUC on different observability layers. The combination beats either component alone. And the award-layer screen functions as a Stage-1 gatekeeper, narrowing forensic-recoverable bid-microdata interrogation to a manageable subpool while preserving most of the joint-scoring recall.

Figure 3 sets up the comparison. The bid layer carries the seven within-tender features the Imhof–Wallimann pipeline aggregates into a discrimination score (within-tender coefficient of variation, its standard deviation across tenders, skewness, kurtosis, normalized spread, log min/max, log-second-lowest distance); the operational layer preserves only the award envelope—winner identity, participant identity, item code, negotiated price—on which the screening statistic operates. The screen the title refers to is the award-layer rule alone; the bid-microdata pipeline appears in this section as the benchmark we read the screen against and as the Stage-2 input the screen triages, never as a screening input itself.

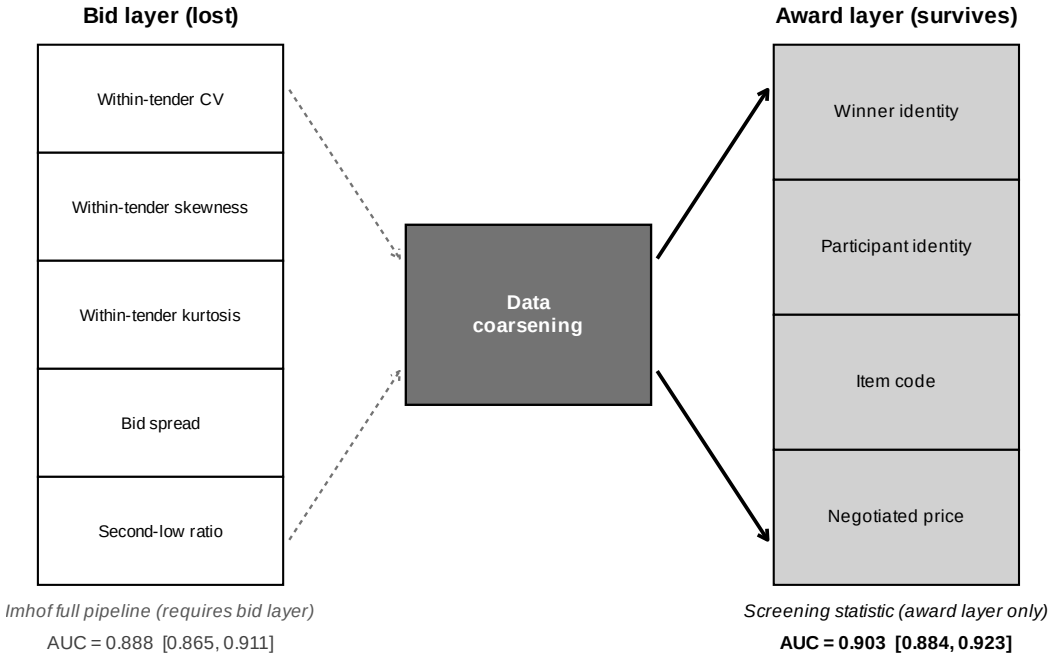

Figure 3: Bid-layer features and award-layer fields. The bid-distribution features the Imhof–Wallimann pipeline operates on (left, dashed arrows) are dropped by the coarsening that defines most enforcement environments at the operational layer; the four award-record fields (right, solid arrows) survive. Boxes on the right report random-forest AUCs against 193 adjudicated CADE-cobidder labels (Table 12).

On the full always-loser pool (N = 16,779 firms, 193 adjudicated cobidder labels; Table 12), the screening statistic discriminates at AUC 0.903 (95% CI [0.884, 0.923]) under the binary FL14 indicator and 0.884 ([0.860, 0.908]) under the continuous log(1 + tenders_count) score. The full Imhof–Wallimann pipeline reaches AUC 0.888 ([0.865, 0.911]) on the same target. The two pipelines reach statistically indistinguish-

