Screening for Bid Rigging with Frequent Losers in Public Procurement*

Darcio Genicolo-Martins† Paulo Furquim de Azevedo‡

Abstract

Minimum-bidder rules in public procurement give cartels a reason to deploy cover bidders—firms that show up repeatedly with no intention of winning. We exploit this behavioral footprint to build a participation-based screen: *frequent losers* (FL), firms that never win yet bid abnormally often. The screen requires only win/loss records, making it deployable in settings where bid microdata are unavailable. In São Paulo's electronic procurement (4.5 million tender-items, 2009–2019), it achieves AUC = 0.94 against competition-authority co-participation (0.54 within participation-count bands, confirming that volume contributes substantially to the unconditional figure), complements bid-level tools (correlation 0.06), and flags environments with 3.6–7.7% higher conditional prices. The price association is concentrated in competitive markets and where cover bidding is voluntary rather than forced by the minimum-bidder constraint—suggesting strategic deployment, not mere rule compliance. We propose a three-stage enforcement pathway (screen, triage, investigate) that allocates investigative resources toward the procurement environments where the association is strongest and oversight weakest.

Keywords: bid rigging, cover bidding, public procurement, cartel screening, frequent losers

*This version: April 2026.

†Corresponding author. INSPER, São Paulo, Brazil. Email: darciogm1@insper.edu.br.

‡INSPER, São Paulo, Brazil.

JEL: D44, D73, H57, K21, L41

1 Introduction

Procurement authorities need triage tools that work with minimal data. Most bid-rigging screens rely on granular bid amounts—within-tender variance, kurtosis, coefficient of variation (Imhof, 2019; Huber and Imhof, 2019)—yet many procurement systems record only who bid and who won. When cover bidders coordinate their bids tightly around the designated winner, dispersion *falls* and these screens lose power—a detection paradox in which the best-coordinated cartels look the most competitive. The gap is practical: how should an agency with participation records but no bid microdata decide where to look?

We propose a screen built on a simpler footprint. Cover bidding requires firms to show up and lose—repeatedly. A firm that bids on 50 tenders over a decade without winning once is either spectacularly unlucky or doing something that warrants scrutiny. We call these firms *frequent losers* (FL) and show that this participation anomaly—observable from win/loss records alone—flags procurement environments with systematically higher prices and strong proximity to known cartel activity. A conceptual framework (Section 4) organizes the intuition by distinguishing two cover-bidding regimes, but the screen's value rests on its empirical performance, not on the framework.

The institutional logic is direct. Brazil's sealed-bid procurement (convite) requires at least three bidders; cartels exploit this rule by deploying cover bidders who show up and lose. The minimum-bidder rule creates the incentive; the FL screen detects the behavioral response; a three-stage enforcement pathway directs scarce resources toward the highest-risk environments. Tellingly, the price association is 3.8% for convite but 9.3% for pregão (electronic auction, no minimum-bidder rule). Where the law forces participation, some cover bidding is mere compliance noise. Where no rule compels entry and firms still show up to lose, the signal is cleaner and the premium more than doubles.

We test this on São Paulo's Bolsa Eletrônica de Compras (BEC), which covers 4.5 million tender-items and 40 million bids over 2009–2019. Among 41,000 participating firms, 16,843 never win a single tender; a simple IQR-based outlier rule identifies 2,735 whose participation frequency defies any profit-maximizing logic. The screen is a first-stage triage device—it flags suspicious environments, not guilty firms.

Validated against CADE (Brazil's competition authority) cartel co-participation, the screen achieves AUC = 0.94—above a random forest trained on seven bid-level features (AUC = 0.90; DeLong $p = 0.04$) and a coarser CV-based proxy (AUC = 0.79; Section 7.3). Volume contributes substantially to this performance—within participation-count bands, the AUC drops to 0.54—but a count-matched permutation test confirms the FL classification adds signal beyond volume alone (2.2×, $p < 0.001$; Section 6), and the zero-win condition is the active ingredient conditional on volume.

FL-flagged tenders show 3.6–7.7% higher conditional prices in within-item, within-year, within-purchasing-unit comparisons—a range spanning cross-fitting (3.6%), OLS with high-dimensional fixed effects (6.4%), and matching (CEM: 7.7%; IPW: 5.5%). After absorbing item fixed effects, the raw FL–non-FL gap collapses from 2.43 to 0.050 log points (overlap 0.978; Table C.1, Appendix C), so the regressions compare observationally similar tenders within tight product categories. An omitted confounder would need to explain 17.5% of residual variation in both FL presence and prices to eliminate the coefficient (Cinelli and Hazlett 2020).

We then ask whether the screen behaves as coordinated cover bidding would predict. Five independent diagnostics point in the same direction: the price association concentrates in competitive markets; Bajari–Ye tests reject exchangeability of FL and non-FL bid residuals; FL–winner pairs recur well beyond chance; the coefficient is larger where cover bidding is voluntary; and in a horse-race regression FL captures largely non-overlapping information relative to bid-level screens (correlation 0.06). No single test is decisive, but the joint pattern is difficult to reconcile with a benign explanation.

1.1 Contributions

The paper makes three contributions. First, we introduce a participation-based screen that runs on win/loss records alone—no bid microdata, no enforcement priors, no supervised training—and that any procurement agency recording who bid and who won can deploy immediately. Second, the screen performs well empirically: AUC = 0.94 against competition-authority co-participation (surviving exact count matching at 2.2×), a 3.6–7.7% conditional price premium, and near-orthogonality with bid-level tools (correlation 0.06). Third, the price association concentrates where cover bidding is a strategic choice (7.6% voluntary premium) and where oversight is weakest (12.5× gradient across purchasing-unit size quartiles)—tying the screen's empirical performance to the institutional features that make cover bidding profitable.

What we do not claim. The price association is not a causal estimate of collusion's effect on prices. The FL classification flags suspicious environments, not guilty firms. The diagnostics are consistent with coordinated cover bidding but do not prove it. What the paper contributes is the screen itself—how to build it, how it performs against external enforcement data, and why the diagnostic pattern supports its use as a first-stage investigative tool.

Scope. The remainder of the paper proceeds as follows. Section 2 reviews related work. Section 3 describes the data and FL definition. Section 4 presents the conceptual framework. Section 5 lays out the empirical strategy. Section 6 reports the CADE consistency check. Sections 7–8 present results and diagnostics. Section 9 reports robustness checks. Section 10 discusses limitations. Section 11 proposes a three-stage enforcement pathway and discusses portability.

2 Related Literature

Detecting bid rigging requires knowing whom to suspect—but the best tools for establishing suspicion require already knowing whom to suspect. Breaking this circularity is the practical problem the FL screen addresses.

Bid-coordination tests and bidder classification. Bajari and Ye (2003) develop powerful tests for collusion—exchangeability of bid functions and conditional independence of bid residuals—but require an *ex ante* partition of bidders into "suspected colluders" and a "competitive fringe." No existing method provides this partition from data alone. Subsequent work detects bid rotation (Porter and Zona, 1993, 1999), collusion in timber auctions (Baldwin et al., 1997), bidder groups from co-bidding patterns (Conley and Decarolis, 2016), and cartel pricing dynamics (Clark and Houde, 2014; Schurter, 2023). Conley and Decarolis (2016) are the closest precedent in data requirements, though they detect group structure from co-bidding patterns rather than flag individual firms exhibiting anomalous participation. Unlike their network-based approach, the FL screen starts from a behavioral anomaly (zero wins at high intensity) and uses co-bidding patterns only as a supporting diagnostic. In every case the partition is constructed *ex post* or by assumption. The FL classification provides a candidate for this partition: a data-driven, *ex ante* classification grounded in a participation anomaly that can feed into Bajari–Ye tests (Section 7.5).

Proactive screens for bid rigging. Imhof (2019) and Imhof et al. (2018) train ML classifiers on tender-level features and achieve high accuracy against known cartels; see also Abrantes-Metz et al. (2006), Harrington (2008), Chassang et al. (2022), Wallimann et al. (2023), and Caoui (2022), who uses structural estimation to quantify umbrella damages from bid-rigging cartels in procurement. All share two requirements: granular bid values and tender-level units of analysis. Many procurement systems, particularly in developing economies, record who bid and who won but not the bid amounts. The FL screen operates at the *firm* level using only participation and outcome records. This supports a two-stage triage workflow: FL flags prioritize environments for closer scrutiny, and bid-level analysis follows on the flagged subset. In a direct comparison on the same ground truth, FL achieves AUC = 0.94 versus 0.79 for an Imhof-style CV proxy, and the two screens are nearly orthogonal (correlation

0.06).[1]

Cover bidding. Cover bidders appear throughout the empirical record of prosecuted cartels (Porter and Zona, 1999; Pesendorfer, 2000; Asker, 2010; Marshall and Marx, 2012) and in theoretical comparisons of auction formats (Athey et al., 2011). In every case the cover bidder is identified *ex post*, after the cartel is caught. Our framework models cover bidding *ex ante*, distinguishing two regimes with different bid distributions and different vulnerabilities to detection. Under the coordinated regime, dispersion-based screens lose power—precisely the setting in which a participation-based screen may be most useful.

Enforcement costs and procurement design. Posner (1970) documented half a century ago that antitrust enforcement concentrates on easily detected conspiracies, leaving a large pool of cartels undetected. The economics of enforcement (Becker, 1968; Polinsky and Shavell, 2000) explain why: investigation is expensive, so screening tools that narrow the search have direct welfare value—even before any case is brought. Ghosal and Sokol (2014) trace the evolution of U.S. cartel enforcement through three institutional stages, each relying more heavily on data-driven detection; the FL screen fits this trajectory as a first-stage filter that needs no supervised training and no enforcement priors. Harrington and Chang (2015) show that leniency programs work best when paired with independent detection capacity—precisely the role a low-cost screen can fill. On the design side, minimum-bidder requirements can facilitate cover bidding (Athey et al., 2011), while electronic formats may reduce but not eliminate coordination (Bajari et al., 2009). Our results speak to this interaction: the FL–price association varies across procurement formats in the direction that differences in participation incentives predict.

Procurement in developing economies. The bid-rigging literature concentrates on developed economies, yet enforcement needs are most acute where investigative resources

[1] Our Imhof comparison uses a CV median split, a coarser proxy for the full implementation; see Section 7.3.

are scarce and bid microdata are unavailable (Sánchez Graells, 2019). Baranek and Titl (2024) find that political favoritism raises Czech procurement prices by ∼6%—a magnitude comparable to our FL–price association (6.4%)—illustrating how institutional distortions leave detectable price footprints. The FL screen's minimal data requirements make it directly portable to low-capacity enforcement settings. BEC's scale (4.5 million tender-items, 40 million bids, 11 years, two auction formats) provides a demanding test bed for a screen designed to operate with less. For enforcement agencies in such settings, the relevant question is not whether a screen identifies the mechanism but whether it reliably flags environments that warrant closer scrutiny with scarce investigative resources.