able discrimination on different data envelopes; the construct does not outperform the bid-distribution pipeline, and we do not claim it does—the architectural point is that comparable accuracy is achievable on a substantially thinner data envelope. FL14 combined with the full Imhof pipeline reaches AUC 0.955 ([0.943, 0.967]); continuous loss intensity combined with the full pipeline reaches 0.962 ([0.954, 0.969]); the combined classifiers gain 5–7 AUC points over either single construct. Restricting attention to the same-sample audit (Table E.2), where all detectors evaluate the identical subset of firms—the subset for which Imhof features can be computed because bid microdata are available—the screening statistic adds +0.035 AUC over the Imhof full pipeline alone (DeLong $p = 0.014$) and the combined model adds $+0.096 - - + 0.098$ AUC over Imhof full alone ($p < 0.001$). The award-layer screen and the bid-layer forensic moments carry non-redundant information about the same target. The subset restriction is for the comparison, not for the screen: the screen itself runs on the four award-record fields whether or not bid microdata are recoverable for any given firm.

Beyond the AUC headline lies the operational test. The award-layer screen functions as a Stage-1 gatekeeper for the bid-layer forensic stage, narrowing the pool of firms whose bid microdata the forensic stage needs to interrogate. The screen itself does not need bid microdata; the gatekeeper claim is about how it shapes the data envelope that Stage-2 must request. Table 13 reports precision, recall, and bid-microdata footprint for four rules on the same evaluation pool of 11,676 always-loser firms (193 adjudicated cobidder positives, base rate 0.0165): the award-layer screen alone, the bid-layer Imhof full pipeline alone, joint single-model scoring, and the sequential gatekeeper that uses the award-layer screen to select a Stage-1 pool of K_1 firms before applying Imhof full within that pool. At top-1,000 flags, joint single-model scoring captures 142 of 193 cobidders (recall 0.736) while requiring bid microdata for the entire 11,676-firm pool; the sequential gatekeeper with $K_1 = 2{,}000$ captures 131 cobidders (recall 0.679) while requiring bid microdata for 2,000 firms only—a 83% reduction in bid-microdata footprint at a 8% recall cost relative to joint scoring. The sequential rule is not the precision champion (joint scoring is, at 0.224 vs. 0.192 at top-500); its in-sample operational case is the data-envelope wedge.

In-sample evaluation invites the natural worry that the score and the labels both

Table 12: Imhof–Wallimann Pipeline vs. Screening Statistic: Discrimination at Different Data Costs

Model	Features required	AUC	95% CI
Participation-only (award-record level)			
FL flag (binary)	participation count > 14, win flag	0.903	[0.884, 0.923]
log(1+tenders_count)	participation count, win flag	0.884	[0.860, 0.908]
Bid-distribution (per-bid microdata)			
Imhof CV only	per-bid offer values	0.585	[0.553, 0.616]
Imhof full pipeline	5 within-tender features (CV, skew, kurt, spread, second-low)	0.888	[0.865, 0.911]
Combined			
FL flag + Imhof full	participation count + 5 bid features	0.955	[0.943, 0.967]
log(1+tenders_count) + Imhof full	participation count + 5 bid features	0.962	[0.954, 0.969]

Notes: 5-fold cross-validated AUC against 193 CADE-cobidder labels within the always-loser pool (N = 16,779). Random forests with 500 trees, otherwise default tuning. Bid-level features are aggregated to the firm level by averaging the per-tender within-tender statistic across each firm's tenders. The two top blocks compare alternative single-construct screens operating on different data envelopes; the bottom block reports the combined classifiers. The combined model gains 5–7 AUC points over either single construct, indicating non-redundant signal in the participation-only and bid-distribution information sources. The participation-only screen achieves discrimination comparable to the full Imhof–Wallimann pipeline at substantially lower data cost, which is the architectural claim of the paper: the screening stage operates on the data layer that survives most enforcement environments, and integrates with the bid-distribution forensic stage where microdata are available. We do not claim the participation-only screen *outperforms* the bid-distribution pipeline; the relevant comparison is at fixed data cost, where the two pipelines reach similar accuracy and the combination dominates.