3 Data and Frequent Losers Definition

3.1 BEC Platform

The data come from the Bolsa Eletrônica de Compras (BEC), São Paulo's centralized electronic procurement platform. BEC serves 1,308 purchasing units (PBUs) and handles two modalities: *convite* (sealed-bid, minimum three bidders, under Brazil's general procurement statute Lei 8.666/93) and *pregão* (electronic reverse auction, no minimum-bidder rule). Under the framework, the convite minimum-bidder rule creates an incentive for cover bidding; under pregão, additional bidders may inflate apparent competition. The sample spans 2009–2019: 4.5 million tender-items, 40 million bids, 41,000 firms (Online Appendix).

3.2 Frequent Losers Definition

The screen is designed for simplicity: any procurement authority with a table of who bid and who won can run it. We define frequent losers (FL) in two steps:

1. **Always-losers**: firms with a zero win rate across all 2009–2019 tenders. 16,843 firms (∼41% of all BEC participants) never win a single tender—a high rate explained by

BEC's low participation cost (electronic, no travel) and, for convite, PBU officials inviting firms to meet the three-bidder minimum. The screen is not "never winning"; it is never winning *at intensities that are economically irrational under competitive entry*. FL firms are a small, extreme tail of this heterogeneous population: among 16,843 always-losers, only 2,735 (16%) participate at intensities difficult to reconcile with profit-maximizing entry. In the CADE validation (Section 6), always-loser status alone adds no signal beyond participation volume, but FL status adds substantial predictive power conditional on both volume and always-loser status—it is the combination that matters.

2. **IQR threshold**: among always-losers, we identify those with abnormally high participation counts using a median + 1.5 × IQR threshold on the distribution of tender participations. The primary justification is economic: at the resulting threshold of 14 or more tenders with a zero win rate, expected profit is strictly negative for any positive bidding cost ($E[\pi] = 0 \times \overline{V} \times \text{margin} - c < 0$; see Section 8). Sustained participation at this intensity is difficult to rationalize as competitive entry and consistent with the patterns documented in prosecuted cartels, where cover bidders receive compensation through side arrangements (Asker, 2010; Marshall and Marx, 2012). The threshold was fixed *a priori* at 1.5×IQR based on this economic rationale, before examining CADE data.[2] Post hoc, the ROC analysis in Section 5.2 shows that the data-driven optimal threshold (Youden's $J = 0.84$) corresponds to a multiplier of 1.45×IQR—close to the 1.5× rule we chose on economic grounds. This convergence is reassuring but was not used to select the threshold; robustness to the multiplier is reported in Section 9. The definition yields 2,735 FL firms.

The screening indicator losers_{igt} equals one if tender-item i in item group g at time t

[2]We use median + 1.5 × IQR rather than the standard Tukey rule (Q3 + 1.5 × IQR) because the participation distribution is highly right-skewed. This produces a lower threshold than Tukey's standard rule and thus a more inclusive FL set, consistent with a first-stage triage tool that errs on the side of flagging more firms for subsequent investigation.

has at least one FL firm among its participants, and zero otherwise. This is a screening construct, not a treatment assignment: it flags an environment-level participation anomaly, not a firm-level collusion status.

The FL classification is static: it pools participation over the full 2009–2019 window. Classifying FL separately in 2009–2013 and 2014–2019, fewer than 9% of early-period FL firms reappear as FL in the second half (108 of 1,240). A fate analysis of the 1,132 disappeared firms finds that most (63.7%) simply exit BEC, and 19.6% reappear with positive win rates (mean 13.3%)—consistent with normal entry and exit rather than cover-bidder rotation. However, new FL firms in 2014–2019 operate in the same purchasing units as disappeared FL firms at rates modestly above chance (83.5% overlap vs. 76% expected, $p < 0.001$), and 18% of new FL firms had positive win rates in the earlier period—a transition from winning to persistent losing that is harder to reconcile with competitive entry alone. The balance of evidence favors a mixture: most FL turnover reflects normal market dynamics, with a smaller component consistent with the rotation documented in prosecution records (Marshall and Marx, 2012). For empirical purposes, FL status is best read as an environment-level marker (a tender with suspicious participation) rather than a stable firm label. From an enforcement standpoint, this is a feature: the screen flags procurement environments for closer scrutiny, not individual firms for sanction. Diagnostics of the classification—continuous measures, permutation placebos, and zero-win sensitivity—are reported in Section 7.1.

3.3 Sample Construction

The participation distribution among always-losers is heavily right-skewed (Online Appendix). The regression sample retains convite and pregão modalities with a recorded winner and item types with at least one FL-present tender. This last restriction conditions on the screening indicator: items that never attract FL firms are excluded, ensuring controls come from item markets where FL firms could plausibly participate. The unrestricted sample (Online Appendix) yields a nearly identical coefficient, indicating that the restriction is empirically innocuous. The resulting sample contains 1,654,447 tender-items across 18,783 item types (defined at the finest product-code level available in BEC, not aggregated to two-digit groups), of which 15,101 (82%) have both FL-present and FL-absent tenders and thus contribute to within-item identification. FL firms participate in 4.8% of tender-items.[3] Table 1 reports descriptive statistics.

Table 1: 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 $\geq$1 FL tender. FL = frequent losers defined by median + 1.5$\times$IQR threshold.

3.4 Structural Estimation Sample

The structural model requires bid-level data: 40 million bids from 41,000 firms, with bid amount, winning price, firm identity, and FL classification. FL losing bids exhibit a mean log-spread of 0.83 (SD = 1.19) above the winning price, versus a genuine-bid SD of 1.64. The ratio $\hat{\sigma}_c/\hat{\sigma}_g = 0.72$ is consistent with Regime 2 (coordinated cover bidding).

3.5 CADE Sample

The CADE consistency check uses a temporal split to prevent information leakage: FL status is defined on 2009–2014 data and evaluated against 2015–2019 enforcement outcomes. The

[3] Sample sizes vary slightly across dependent variables: 1,654,401 for price regressions (46 observations lack a recorded negotiated price), and 1,653,156 for non-FL firm counts (291 tenders where all participants are FL, making the non-FL count undefined).

CADE data comprise 65 cartel cases, of which 47 convicted firms match BEC identifiers. Among 2,735 FL firms, 193 (7.1%) co-participate with CADE-convicted cartelists—3.5 times the expected rate ($p < 0.001$; Section 6).

Section 4 develops a conceptual framework that organizes the diagnostic implications assessed in the empirical analysis; the screen's value rests on its empirical performance, not on the framework.

4 Conceptual Framework

A good cover bidder is invisible to the tools designed to catch him. If he coordinates his bid tightly enough with the cartel's designated winner, variance-based screens read the tender as competitive. This section formalizes that intuition: when does coordination defeat detection, and what trace does the cover bidder leave behind? The answer generates five testable implications. Full assumptions, proofs, and the structural likelihood are in Appendices A and B.

Consider a cartel that controls a designated winner and deploys $m \geq 0$ cover bidders (FL firms), each bidding above the winning bid b^*. The optimal m^* falls with detection probability θ_k and per-bidder cost c_1. Under convite, the minimum-bidder rule ($\underline{n} = 3$) can force participation even when market conditions would not justify it, mixing mandatory and voluntary deployments and diluting the FL–price signal (Section 8).

Two information structures generate two cover-bidding regimes. Under **Regime 1** (complementary), cover bids scatter uniformly above b^*, raising within-tender dispersion—and handing variance-based screens an easy target. Under **Regime 2** (coordinated), cover bids follow a truncated normal centered near b^*, *lowering* dispersion. This is the paradox at the heart of the paper: the better the cartel coordinates, the more its tenders *look competitive* to dispersion-based tools. A participation-based screen sidesteps the trap entirely—it does not care how cover bids are distributed, only that cover bidders must show up.

The model permits **strategic complementarity**: if the marginal return to cover bidding is higher in competitive tenders (many genuine bidders), then m^* *increases* in n. The calibrated model confirms this pattern ($\hat{\gamma} = 0.69 > 0$; Figure 1 and Appendix B), indicating that cartels deploy more cover bidders precisely where genuine competition is strongest. This generates a **market-selection implication**: FL presence should concentrate in competitive markets (low HHI).

Table 2 maps the framework's diagnostic implications to the empirical tests reported below.

Table 2: Diagnostic implications and empirical tests

#	Diagnostic implication	Diagnostic test	Section
D1	FL associated with higher prices	Conditional price comparison	5.1
D2	FL adds to, not displaces, genuine bidders	Non-FL bidder count	5.1
D3	Coordinated regime: $\sigma_c < \sigma_g$	BIC model selection	5.3
D4	FL association ↓ in HHI	Network split	5.5
D5	FL residuals non-exchangeable	Bajari–Ye KS + pairwise	5.4

These are predictions, not conclusions. The next sections confront them with data: does FL presence actually predict higher prices? Do FL bids really cluster where the framework says they should? And does the screen find cartels that existing tools miss? Illustrative welfare calculations appear in Online Appendix.

5 Empirical Strategy

A screening tool does not need a causal estimate—it needs a reliable signal. We build the empirical case in three tiers, in descending order of evidentiary weight.

Tier 1: Conditional price association (Section 5.1). Do FL-present tenders have higher prices, after absorbing item, year, and purchasing-unit heterogeneity? OLS with high-dimensional fixed effects, cross-fitting, and matching all say yes.