Table 13: Architecture / Gatekeeper: Operational Precision and Bid-Microdata Footprint by Rule

Rule	k	TP	Precision	Recall	Lift	Bid-microdata firms needed
Award-layer only (FL log_tc)	250	40	0.160	0.207	9.68×	0
	500	66	0.132	0.342	7.99×	0
	1,000	97	0.097	0.503	5.87×	0
Bid-layer only (Imhof full)	250	50	0.200	0.259	12.10×	11,676
	500	73	0.146	0.378	8.83×	11,676
	1,000	97	0.097	0.503	5.87×	11,676
Joint scoring (FL + Imhof, single model)	250	74	0.296	0.383	17.91×	11,676
	500	112	0.224	0.580	13.55×	11,676
	1,000	142	0.142	0.736	8.59×	11,676
Sequential: FL → Imhof (Stage-1 keeps top 1000)	250	62	0.248	0.321	15.00×	1,000
	500	87	0.174	0.451	10.53×	1,000
	1,000	97	0.097	0.503	5.87×	1,000
Sequential: FL → Imhof (Stage-1 keeps top 2000)	250	65	0.260	0.337	15.73×	2,000
	500	96	0.192	0.497	11.62×	2,000
	1,000	131	0.131	0.679	7.93×	2,000
Sequential: FL → Imhof (Stage-1 keeps top 4000)	250	59	0.236	0.306	14.28×	4,000
	500	87	0.174	0.451	10.53×	4,000
	1,000	127	0.127	0.658	7.68×	4,000

Reference: evaluation pool size = 11,676 *always-loser firms with* ≥ 2 *priced bids per tender;* 193 *cobidder positives; base rate* = 0.0165.

Notes: Always-loser firms in the BEC sample, restricted to those for whom the within-tender bid-distribution moments (CV, SD, skewness, kurtosis, spread, min/max log, second-lowest distance — the seven Imhof full-pipeline features) can be computed from `bid_level_with_prices.parquet`. Five-fold cross-validated random forest scores; no out-of-sample temporal holdout in this table (see Online Appendix C for the temporal-holdout audit). The single-model joint scoring trains a random forest on $\log(1 + \text{tenders\_count})$ together with the seven Imhof features in one estimator; the sequential gatekeeper trains separately, using FL-stage rank to filter, then Imhof-stage rank inside the survivor pool. "Bid-microdata firms needed" is the number of always-loser firms for which per-bid *Valor Unitário Proposta* must be retrieved to compute the Imhof features under each rule. The award-layer-only rule needs zero bid-microdata extractions because $\log(1 + \text{tenders\_count})$ uses participation counts only. Lift is precision-at-k divided by the unconditional cobidder rate in the evaluation pool. The sequential pipeline does not raise top-k precision above joint scoring; its operational claim is that comparable precision is reachable while requesting bid microdata for only the Stage-1 survivor pool, not the full pool.
Source: `scripts/63_architecture_gatekeeper.R`.

draw on the full 2009–2019 window: the 5-fold cross-validated precision-at-k in Table 13 confounds prediction with classification. Table 14 restricts every per-firm feature to participation observed in 2009–2016, leaving the cobidder ground truth as anchored on post-sample CADE adjudications (which cannot be temporally truncated without discarding nearly all positives); the temporal-holdout pool is 8,257 always-losers with both train-window participation > 0 and complete train-window Imhof features (142 cobidder positives). Two patterns matter for the architectural reading. Precision compresses asymmetrically across the four rules: the award-layer screen alone loses 47% at top-500 (consistent with the precision-at-k audit on FL alone in Online Appendix C), joint scoring loses 24%, and the sequential gatekeeper loses 9%. In absolute true-positive count, the sequential gatekeeper matches or exceeds joint scoring under temporal holdout: at top-500 the sequential rule recovers 87 adjudicated cobidders versus 85 for joint scoring, and at top-1,000 it recovers 114 versus 111. Stage-2 Imhof moments derived from train-window bid distributions are more temporally stable than the FL participation rank applied across the full pool, so restricting bid-microdata interrogation to a Stage-1 survivor pool both compresses the data envelope and preserves out-of-time TP capture better than running joint scoring on the full pool. The 83% data-envelope reduction does not trade off against out-of-time recall.