Tier 2: Detection performance (Section 5.2). Does the screen find real cartels? ROC

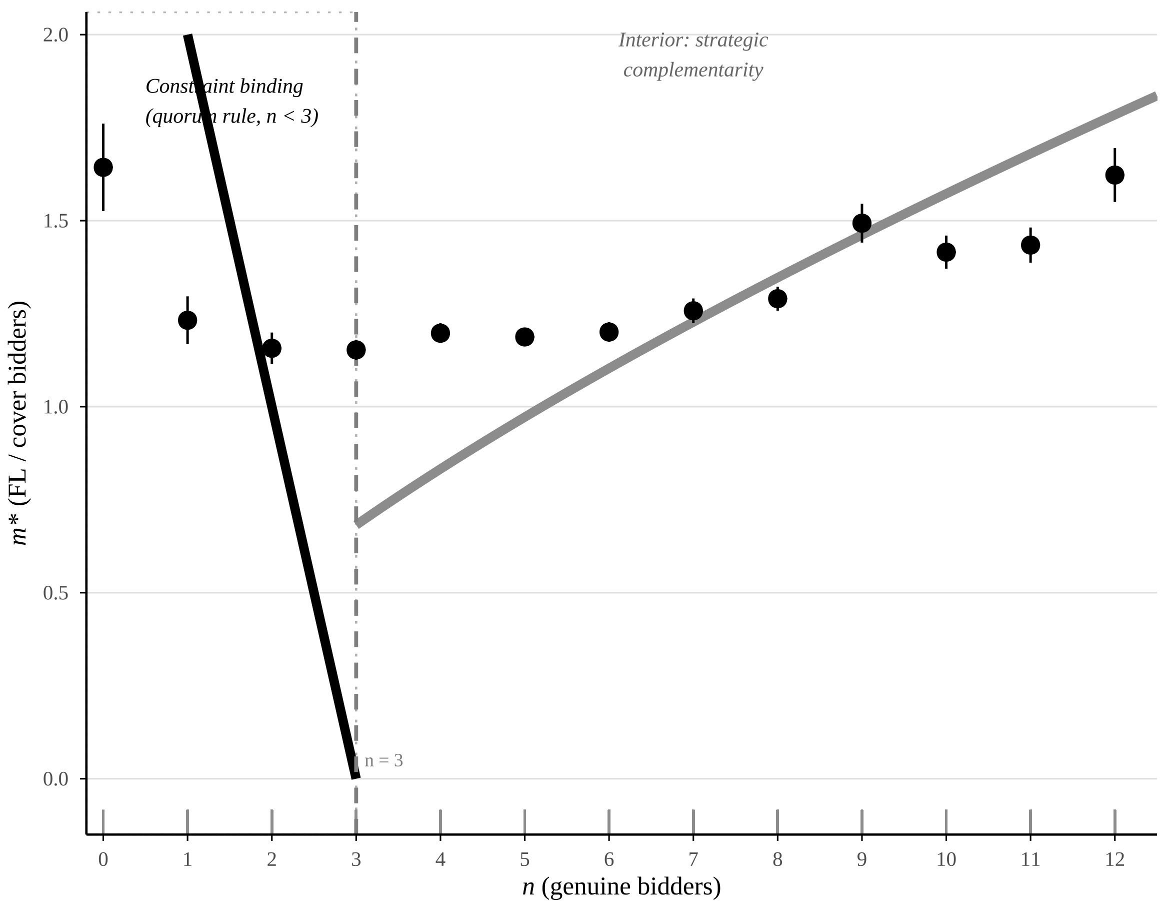

Figure 1: Optimal number of cover bidders m^* as a function of genuine bidders n. Dotted region ($n < 3$): constraint-binding corner solution under the minimum-bidder rule ($\underline{n} = 3$, Lei 8.666/93). Solid curve: calibrated interior solution ($\hat{\gamma} = 0.69$, strategic complementarity). Points: empirical binned means with 95% CIs.

analysis against CADE convictions, benchmarked against bid-level screens.

Tier 3: Supporting diagnostics (Sections 5.3–5.5). Does the screen behave the way coordinated cover bidding predicts? Structural estimation, Bajari–Ye tests, and network-split heterogeneity probe the mechanism without claiming to identify it.

The hierarchy matters: the screen's value as a triage tool rests on Tiers 1 and 2. Tier 3 diagnostics strengthen the economic rationale but are not necessary for the screen to be useful.

5.1 Conditional Price Association

OLS baseline. The baseline specification regresses tender outcomes on FL presence:

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

where y_{igt} is the outcome for tender-item i in item group g at time t and purchasing unit k; $\text{losers}_{igt} \in \{0, 1\}$ indicates FL presence; $\mathbf{x}_{igt}$ includes a convite indicator; α_g, λ_t, and γ_k are item, year, and PBU fixed effects; and ε_{igt} is the error term clustered at the item level.

The OLS coefficient $\hat{\beta}$ captures a conditional association—the average difference in outcomes between FL-present and FL-absent tenders within the same item type, year, and purchasing unit. Under the framework's first diagnostic implication (Table 2), $\hat{\beta} > 0$ for prices. Item, year, and PBU fixed effects absorb time-invariant cross-market heterogeneity; the residual threat is time-varying within-market confounders (Section 10). We assess the sensitivity of $\hat{\beta}$ to such confounders using the Cinelli and Hazlett (2020) framework, which yields a robustness value of $RV_{q=1} = 17.5\%$: an unobserved confounder would need to explain at least 17.5% of the residual variation in both FL presence and log prices to reduce $\hat{\beta}$ to zero.

Matching estimators. To address potential selection on observables, we estimate the price association using two matching approaches: coarsened exact matching (CEM) on year, procedure type, item group, and PBU size quartile ($\hat{\beta}_{\text{CEM}} = 0.077$, $N = 969{,}751$), and inverse probability weighting (IPW, $\hat{\beta}_{\text{IPW}} = 0.055$, $N = 830{,}194$). These estimates bracket the OLS baseline (0.064) and the cross-fit average (0.036), providing a non-parametric check that the price association survives reweighting on observables.

Instrumental variable (measurement-error diagnostic). The binary screening indicator contains misclassification noise: firms near the IQR threshold may not have collusion-related participation, and some collusion-related firms may escape the threshold. Under the plausible assumption that false positives (non-collusive firms above the IQR threshold) outnumber false negatives (collusive firms below it), binary misclassification attenuates the OLS coefficient toward zero. We use a leave-one-out FL supply instrument primarily to diagnose this attenuation:

$$Z_{kgt} = \sum_{j \neq k} \mathbf{1}[\text{FL firm active at PBU } j \text{ in group } g,\ \text{year } t] \tag{2}$$

The instrument exploits aggregate FL supply variation at other purchasing units in the same product market and year. We treat $\hat{\beta}_{\text{IV}}$ as a measurement-error diagnostic: it indicates the direction of bias in the binary FL indicator (attenuation toward zero) but not the magnitude of any underlying association, because balance tests reveal systematic imbalance across observables (Online Appendix). The IV is consistent with OLS understating the association; our primary range is cross-fit (0.036), OLS (0.064), and matching (0.055–0.077).

5.2 Detection Performance Validation

The detection assessment uses CADE cartel convictions as ground truth, with FL status defined on 2009–2014 data and evaluated against 2015–2019 enforcement outcomes (the temporal split described in Section 3.5).

ROC analysis. We vary the IQR multiplier from 0.5 to 5.0 in steps of 0.1, producing 46 FL thresholds. For each, we compute the true-positive and false-positive rates against CADE co-participation and trace the ROC curve. The optimal multiplier maximizes Youden's J (TPR − FPR), connecting the IQR rule to a formal detection-theoretic criterion.

Benchmarking. We benchmark FL against Imhof (2019)-style bid-level features on the same CADE ground truth, and include both screens in a horse-race regression to test orthogonality (Section 7.3).

5.3 Structural Diagnostic

As a Tier 3 supporting diagnostic, we estimate a two-stage mixture model on bid-level data and use BIC to select between the two cover-bidding regimes. The test distinguishes which *form* of cover bidding best fits the data; it does not test cover bidding against a competitive baseline. Details are in Appendix B.

5.4 Bajari–Ye Bid Coordination Tests

We implement the Bajari and Ye (2003) tests using corrected first-stage residuals from the bid equation:

$$\log b_{jt} = \mathbf{x}_j' \boldsymbol{\phi} + \alpha_i + \lambda_t + u_{jt} \tag{3}$$

where $\mathbf{x}_j$ includes firm size, firm age, and CNAE sector dummies. We exclude n_{bids} from the first stage: it is jointly determined with bid levels and mechanically increased by FL presence (Bajari and Ye, 2003, p. 978).

Using the residuals $\hat{u}_{jt}$, we test (i) *exchangeability* (KS test comparing FL and non-FL residual distributions; Diagnostic 2) and (ii) *conditional independence* (mean pairwise product of FL residuals within tenders; bootstrap with 1,000 iterations). A fake-groups placebo splits non-FL firms randomly within each tender; the economically relevant statistic is the FL–non-FL pairwise *difference*, which isolates excess FL correlation beyond common tender-level shocks. We also report a tender-FE variant as a conservative specification absorbing all tender-level variation.

5.5 Network-Split Heterogeneity Test

To assess the framework's market-selection implication—that cover bidding concentrates in competitive markets—we split FL firms into two subgroups based on co-bidding network characteristics.

Winner HHI classification. For each FL firm j, we compute the HHI of winning firms across all tenders in which j participates. Firms above the sample-median HHI are classified as *concentrated-market*; those below as *competitive-market*. We interact losers_{igt} with the market-structure indicator in Equation (1).

Before presenting the price regressions, we check the FL classification against external enforcement data (Section 6). This provides an initial check that FL presence flags economically meaningful environments—a necessary condition for the screen to be useful as a triage tool and for the price association to be informative.

6 External Consistency Check: CADE Cartel Convictions

Before turning to prices, we ask a simpler question: do FL firms show up near known cartels? CADE (Conselho Administrativo de Defesa Econômica) convicted 65 procurement cartel cases during 2009–2019; 47 convicted firms match to BEC via CNPJ. Among the 2,735 FL firms, 193 (7.1%) co-participate with at least one convicted firm—3.5 times the expected rate under a permutation null stratified by participation quartile ($p < 0.001$; 1,000 iterations). That 92.9% show no CADE overlap is unsurprising: cartel prosecution is rare (Connor, 2007), and convictions represent only the visible tip. The screen is designed to reach the rest.

Interpretation. Co-participation measures proximity to suspicious procurement environments, not cartel membership. Higher participation volume mechanically raises co-participation opportunities, and the unconditional excess is partly driven by this channel. Within the always-loser population, conditioning on participation count renders FL and non-FL overlap rates indistinguishable ($p = 0.93$; Section 7.3)—as expected, since FL *is* the high-participation tail of always-losers. The relevant question is whether the FL classification adds signal beyond volume alone.

Disentangling volume from the zero-win condition. Three tests address this. First, a logistic regression of CADE co-participation on log(participation count), always-loser status, and FL status across all 41,444 BEC firms shows that always-loser status alone adds nothing beyond volume (LR $\chi^2 = 0.2$, $p = 0.64$), but FL adds substantial signal conditional on both volume and always-loser status (LR $\chi^2 = 46.3$, $p < 10^{-6}$). Once FL is included, always-loser status turns *negative* (-0.354, $p < 0.001$): non-FL always-losers are *less* likely to co-participate with CADE firms than non-always-losers with the same volume. It is the combination of zero wins and abnormally high participation—not either component alone—that predicts CADE proximity.

Second, a permutation test matching each FL firm to a non-FL always-loser with the same participation count (± 2 tenders; 1,000 iterations) yields an FL co-participation rate of 7.2% versus a matched mean of 3.4% ($2.2\times$, $p < 0.001$). The signal survives exact count matching, ruling out a purely mechanical volume explanation.

Third, the stratified permutation test ($3.5\times$, $p < 0.001$; 1,000 iterations preserving participation quartiles) confirms the association is not an artifact of the stratification granularity.

The AUC reflects the screen's ability to identify high-participation zero-win firms that operate near known cartel environments. Volume contributes to the signal—as it should, since abnormally high participation is the behavioral anomaly the screen exploits—but does not exhaust it. For enforcement purposes, this is the relevant target: a triage rule that flags environments warranting closer scrutiny, not a tool that adjudicates firm-level liability.

Dual pattern. Most CADE-convicted firms in BEC are *winners*—the designated recipients of cartel rents. Yet three convicted firms are themselves classified as frequent losers (Table 3): they participated across multiple item groups and years without ever winning.

Table 3: CADE-convicted firms classified as frequent losers

Firm	Cartel	CADE process	Tenders	Wins
Sol Tecnologia	Solar water heaters	08012.001273/2010-24	84	0
Nova Esperança Locadora	School transport.	08700.005876/2019-85	97	0
Jofran Comércio	Trash bags (Op. Colludium)	08700.005789/2015-02	65	0

These three firms are definitional rather than independent validation: they are convicted cartel members who exhibit the FL pattern, which is expected given the screen's design rationale. The relevant validation comes from the 190 non-convicted FL firms that co-participate with CADE cartelists at rates exceeding volume-matched controls (2.2×, $p < 0.001$).

Co-bidder analysis. FL firms appear in 5.6% of CADE-firm tenders versus a 2.1% baseline (2.6×). For individual convicted firms the rate is much higher—up to 69% for Mayfran (school transportation). Excluding the three directly convicted FL firms and their 246 tender-items leaves $\hat{\beta}$ unchanged at 0.064.

Beyond known cartels. Dropping all 31,447 CADE-involved tender-items barely moves the coefficient: $\hat{\beta} = 0.062$ ($N = 1{,}622{,}954$) versus 0.064 in the full sample (Table C.3). The price association is not driven by tenders CADE has already flagged—which is precisely the screen's practical value. It reaches markets that existing enforcement records do not. The question now is whether this association holds up under scrutiny.

7 Results

We evaluate the screen in four stages. First, classification diagnostics (Section 7.1) verify that the FL rule captures a real behavioral pattern, not a threshold artifact. Second, conditional price comparisons (Section 7.2) measure the association between FL presence and tender outcomes. Third, detection performance (Section 7.3)—the paper's primary validation—tests the screen against external enforcement data. Fourth, supporting diagnostics (Sections 7.4–7.7) ask whether the screen behaves as coordinated cover bidding would predict. A useful benchmark throughout: after absorbing item fixed effects, the raw FL–non-FL price gap collapses from 2.43 to 0.050 log points (overlap 0.978). Within narrow product categories, FL-present and FL-absent tenders look nearly identical on observables. The conditional price association is 3.6–7.7% across four estimation approaches.

7.1 Classification Diagnostics

Any binary classification built on a statistical threshold invites the question: is the cutoff doing real work, or just splitting noise? Before turning to prices, we check. The zero-win condition is strict: relaxing it to include firms with win rates below 1% or 2% attenuates the price coefficient substantially (Online Appendix), so the bright line at zero is doing real work. A continuous measure—the log of the most active always-loser's participation count—yields a coefficient of 0.022 (SE = 0.005, $p < 0.01$) per log-point.[4] At the FL threshold ($\log 14 \approx 2.64$), the continuous implied value is $2.64 \times 0.022 \approx 5.8\%$—close to the binary OLS of 6.4%, indicating that the binary coefficient reflects the underlying continuous relationship evaluated at the threshold rather than an artifact of the cutoff. A permutation placebo that randomly reassigns the FL indicator across tenders (preserving the 4.8% FL-presence rate) produces coefficients centered at zero (mean 0.001, SD 0.003 over 20 replications), with no permutation reaching 0.064. The FL–price association reflects the specific allocation of FL

[4] The continuous measure correlates moderately with the number of always-losers in the tender ($r = 0.33$), so part of the coefficient may capture always-loser count rather than individual-firm intensity. Using the mean (not max) participation count reduces the correlation to 0.15 and yields qualitatively similar results.

presence across tenders, not chance.

7.2 OLS and Matching Results

Table 4 reports the baseline OLS estimates. FL presence is associated with significantly higher prices across all four specifications: 6.8% without PBU fixed effects, 6.4% with them, 9.3% for pregão, and 3.8% for convite. That PBU fixed effects barely move the coefficient matters: FL presence does not proxy for purchasing units that happen to pay more. The unrestricted sample (including items that never attract FL firms) yields a nearly identical coefficient, so the item-type restriction does not inflate the baseline.

Table 4: 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-squared	0.8790	0.8860	0.8821	0.8932
Item Dummies	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES
PBU Dummies	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.

FL-present tenders attract 0.19 log points more non-FL firms (Diagnostic 2; PBU FE specification), contradicting crowding-out and consistent with selection into competitive markets (Online Appendix).

Cross-fit estimate. FL status is defined on the same 2009–2019 window used for estimation, potentially creating a mechanical link between classification and outcomes. To break this link, we define FL using only odd years and estimate on even years, and vice versa. The cross-fit average is $\hat{\beta}_{\mathrm{CF}} = 0.036$ (3.6%), attenuated relative to full-sample OLS. A decomposition exercise—applying the odd-year FL list to the *full* sample—yields 0.043 (SE = 0.019), indicating that roughly two-thirds of the attenuation reflects classification noise from the smaller subsample (1,885–2,153 vs. 2,735 FL firms), not a mechanical bias in the baseline.[5] The cross-fit is valuable as a conservative estimate that eliminates any within-sample mechanical correlation; the OLS provides the full-sample baseline. The robust range is 3.6–6.4%.

Matching estimators. Coarsened exact matching (CEM: 0.077, $N = 969{,}751$) and inverse probability weighting (IPW: 0.055, $N = 830{,}194$) bracket the OLS and cross-fit estimates: the price association holds up under non-parametric reweighting.

Measurement-error diagnostic (IV). A leave-one-out IV suggests OLS understates the association (0.194 vs. 0.064), consistent with binary misclassification attenuation, but balance violations preclude a preferred-estimate interpretation (Appendix D).

Summary. The price association is remarkably stable: four approaches that differ in weighting, functional form, and information sets all land between 3.6% and 7.7%. They cluster rather than scatter—the hallmark of a real association, not a methodological artifact. To nullify the coefficient, an omitted confounder would need to explain 17.5% of residual variation in *both* FL presence and log prices, after item, year, and PBU fixed effects have already absorbed most cross-market heterogeneity (Cinelli and Hazlett, 2020). None of these estimates is causal; they document a conditional association robust enough to motivate deploying the screen.

[5] The difference $0.064 - 0.043 = 0.021$ has approximate SE $\sqrt{0.019^2 + 0.020^2} \approx 0.028$, yielding $t \approx 0.75$ ($p \approx 0.45$).

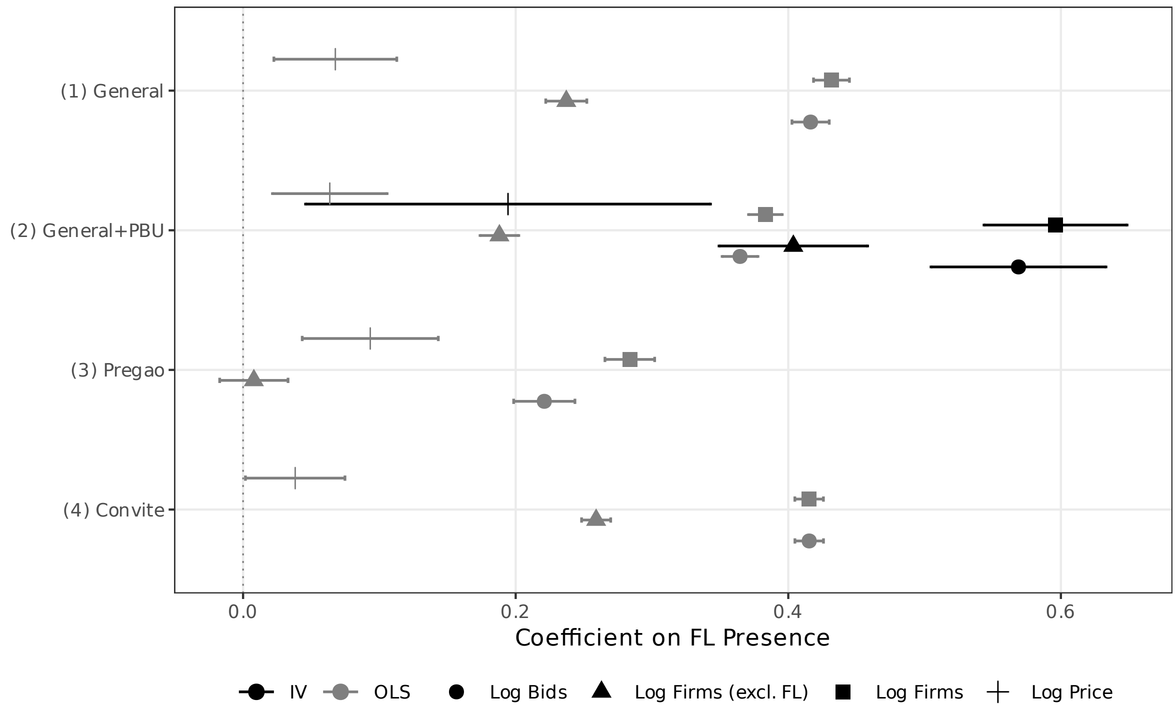

Figure 2: Coefficient on FL presence across outcomes, specifications, and estimation methods. OLS estimates in gray; IV estimates in red.

7.3 Detection Performance

The real test comes from outside the data used to build the screen. Figure F.1 (Appendix D) traces the FL ROC curve against CADE cartel co-participation: AUC = 0.94 (bootstrapped 95% CI: $[0.93, 0.95]$; Table F.1, Appendix D).