Taken together, the three findings identify the sequencing structure the paper proposes. Under incomplete operational observability, the deployable architecture is an award-layer screening stage running on operational data, followed by a bid-layer forensic stage running the bid-distribution moment battery on the candidates the screening stage flags, with bid microdata recovered through administrative request only for the screening-flagged subpool. Jurisdictions whose operational layer carries award records but not per-bidder bid amounts can deploy the screening stage immediately. Jurisdictions whose forensic-recoverable layer is also analytically queryable gain the forensic stage on top, with the screen already constraining the pool of cases that warrant the forensic step. Reforms that mandate operational bid-microdata archival (§3) are best read as expanding the forensic stage rather than substituting for the screening one—a reading that reverses the standard policy intuition that operational bid-archival is a precondition for proactive

Table 14: Architecture / Gatekeeper Under Temporal Holdout: Features Trained on 2009–2016, Evaluated on the Same Cobidder Labels

		Precision @ k		Recall @ k		TP @ k	
Rule	k	In-samp.	T. holdout	In-samp.	T. holdout	In-samp.	T. holdout
Award-layer only	250	0.160	0.076	0.207	0.134	40	19
	500	0.132	0.070	0.342	0.246	66	35
	1,000	0.097	0.066	0.503	0.465	97	66
Bid-layer only (Imhof full)	250	0.200	0.156	0.259	0.275	50	39
	500	0.146	0.112	0.378	0.394	73	56
	1,000	0.097	0.082	0.503	0.577	97	82
Joint scoring (FL + Imhof)	250	0.296	0.244	0.383	0.430	74	61
	500	0.224	0.170	0.580	0.599	112	85
	1,000	0.142	0.111	0.736	0.782	142	111
Sequential FL → Imhof, $K_1 = 2{,}000$	250	0.260	0.248	0.337	0.437	65	62
	500	0.192	0.174	0.497	0.613	96	87
	1,000	0.131	0.114	0.679	0.803	131	114

Notes: Temporal-holdout column restricts every per-firm feature to participation observed in 2009–2016 before random-forest scoring. The cobidder ground truth (193 always-losers co-bidding with direct CADE defendants) is anchored on adjudications post-sample and cannot be temporally truncated without discarding nearly all positives, so the holdout applies to features rather than labels. Evaluation pool: in-sample is the 11,676-firm pool of Table 13; temporal-holdout pool is the 8,257 always-losers with both train-window participation > 0 and complete train-window Imhof features. The pattern parallels the precision-at-k audit on FL alone reported in Online Appendix C: precision compresses by roughly half, recall by less, lift by roughly half. The relative ordering of the four rules is preserved.
Source: `scripts/64_gatekeeper_temporal_holdout.R`.

cartel screening.

11. Limitations

11.1. Identification scope

Identification is descriptive throughout. The two design-based strategies that would extract a causal estimate return null at the available bandwidths. A sharp regression discontinuity at the convite/pregão statutory caps is gated by the absence of bunching in the McCrary density test, and a difference-in-differences exploiting the 2018 cap raise is gated by the failure of parallel trends. The within-PBU coefficient 6.4% on the broad sample reverses to −9.72% under overlap-restricted matching and −30.67% under propensity-score trimming; the segment-level decomposition in §7.2 attributes the aggregate reversal to the population-weighted average of a positive Q4-tender-value coefficient and three negative Q1–Q3 coefficients, all robust to design choice. The sensitivity bounds (Cinelli and Hazlett (2020) robustness value 17.5%, Oster (2019)

$\hat{\delta} = 261.6$) discipline the broad-sample association against selection on observables already absorbed by item, year, and procuring-unit fixed effects, but do not extend to selection on unobservables that the overlap restriction surfaces. The +3.6% – + 7.7% headline range is a conditional descriptive association, not an estimate of cover bidding's causal effect on prices, and the 12.6× extreme-quartile gradient is observationally consistent with the framework's detection-cost comparative static without identifying the causal channel. Under the screening framing of §5.1, this descriptive scope is sufficient: the screening object is the loser-side participation footprint where the cartel deploys it, not a counterfactual treatment effect on prices.

11.2. Construct scope

The construct flags loser-side participation patterns, not cartel members. The validation discipline established in §6—cobidders inside the always-loser stratum as the structurally appropriate object given the data layer, direct-defendant AUC of 0.49 as the design's empirical signature—defines the limit. An enforcement program that requires firm-level membership identification will need a different statistic. Within the always-loser stratum, the participation footprint is observationally compatible with three non-competitive readings: cover bidding under cartel rotation, capacity-transfer signaling for subcontracting or affiliate consortia, and market-intelligence accumulation. The framework of Online Appendix A is built around the cover-bidding reading, which is the reading the strongest heterogeneity dimensions match (modal asymmetry in §8.2, detection-regime gradient in §7.3); the bid-level bridge in §6.2 narrows it further to the credible-cover-bidding variant. The other two readings are not refuted, and the empirical work is organized around the cover-bidding reading without claiming it as unique. Two finer mechanism predictions—a cell-heterogeneity grid and a discrete first-time-FL imprint—move in the predicted direction but the data do not adjudicate them under disciplined estimation (Online Appendix C); neither is load-bearing for the screening interpretation. Bid-aggressiveness against the reference price remains an untested prediction of the cover-bidder type: *Valor Unitário Referência* is populated for only 22% of convite and 32% of pregão losing bids, leaving most firms without per-firm coverage;

that part of the bridge is reserved for jurisdictions with universal reference-price archival.