Participation volume contributes substantially to this figure: both the FL score (participation count) and the ground truth (CADE co-participation) are volume-driven. Decomposing the AUC within participation-count deciles yields a pooled within-band AUC of 0.54 (Table D.5, Appendix D), confirming that most of the unconditional discrimination reflects the volume channel. The within-band AUC exceeds chance but is modest; the screen's detection power comes primarily from identifying the high-participation tail of always-losers, not from discriminating among firms with similar volume.

For operational screening, the unconditional AUC remains the relevant metric: enforcement agencies care about total signal, and participation volume is part of the behavioral anomaly the screen exploits. The data-driven optimal threshold (Youden's $J = 0.84$) falls at 1.45×IQR—nearly identical to the 1.5× rule we chose *a priori* on economic grounds—achieving TPR = 1.00 with FPR = 0.16. A count-matched permutation test (2.2× at ±2 tenders, $p < 0.001$; Section 6) confirms that the FL classification adds signal beyond volume alone, consistent with the logistic decomposition showing that the zero-win condition is the active ingredient (Section 6).

Complementarity with bid-level screens. Against the same CADE ground truth, the FL screen achieves AUC = 0.94 versus 0.90 for a random forest trained on seven bid-level features (DeLong $p = 0.04$; Table F.2, Appendix D) and 0.79 for a coarser Imhof (2019)-style CV proxy (CV median split; Table D.6, Figure D.1 in Appendix D). The CV proxy understates what a full Imhof implementation (CV, kurtosis, spread, and a trained classifier) would achieve, so the random forest is the fairer benchmark.[6] What FL adds is reach: it works where bid-level records are unavailable—a common constraint in developing-country procurement—and produces firm-level flags that triage investigations before detailed bid analysis begins. The screen does not replace bid-level tools; it tells enforcement agencies where to point them.

Horse-race regression. Including both FL presence and an Imhof-style CV flag as regressors reveals that the two screens capture largely non-overlapping information (correlation 0.060; the theoretical maximum for two binary indicators with prevalences 4.8% and 50% is approximately 0.30, so the observed correlation is about one-fifth of this ceiling). The FL coefficient *rises* from 0.064 to 0.084 ($p < 0.01$) when the CV flag is added—a suppression effect predicted by the framework: coordinated cover bidding may produce low within-tender dispersion while maintaining FL participation. The Imhof flag enters at 0.021 ($p < 0.01$).

[6] Two structural differences also affect the comparison. First, the CADE ground truth, based on co-participation, partly favors participation-based screens through a volume channel (Section 6). Second, FL operates at the firm level while bid-level screens operate at the tender level; aggregating tender-level scores to the firm level introduces noise that penalizes bid-level approaches.

Both screens contribute marginally to R^2 given the dominance of high-dimensional fixed effects (Table D.7, Appendix D); the coefficient changes, rather than the R^2 increments, document their independent information content. The two screens complement each other: FL triages environments using participation records alone, and bid-level tools can then be deployed on the flagged subset.

Why combination degrades. Combining the two screens into a single index is tempting but backfires: AUC drops to 0.61, down from 0.94 for FL alone (Table D.6). The reason traces to the framework's central insight: under coordinated cover bidding, FL firms enter tenders where they *raise* within-tender dispersion (Table D.4, coefficient 0.47–0.55 on log bid SD), causing Imhof-style features to classify those environments as *less* suspicious. For the same firms, the two screens point in opposite directions. This is not a failure—it is the reason sequential deployment (screen $\rightarrow$ triage $\rightarrow$ investigate) dominates score combination.

Three additional checks—FL bid spreads in CADE-present tenders, volume-conditional overlap rates, and zero overlap among non-FL always-losers with cover-bid-like spreads—confirm that the participation anomaly, not the bid pattern, drives the CADE association (Appendix D).

The remaining subsections report supporting diagnostics. Their role is to assess whether the screen behaves as the cover-bidding interpretation predicts; they do not independently establish the mechanism.

7.4 Network-Split Heterogeneity

Where does the price association come from? We split FL firms into concentrated-market and competitive-market subgroups based on co-bidding network metrics (Section 5.5; Table 5).[7] Nearly all of the association comes from competitive-market FL firms (0.126, $p < 0.01$, PBU FE); in concentrated markets the coefficient is indistinguishable from zero (-0.018, $p > 0.3$).

[7] Markets above the winner-HHI sample median are classified as concentrated; firms sharing at least two tenders are classified as repeat co-participants.

This is exactly what the framework predicts (Appendix A): where no single firm dominates, manufactured competition has the highest marginal return; where market power already sustains high prices, cover bidders are redundant. The enforcement implication is striking: the screen is most informative precisely where undetected collusion would cost taxpayers the most.

The network-split classification is endogenous: market concentration is not randomly assigned, and FL firms that appear in concentrated markets differ systematically from those in competitive markets. Noisier FL classification in concentrated markets or thinner fixed-effect cells could also attenuate the coefficient. The monotonic pattern across HHI quartiles is suggestive of the framework's prediction but does not establish that competition level causally moderates the FL–price association.

Table 5: Price Effects by FL Suspicion Level

	(1) General	(2) General	(3) Pregão	(4) Convite
High-suspicion FL	-0.0132	-0.0181	-0.0004	-0.0288*
	(0.0195)	(0.0178)	(0.0229)	(0.0149)
Low-suspicion FL	0.1304***	0.1257***	0.1634***	0.0911***
	(0.0329)	(0.0312)	(0.0346)	(0.0281)
Observations	1,654,401	1,654,401	546,549	1,107,852
Item + Year FE	YES	YES	YES	YES
PBU FE	NO	YES	YES	YES

Notes: DV: log negotiated price. High-suspicion FL = winner HHI above median and $\geq$2 repeat co-bidding partners. SE clustered at item level. *** p<0.01, ** p<0.05, * p<0.1.

7.5 Bajari–Ye Tests

The price association concentrates in competitive markets. But are FL bids actually different from genuine bids, or do they just happen to come from firms that lose? The Bajari and Ye (2003) framework provides two formal tests (Table D.2, Appendix D).

Exchangeability. If FL firms were ordinary competitors, their bid residuals would be statistically indistinguishable from non-FL residuals. They are not: the KS test rejects exchangeability ($D = 0.15$, $p < 0.001$). FL bids appear to be drawn from a different process.

Conditional independence. If FL firms bid independently, the pairwise products of their residuals within a tender would average near zero. Instead, the mean is 4.28 ($t = 81.0$, $p < 0.001$), with the bootstrap FL–non-FL difference excluding zero—consistent with coordinated bidding rather than independent cost draws. Enriching the first stage with firm age and CNAE sector dummies alongside firm size leaves R-squared virtually unchanged (0.770) and does not alter the KS or pairwise-product results, reducing the concern that the exchangeability violation is driven by omitted firm-level cost shifters (Table D.2).

Under tender FE, the FL pairwise product drops below non-FL—a reversal predicted by Regime 2, though also consistent with purely between-tender differences (Table F.3, Appendix D).

7.6 Additional Diagnostic Evidence

FL entry event study. Within-PBU event studies around first FL entry (Online Appendix) show positive pre-event coefficients at $t = -2$ and $t = -3$: PBUs receiving FL entry *already had higher prices*. This is a disciplined non-result. The pre-trends are consistent with strategic market selection but equally compatible with unobserved confounding; the reverse-causality test (M3, elasticity = 0.002) and sensitivity analysis ($RV = 17.5\%$) constrain but do not eliminate the confounding channel. The pre-trends reinforce the paper's framing: the conditional price association should not be read as a causal effect of FL presence on prices.

Minimum-bidder constraint variation. Convite's three-bidder minimum creates variation in the incentive to deploy cover bidders. Interacting FL with a constraint-binding indicator yields $\hat{\beta}_{\text{FL}}(n \geq 3) = 0.076$ (SE = 0.021, $p < 0.001$)—a 7.6% voluntary premium—and $\hat{\beta}_{\text{FL}\times(n<3)} = -0.160$ ($p < 0.001$). The price association concentrates where cover bidding is a strategic choice. A Callaway–Sant'Anna DiD exploiting CADE conviction timing shows that FL presence drops sharply post-conviction (ATT $= -0.275$, $p < 0.001$; Appendix E).

7.7 Structural Diagnostic

As a final supporting diagnostic, we report the structural estimation results from the bid-level mixture model (Appendix B). The regime test distinguishes between two forms of cover bidding—complementary (uniformly dispersed above the winning bid) and coordinated (tightly clustered near the winning bid). It does not test cover bidding against competition; that comparison is addressed by the cross-sectional price association (Section 7.2) and the CADE validation (Section 6). BIC strongly favors the coordinated regime (ΔBIC $= -91{,}473$; Table B.1). The dispersion ratio $\hat{\sigma}_c/\hat{\sigma}_g = 0.72$: FL bids are 28% *less* dispersed than non-FL bids—as expected under coordinated cover bidding, where variance-based screens lose power. The n-conditional markup (6.4%) is close to the OLS baseline. The enforcement implication is direct: participation-based screens complement dispersion-based tools precisely because the dominant cover-bidding regime defeats the latter. Where coordination is tightest, variance screens see nothing; the FL screen sees everyone who showed up.

Additional specification tests are in Appendix B; a companion paper (Genicolo-Martins and Furquim de Azevedo, 2026) develops the full structural model, calibration, and policy counterfactuals.

8 Supporting Diagnostics and Alternative Explanations

A price association, by itself, does not tell an enforcement agency *why* prices are higher. This section asks a narrower question: does the screen behave the way coordinated cover bidding would predict? Five diagnostics probe different facets of this question. None is dispositive alone; the weight comes from the joint pattern (Online Appendix).

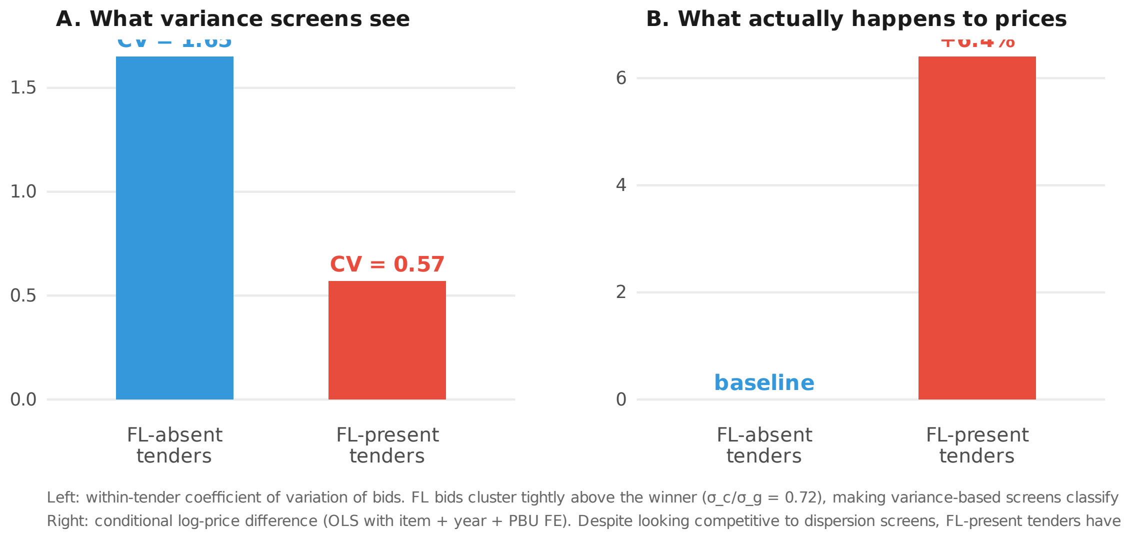

Figure 3: The detection paradox. *Left*: FL-present tenders have *lower* within-tender bid dispersion (CV 0.57 vs. 1.65), causing variance-based screens to classify them as competitive. *Right*: the same tenders have 6.4% higher conditional prices. Coordinated cover bidding (Regime 2) defeats dispersion-based detection precisely when coordination is tightest—the setting where a participation-based screen is most valuable.

M1: Do FL firms crowd out genuine competitors? If cover bidders displaced real competitors, FL-present tenders would have fewer genuine participants. The opposite happens: +0.19 more non-FL firms under PBU fixed effects ($p < 0.01$). FL firms add to the headcount rather than replace anyone—exactly what the framework predicts (Appendix A).

M2: Are winning bids anchored to the reference price? BEC posts reference prices before bidding opens. In FL-present tenders, winners bid 4% closer to that anchor (-0.041, $p < 0.01$)—a pattern one would expect if the winning bid and the cover bids are calibrated around the same public price (Marshall and Marx, 2012).

M3: Do FL firms chase high prices? Within item markets, a 10% lagged-price increase raises FL entry probability by 0.02 pp ($\hat{\beta} = 0.0021$, SE $= 0.0008$). The elasticity is too small to explain much of the 6.4% price association, though it cannot fully separate reverse causality from strategic selection.

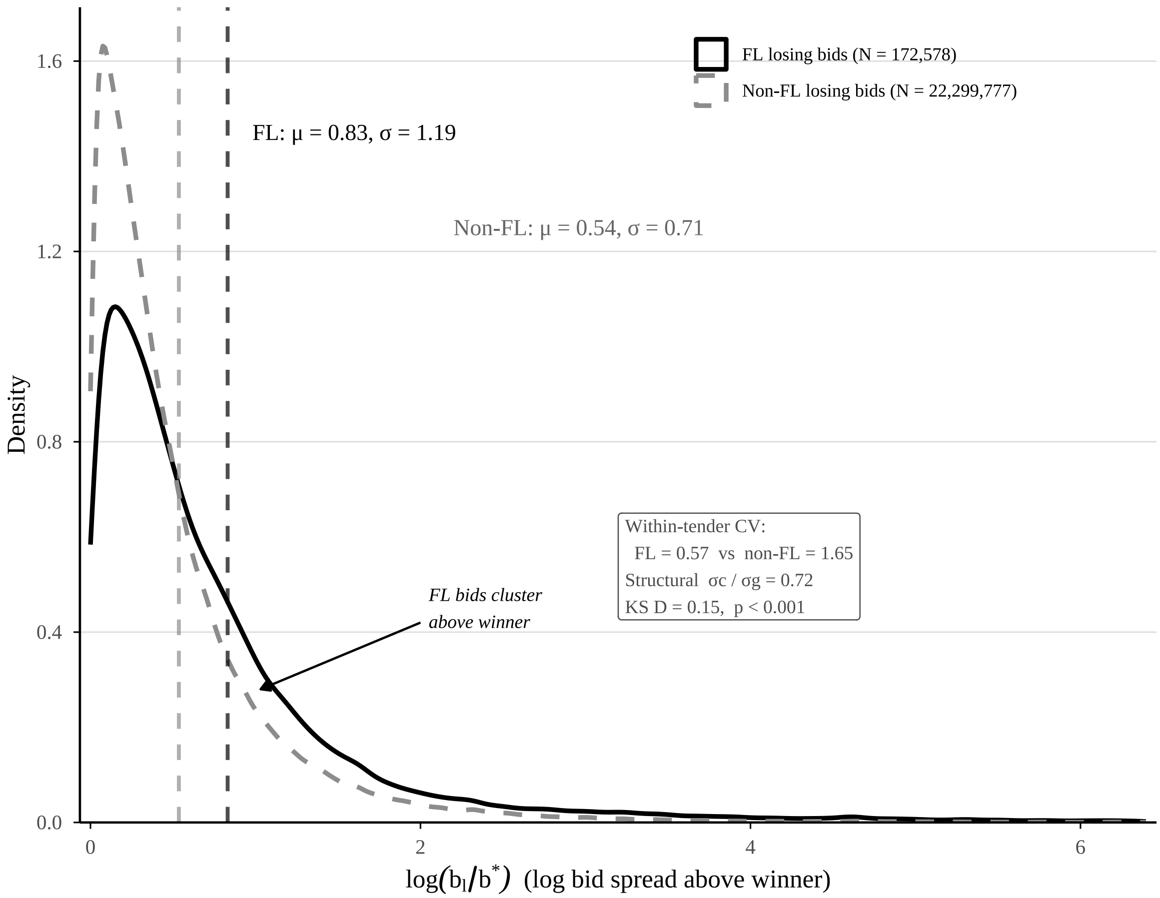

Figure 4: Distribution of log bid spread above winning price for FL and non-FL losing bids. FL bids concentrate above the winner ($\bar{\epsilon} = 0.83$) with overall $\sigma = 1.19$. Within-tender dispersion is *lower* for FL bids (structural $\hat{\sigma}_c/\hat{\sigma}_g = 0.72$).

Alternative explanations. FL firms are roughly one year younger than non-FL firms—consistent with cartels creating short-lived shells (Marshall and Marx, 2012), but equally compatible with young firms exploring procurement markets before giving up. More telling is the pregão–convite asymmetry: the premium is 9.3% in electronic auctions (no minimum-bidder rule) versus 3.8% in sealed-bid tenders (three-bidder minimum). Where the law forces participation, the FL signal is diluted by compliance noise. When we interact FL with a constraint-binding indicator (Section 7.6), the entire price association comes from voluntary participation (7.6% premium)—not from rule-forced entry. None of these patterns pins down the mechanism; each admits more than one reading.

Rationality of FL participation. With zero wins, expected profit per bid is $-c < 0$ for any positive bidding cost. At the FL threshold (14 tenders), cumulative losses are R$700–7,000—bad luck, perhaps. But at the 90th percentile (50+ tenders), losses reach R$2,500–25,000, and a Bayesian learner updates Pr(win) to 0.019. At that point, exit is the only rational response—unless someone is paying you to stay.[8] The rationality calculation documents anomalous participation, not proven side payments; its role is to strengthen the case that FL captures behavior worth screening for, not to establish the mechanism. Full calculations are in the Online Appendix.

M4: Dyadic linkage concentration. A stratified permutation test (1,000 iterations, preserving participation quartiles) shows FL firms form 4,696 high-frequency co-bidding pairs ($\geq$5 shared tenders) versus a permuted mean of 3,271 ($p < 0.001$). The excess of stable FL–winner pairs is consistent with the persistent co-bidding relationships the framework emphasizes (Online Appendix).

[8]The calculation assumes a beta-binomial updating model with a uniform prior (Beta(1,1)), constant bid cost per tender (R$50–500), and average tender value $\bar{V}$ = R$86,000 with 10% margin. Under these assumptions, expected gain per bid is R$163—below the lowest bidding cost. Alternative priors (Beta(2,2)) yield posterior Pr(win) = 0.038 after 50 losses, still making exit rational. Win rate zero in BEC does not mean zero elsewhere, but 50+ zero-win BEC participations require explanation even for firms profitable outside the platform (Asker, 2010; Marshall and Marx, 2012).

M5: Firm exit in FL-exposed markets. A Cox model shows FL exposure is associated with *lower* exit hazard (HR = 0.60, $p < 0.01$): firms in FL-exposed markets survive longer, the opposite of crowding-out.[9] Together with M1, the result suggests FL firms add noise rather than competition (Online Appendix).

Strategic adaptation. A sophisticated cartel could rotate its cover bidders or let them win occasionally to stay below the threshold—and the threshold sensitivity analysis (Section 9) already suggests some do. Any operational deployment would therefore need periodic recalibration alongside bid-level tools (Harrington, 2008), which is why the three-stage workflow in Section 11 treats the screen as a first stage rather than a final word.

Joint assessment. Each diagnostic, taken alone, admits other readings. Taken together, the pattern is hard to square with a competitive account: entry without displacement, reference-price anchoring, negligible reverse-causality elasticity, excess dyadic linkage, and lower exit hazard in FL-exposed markets all point in the same direction. The diagnostics do not prove coordinated cover bidding. They do make a compelling case that the screen is worth deploying.

8.1 Bid Rotation and Bid Inflation

Three additional patterns round out the diagnostic picture. FL firms co-participate with a wider range of winners (HHI 0.178 vs. 0.303 for non-FL always-losers), form persistent co-bidding pairs far beyond chance (4,696 pairs sharing ≥ 5 tenders vs. 494 for non-FL), and submit bids 85% above the winner versus 43% for non-FL losers. Each pattern is consistent with coordinated cover bidding; each also admits a competitive reading. Full details are in Appendix D.

[9] The PH assumption is rejected ($p < 0.001$); the hazard ratio is descriptive. "Exit" means cessation of BEC participation.

9 Robustness and Extensions

We stress-test the baseline from five angles: FL-definition variants (IQR multiplier, win-rate cutoff, temporal window); sample and specification variants (unrestricted sample, homogeneous subsamples, item×year FE, matching); alternative clustering; sensitivity to unobservables (Cinelli–Hazlett); and staggered DiD. Table 6 collects the point estimates alongside the identifying assumption each estimator requires.