The construct's behavior under deliberate adversarial adaptation is bounded but not invariant (Online Appendix G, Table F.1). The screening statistic is resilient to periodic rotation of cover bidders and to occasional wins evading the bright-line filter (AUC essentially unchanged), degrades modestly under CNPJ splitting (-3.7 points), and is substantially vulnerable to threshold-aware capping (-12.7 points); combining capping with rotation drives AUC from 0.928 to 0.632. The threshold rule is the most exposed margin, and any application of the construct requires periodic recalibration alongside integration with bid-distribution methods where those data are available (§10).

11.3. External validity and portability

The empirical evidence comes from a single jurisdiction and a single time window. The construct's portability rests on three structural features explicit in BEC and identifiable in any candidate jurisdiction: a commodity-heavy item composition that supports within-product-code price comparisons; an electronic platform with publicly auditable participant lists; and an institutional asymmetry between winner and loser sides that makes persistent zero wins informative about strategic non-competitive participation. The architectural test in §10 sharpens the portability claim: award-only enforcement environments—those whose operational layer preserves participants and winners but not bid-by-bid records—are the construct's natural deployment environment, real-time-electronic environments (pregão-style auctions) are the regime in which the screening signal is sharpest, and sealed-bid environments without real-time interaction are the scope-restricted boundary rather than the natural deployment one. Three settings the paper does not test: the federal Brazilian procurement registry (ComprasNet) and other state-level registries preserve the same award-record envelope as BEC and are accessible through the same legal regime; non-commodity settings (services, infrastructure, public works) are within the framework's logical scope but the within-item identifying variation that underwrites the headline price gap is not equally available; and time periods outside the sample window are untested. The 2021 reform (Lei 14.133/2021) consolidates pregão-style auctions as the institutional default—combined with the modal contrast in §8.2, the regulatory direction

is favorable to portability rather than adversarial.

12. Conclusion

Cheap data can triage cartel investigations before costly forensics. On São Paulo's BEC (2009–2019), routing forensic interrogation through a frequent-loser flag built on contract-award records cuts the bid-microdata pool the forensic stage must work on by 83% while still flagging 131 of 193 adjudicated cartel cobidders. Using only contract-award data, the flag matches the seven-feature Imhof–Wallimann bid-distribution pipeline that requires bid microdata, and adds non-redundant signal when the two are combined (+0.035 AUC, DeLong $p = 0.014$). Built on 2009–2016 participation only, the screen prospectively flags adjudicated cobidders in 2017–2019 at firm-level AUC 0.864. The screen triages; it does not adjudicate cartel membership—it ranks loser-side firms and procurement environments for costly bid-layer interrogation, not cartel members for legal sanction. A simple separating-equilibrium argument with cover bidders motivates endogenous loser-side participation as the ranking primitive on the award layer; the binary median + 1.5×IQR rule on always-losers operationalizes it.

The screening and forensic stages are informational complements, not substitutes. They discriminate the same cartel-adjacency target on different observability layers and carry non-redundant signal when both are accessed; the screening stage functions as a credible Stage-1 gatekeeper for the forensic stage. Reforms that mandate operational bid-microdata archival expand the forensic stage, not the screening one. Jurisdictions whose operational layer carries award records but not per-bidder bid amounts can deploy the screening stage on data they already maintain, which reverses the standard policy intuition that operational bid-archival is the rate-limiting precondition for proactive cartel screening (Becker, 1968; Stigler, 1970; Baker, 2003; Harrington, 2008).