Table 6: Robustness summary: FL conditional price coefficient across checks

Check	Coeff.	SE	N	Assumption	Note
Baseline and matching					
OLS (item + year + PBU FE)	0.064	0.020	1,654,401	Sel. on obs. + FE	Baseline
CEM	0.077	0.024	969,751	Sel. on obs.	Sec. 7.2
IPW	0.055	0.021	830,194	Sel. on obs.	Sec. 7.2
Cross-fit	0.036	0.019	1,654,401	Sel. on obs. + FE; no mechanical link	Out-of-sample
FL definition variants					
IQR 1.0×	0.079	0.023	1,654,401	Sel. on obs. + FE	3,442 FL
IQR 1.5×	*(= OLS baseline above)*				
IQR 2.0×	0.060	0.022	1,654,401	Sel. on obs. + FE	2,093 FL
IQR 3.0×	0.050	0.024	1,654,401	Sel. on obs. + FE	1,456 FL
Specification variants					
Item×year FE	0.074	0.022	1,654,401	Sel. on obs. + FE	Tighter controls
Two-way clustering	0.064	0.024	1,654,401	Sel. on obs. + FE	Item + PBU
Alternative specifications					
Continuous (log max AL)	0.022	0.005	1,654,401	Sel. on obs. + FE	Per log-point
Decomposition (odd-yr FL)	0.043	0.019	1,654,401	Sel. on obs. + FE	Intermediate
Horse-race (FL + Imhof CV)	0.084	0.023	1,654,401	Sel. on obs. + FE	Suppression
IV (leave-one-out)	0.194	0.077	1,654,401	Excl. restriction	Meas. error
Staggered DiD					
Stacked DiD	−0.006	0.014	715,116	Parallel trends	Pre-trends
Oster $\hat{\delta}$	degen.		1,654,401	Proportional sel.	R^2 gap ≈ 0

Notes: "Sel. on obs. + FE" = selection on observables absorbed by item, year, and PBU fixed effects. Cross-fit removes any mechanical link between classification and outcome. The IV is a measurement-error diagnostic; exclusion-restriction concerns (Section 7.2) keep it off the primary range.

Threshold sensitivity. Stricter thresholds yield smaller coefficients (0.079 at 1.0×, 0.064 baseline, 0.060 at 2.0×, 0.050 at 3.0×), likely because classification noise increases as the treated group shrinks. A two-dimensional sensitivity analysis varying both multiplier and win-rate cutoff produces significant coefficients across all 36 cells (Online Appendix). For a deploying authority, this means the screen's triage signal is not sensitive to the exact threshold chosen.

Sensitivity to unobservables. Cinelli and Hazlett (2020) sensitivity analysis yields $RV_{q=1} = 17.5\%$: an unobserved confounder would need to explain 17.5% of residual variation in *both* FL presence and log prices to nullify the coefficient—stringent given the item, year, and PBU fixed-effects structure. The bound assumes a single additive confounder; if selection operates through multiple correlated channels—each individually below 17.5%—the required confounder strength may be lower. The positive pre-trends at $t = -2$ and $t = -3$ suggest multi-channel selection may be relevant (Section 10). Full sensitivity plots are in Appendix D.

Staggered difference-in-differences. The Callaway & Sant'Anna ATT is 0.014 (SE = 0.042) and the stacked DiD (Cengiz et al., 2019) gives -0.006 (SE = 0.014): both are statistically indistinguishable from zero. The minimum detectable effect at 80% power is 0.117 for the C&S estimator—nearly twice the OLS baseline of 0.064—so the DiD lacks statistical power to detect the cross-sectional association even if it were entirely causal. Moreover, positive pre-event coefficients at $t = -2$ and $t = -3$ (Figure E.1, Appendix E) indicate that PBUs receiving FL entry already had higher prices, consistent with the framework's prediction that cover bidders target profitable environments.

The DiD null is uninformative about the FL–price association: it is compatible with zero causal effect, insufficient power, or—most likely—violation of parallel trends through strategic market selection. The paper's evidence hierarchy does not rely on DiD identification; the cross-sectional association documented in Section 7.2 rests on selection-on-observables absorbed by high-dimensional fixed effects (Tables E.2–E.1, Appendix E).

Coefficient stability and cost-benefit. The Oster (2019) bound is degenerate ($\hat{\delta} = 261.6$) because PBU FE barely move R^2 (0.879 to 0.886); the Cinelli and Hazlett (2020) framework ($RV = 17.5\%$) is the appropriate sensitivity metric. Cost-benefit calculations for screening deployment appear in Online Appendix.

Oversight heterogeneity. The FL–price association varies sharply with PBU size: 0.214 in the smallest quartile, 0.098 in Q2, 0.045 in Q3, and 0.017 in Q4—a 12.5× gradient (Table F.4, Appendix D). This matches the framework's comparative static ($\partial m^*/\partial\theta_k < 0$): where oversight is weaker, the screen bites harder. Smaller PBUs also have thinner fixed-effect cells, so part of the gradient could be mechanical, but the monotonic decline across all four quartiles is hard to attribute to noise alone. Threshold stability across IQR multipliers and a comparison between the IQR rule and model-based classifiers are in Appendix D.

10 Limitations

The screen is an investigative prioritization device: it flags suspicious procurement environments for closer scrutiny. We document a robust conditional association and a consistent diagnostic pattern. We do not claim causal identification or proof of the cover-bidding mechanism—nor does a triage tool require either.

Diagnostics are not identification. The five diagnostics are consistent with coordinated cover bidding but do not pin down the mechanism. Each admits alternative readings in isolation, and even the joint pattern does not rule out every non-collusive story. Their role is to strengthen the case for deployment, not to substitute for forensic investigation.

Within-market selection. Item, year, and PBU fixed effects absorb time-invariant cross-market heterogeneity; the cross-fit removes any mechanical link between classification and outcomes. What remains is the possibility of time-varying within-market confounders that attract both FL firms and higher prices. The Cinelli–Hazlett bound ($RV = 17.5\%$) puts a number on how strong such a confounder would need to be, but the bound assumes a single linear additive confounder; if selection operates through multiple correlated channels—as the positive pre-trends at $t = -2$ and $t = -3$ suggest it may—the bound is less informative. The diagnostics in Section 8 constrain the set of compatible stories, but constraining is not identifying.

Event-study pre-trends and DiD null. Positive coefficients at $t = -2$ and $t = -3$ before first FL entry are consistent with strategic market selection (Section 7.6) and equally compatible with unobserved confounding. The staggered DiD estimators yield statistically insignificant ATTs (C&S: 0.014, SE = 0.042; stacked: -0.006, SE = 0.014), and the minimum detectable effect exceeds the OLS baseline. The DiD provides no evidence that FL entry causes price increases. The pre-trends preclude a causal reading and reinforce the paper's non-causal framing; the paper relies on within-item conditional comparisons for this reason.

Classification boundaries. The screen depends on a zero-win bright line and a participation threshold. Relaxing either weakens the coefficient. Sophisticated cartels could evade by allowing cover bidders occasional wins or staying below the threshold—hence the three-stage workflow (Section 11) that combines FL screening with bid-level forensics and periodic threshold recalibration. Any operational deployment would require updating the threshold as procurement environments evolve.

Screen scope. The screen flags suspicious environments, not individual firms. It does not identify who within a flagged environment is colluding, does not estimate the causal effect of collusion on prices, and does not produce evidence sufficient for sanctions. These are inherent boundaries of a triage tool.

External validity. All results come from one platform (BEC). The minimal data requirements make replication straightforward, but whether the screen performs comparably in other procurement systems, institutional settings, and enforcement regimes is an open question.

11 Conclusion

The hardest cartels to detect are the ones that look competitive. When cover bidders coordinate tightly, the tenders they contaminate produce textbook-looking bid distributions—low variance, smooth spacing—while prices quietly climb. The frequent-loser screen bypasses this trap by ignoring what firms bid and focusing on a simpler signal: who keeps showing up to lose. In São Paulo's procurement, this signal flags environments with 3.6–7.7% higher conditional prices, achieves AUC = 0.94 against known cartels (surviving exact count matching at 2.2×), and requires nothing beyond win/loss records.

We propose a three-stage enforcement pathway—addressing the resource-allocation problem at the heart of antitrust enforcement since Posner (1970). (i) *Screen*—apply the FL rule to participation records. The computation requires a single table (firm, tender, outcome), no bid microdata, no enforcement priors, and no supervised training; a procurement agency can implement it with a SQL query or spreadsheet. (ii) *Triage*—cross-reference FL-flagged markets with network metrics (co-bidding concentration, winner HHI) and oversight-capacity indicators (PBU size quartile) to prioritize the highest-risk environments. The 12.5× gradient in the FL–price association between the smallest and largest purchasing units suggests that triage resources are best directed toward low-oversight settings. (iii) *Investigate*—deploy bid-level forensic tools *on the triaged subset*. The screen's near-orthogonality with bid-level screens (correlation 0.06) means the two stages capture different collusion signatures, and combining them reduces both false positives and missed cases.