The paper stops at the screening stage by design. Award-record data does not contain what membership identification requires, and cobidders are the validation object the data layer supports. Adversarial adaptation degrades the screen's discrimination predictably (Online Appendix G); replication rests on three diagnostic conditions observable *ex ante* in any candidate jurisdiction—a winner/loser participation registry, an electronic

platform, and an institutional asymmetry between the two sides of the bidding pool. The conceptual content travels: wherever an enforcement environment exposes the award layer routinely while reserving per-bidder bid amounts for forensic-recoverable access, screening should be sequenced before forensics, not the other way around.

References

Abrantes-Metz, R. M., L. M. Froeb, J. F. Geweke, and C. T. Taylor (2006). A variance screen for collusion. *International Journal of Industrial Organization 24*(3), 467–486.

Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review 100*(3), 724–762.

Athey, S., J. Levin, and E. Seira (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *Quarterly Journal of Economics 126*(1), 207–257.

Bajari, P. and L. Ye (2003). Deciding between competition and collusion. *Review of Economics and Statistics 85*(4), 971–989.

Baker, J. B. (2003). The case for antitrust enforcement. *Journal of Economic Perspectives 17*(4), 27–50.

Baldwin, L. H., R. C. Marshall, and J.-F. Richard (1997). Bidder collusion at forest service timber sales. *Journal of Political Economy 105*(4), 657–699.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Chassang, S. and J. Ortner (2022). Robust screens for non-competitive bidding in procurement auctions. *Econometrica 90*(1), 315–346.

Cinelli, C. and C. Hazlett (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B 82*(1), 39–67.

Clark, R., J.-F. Houde, and J. Kastl (2021). The dynamics of bidder collusion: Evidence from a long-lived cartel. *Journal of Political Economy 129*(8), 2353–2391.

Conley, T. G. and F. Decarolis (2016). Detecting bidder groups in collusive auctions. *American Economic Journal: Microeconomics 8*(2), 1–38.

Coviello, D. and S. Gagliarducci (2017). Tenure in office and public procurement. *American Economic Journal: Economic Policy 9*(3), 59–105.

Decarolis, F., R. Fisman, P. Pinotti, and S. Vannutelli (2025). Rules, discretion, and corruption in procurement: Evidence from italian government contracting. *Journal of Political Economy Microeconomics 3*(2), 213–254.

Decarolis, F., L. M. Giuffrida, E. Iossa, V. Mollisi, and G. Spagnolo (2020). Bureaucratic competence and procurement outcomes. *Journal of Law, Economics, and Organization 36*(3), 537–597.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics 44*(3), 837–845.

Harrington, J. E. (2008). Detecting cartels. In P. Buccirossi (Ed.), *Handbook of Antitrust Economics*, pp. 213–258. MIT Press.

Huber, M. and D. Imhof (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization 65*, 277–301.

Imhof, D. (2019). Detecting bid-rigging cartels with descriptive statistics. *Journal of Competition Law & Economics 15*(4), 427–467.

Imhof, D., Y. Karagök, and S. Rutz (2018). Screening for bid rigging—does it work? *Journal of Competition Law & Economics 14*(2), 235–261.

Kawai, K. and J. Nakabayashi (2022). Detecting large-scale collusion in procurement auctions. *Journal of Political Economy 130*(5), 1585–1629.

Marshall, R. C. and L. M. Marx (2012). *The Economics of Collusion: Cartels and Bidding Rings*. Cambridge, MA: MIT Press.

OECD (2019). Government at a glance 2019. Technical report, Organisation for Economic Co-operation and Development, Paris.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Pesendorfer, M. (2000). A study of collusion in first-price auctions. *Review of Economic Studies 67*(3), 381–411.

Porter, R. H. and J. D. Zona (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy 101*(3), 518–538.

Porter, R. H. and J. D. Zona (1999). Ohio school milk markets: An analysis of bidding. *RAND Journal of Economics 30*(2), 263–288.

Sánchez Graells, A. (2019). 'screening for cartels' in public procurement: Cheating at solitaire to sell fool's gold? *Journal of European Competition Law & Practice 10*(4), 199–211.

Schurter, K. (2020). Identification and inference in first-price auctions with collusion. Working Paper, Pennsylvania State University.

Stigler, G. J. (1970). The optimum enforcement of laws. *Journal of Political Economy 78*(3), 526–536.

Wallimann, H., D. Imhof, and M. Huber (2023). A machine learning approach for flagging incomplete bid-rigging cartels. *Computational Economics 62*(4), 1669–1720.