Two features make the screen administrable. First, the threshold is robust: varying the IQR multiplier from 1.0× to 3.0× and the win-rate cutoff from 0% to 5% yields significant

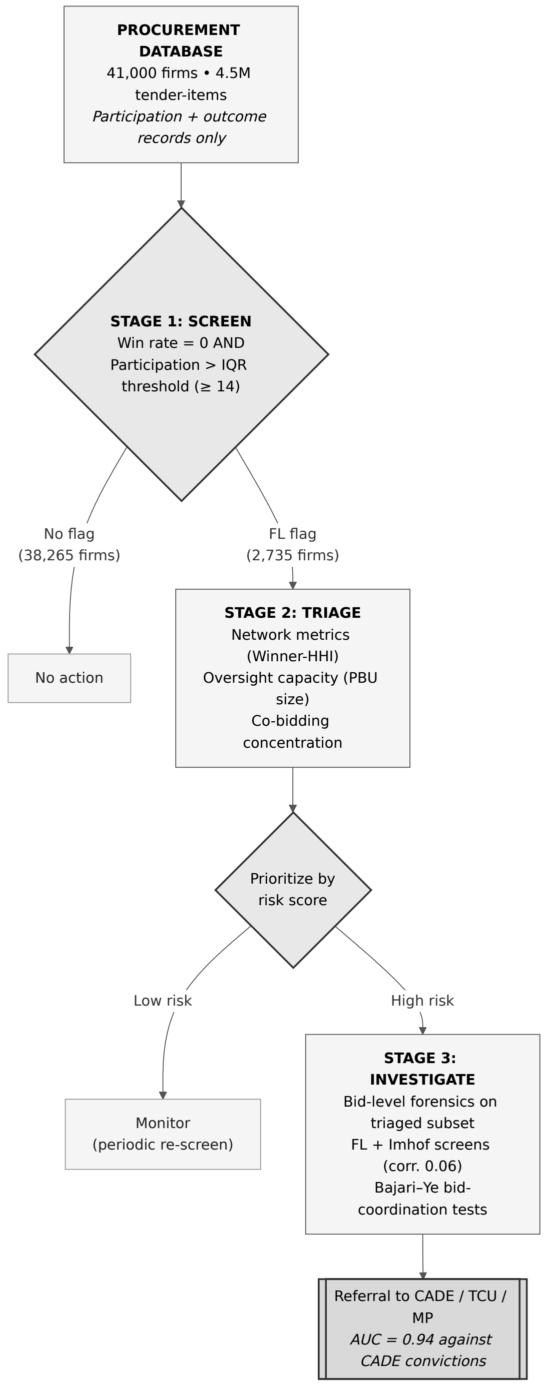

Figure 5: Three-stage enforcement pathway. Stage 1 reduces 41,000 firms to 2,735 FL-flagged firms using participation records alone. Stage 2 prioritizes by network structure and oversight capacity. Stage 3 deploys bid-level forensics on the triaged subset.

coefficients across all 36 combinations, so the signal does not hinge on a precise calibration choice. Second, the data requirements are minimal—participation and outcome records that every procurement system already collects—making the screen portable to developing-country settings where bid-level forensics are out of reach.

The screen's main limitation is also its defining boundary: it flags environments, not firms, and the price association is not causal. For a triage tool, these are features, not defects. An enforcement agency would use it to allocate investigative effort, not to adjudicate liability. The screen does not replace forensic investigation; it tells investigators where to start.

The most pressing next step is replication. The economic logic—cover bidders must participate to perform their role—is institution-general, but the screen's empirical performance outside BEC remains to be tested. The FL rule, the ROC analysis, and the horse-race specification are fully replicable with participation and outcome data alone.

Institutional change and the new procurement law. Brazil's recent procurement reform (Lei 14.133/2021) offers a natural laboratory. The reform eliminates the convite modality and with it the minimum-bidder rule that our framework identifies as a direct incentive for cover bidding. Two predictions follow. The constraint-binding channel documented in Section 7.6—where the quorum rule forces participation and the FL interaction is negative ($\hat{\beta} = -0.160$)—should disappear. The voluntary channel (7.6% premium where $n \geq 3$) should survive, since it reflects strategic choice rather than rule compliance. Comparing pre- and post-reform procurement within the same item markets would provide a clean test. The reform also introduces new modalities (diálogo competitivo, concorrência eletrônica) with different participation incentives, and whether the screen works in those formats is an open question—though it bears noting that the behavioral footprint the screen exploits (cover bidders must participate) does not depend on the procurement format.

The reform sharpens the screen's relevance in a broader sense. The FL–price association is *larger* in pregão (9.3%) than in convite (3.8%), which suggests that cover bidding is not merely a response to the quorum constraint but a strategic practice that persists wherever coordination pays. As procurement migrates to electronic formats under Lei 14.133, participation records will become richer and more standardized—exactly the data environment where a participation-based screen has the widest reach. The transition period, with federal, state, and municipal entities adopting the new law at different speeds, creates quasi-experimental variation that future work can use to test portability across institutional regimes.

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., R. McMillan, and S. Tadelis (2009). Auctions versus negotiations in procurement: An empirical analysis. *Journal of Law, Economics, and Organization 25*(2), 372–399.

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

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.

Baranek, B. and V. Titl (2024). The cost of favoritism in public procurement. *Journal of Law and Economics 67*(2), 445–477.

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

Caoui, E. H. (2022). A study of umbrella damages from bid rigging. *Journal of Law and Economics 65*(2), 239–277.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics 134*(3), 1405–1454.

Chassang, S., K. Kawai, J. Nakabayashi, 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. and J.-F. Houde (2014). The effect of explicit communication on pricing: Evidence from the collapse of a gasoline cartel. *Journal of Industrial Economics 62*(2), 191–228.

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

Connor, J. M. (2007). Price-fixing overcharges: Legal and economic evidence. In *Research in Law and Economics*, Volume 22, pp. 59–153. Elsevier.

Genicolo-Martins, D. and P. Furquim de Azevedo (2026). Cover bidding in public procurement: A structural model of the dispersion paradox. Working Paper, INSPER.

Ghosal, V. and D. D. Sokol (2014). The evolution of U.S. cartel enforcement. *Journal of Law and Economics 57*(S3), S51–S65.

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

Harrington, J. E. and M.-H. Chang (2015). When can we expect a corporate leniency program to result in fewer cartels? *Journal of Law and Economics 58*(2), 417–449.

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.

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

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.

Polinsky, A. M. and S. Shavell (2000). The economic theory of public enforcement of law. *Journal of Economic Literature 38*(1), 45–76.

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.

Posner, R. A. (1970). A statistical study of antitrust enforcement. *Journal of Law and Economics 13*(2), 365–419.

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. (2023). Identification and inference in first-price auctions with collusion. Working Paper.

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

A Proofs of Propositions

Proof of Proposition 1 (Optimal number of cover bidders). Let $\pi(m)$ denote the cartel's gross benefit from deploying m cover bidders—reflecting the higher price sustained by the appearance of competition. Each cover bidder costs c_1 (direct compensation) and raises the expected detection penalty by $\theta_k\phi_0$ (where θ_k is the purchasing unit's oversight capacity and ϕ_0 is the per-bidder detection sensitivity). Relaxing the integer constraint, the first-order condition equates the marginal benefit to the marginal cost:

$$\frac{\partial\pi}{\partial m} = c_1 + \theta_k\phi_0 \tag{4}$$

The gross benefit is strictly concave in m ($\partial^2\pi/\partial m^2 < 0$): each additional cover bidder contributes less to the illusion of competition than the last. The second-order condition is therefore satisfied and the interior solution $\tilde{m}$ defined by (4) is therefore unique. When the constraint $n + m \geq \underline{n}(\tau)$ binds (i.e., $\underline{n}(\tau) - n > \tilde{m}$), the optimal solution is the corner $m^* = \underline{n}(\tau) - n$. Otherwise, $m^* = \lfloor \tilde{m} \rfloor$ (rounding to the nearest integer below $\tilde{m}$). The stated formula follows.

For the comparative statics, differentiate (4) implicitly. Let $H(m, \theta_k) \equiv \partial\pi/\partial m - c_1 - \theta_k\phi_0 = 0$ define the interior solution. Then:

$$\frac{\partial m^*}{\partial\theta_k} = -\frac{\partial H/\partial\theta_k}{\partial H/\partial m} = -\frac{-\phi_0}{\partial^2\pi/\partial m^2} = \frac{\phi_0}{\partial^2\pi/\partial m^2} < 0$$

since $\phi_0 > 0$ and $\partial^2\pi/\partial m^2 < 0$ (concavity). The argument for $\partial m^*/\partial c_1 < 0$ is identical, replacing

ϕ_0 with 1. Property (iii) follows from the definition of m^* as the maximum of the corner and interior solutions: convite's higher $\underline{n}$ weakly raises the corner solution. $\square$

Proof of Proposition A (Complementary cover bidding). Cover bidders under Regime 1 lack information about genuine bid levels—they know only that their bids must exceed the cartel's winning bid b^* and fall within a plausible range $[b^*,\ b^*+\delta]$. Without a coordination signal, the best the cartel can do is scatter cover bids as randomly as possible to avoid creating detectable patterns. This is equivalent to maximizing the entropy of the cover-bid distribution G, or equivalently, minimizing the KL divergence from the maximum-entropy distribution $U = \text{Uniform}[b^*, b^* + \delta]$. By Gibbs' inequality, $D_{\text{KL}}(G\|U) \geq 0$ with equality if and only if $G = U$ almost everywhere. Since U belongs to the feasible set $\{G : \text{supp}(G) \subseteq [b^*, b^* + \delta]\}$, the minimum is attained at $G_1 = U[b^*, b^* + \delta]$. $\square$

Proof of Proposition 3 (Regime 2: Coordinated cover bidding). Cover bids are modeled as $b_\ell \sim \text{TruncNormal}(\mu_c, \sigma_c^2, b^*, \infty)$ with $\mu_c = b^* + \epsilon$, $\epsilon > 0$. Truncation at b^* ensures $\Pr(b_\ell > b^*) = 1$ by construction. Let $\alpha \equiv (b^* - \mu_c)/\sigma_c = -\epsilon/\sigma_c < 0$. The variance of the truncated distribution is $\sigma_c^2[1 - \delta(\alpha)]$, where $\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha] > 0$ and $\lambda(\alpha) = \phi(\alpha)/(1 - \Phi(\alpha))$ is the inverse Mills ratio. Since $\alpha < 0$, the term $\lambda(\alpha) - \alpha > \lambda(\alpha) > 0$, so $\delta(\alpha) > 0$ for all finite α and the observed variance is strictly less than σ_c^2. If the cartel sets $\sigma_c = \sigma_{\text{genuine}}$ to mimic genuine bid dispersion, the left-truncation compresses the effective standard deviation below σ_{genuine}, yielding $\text{sd}(b_\ell \mid b_\ell > b^*) < \sigma_{\text{genuine}}$. This is the mechanism through which coordinated cover bidding defeats dispersion-based screens: the screen observes low within-tender variance and reads it as competitive, when in fact it reflects tight coordination around the winning bid. $\square$

Proof of Proposition 4 (Market-selection prediction). From the FOC (4) at the interior solution: $\partial\pi/\partial m = c_1 + \theta_k\phi_0$. The right-hand side does not depend on HHI. Differentiating implicitly:

$$\frac{\partial^2\pi}{\partial m^2} \cdot \frac{\partial m^*}{\partial \text{HHI}} + \frac{\partial^2\pi}{\partial m\,\partial \text{HHI}} = 0$$

Solving:

$$\frac{\partial m^*}{\partial \text{HHI}} = -\frac{\partial^2\pi/\partial m\,\partial \text{HHI}}{\partial^2\pi/\partial m^2}$$

By concavity, $\partial^2\pi/\partial m^2 < 0$ (denominator negative). By assumption, $\partial^2\pi/\partial m\,\partial\text{HHI} < 0$: the marginal benefit of cover bidding decreases with market concentration. Therefore $\partial m^*/\partial\text{HHI} < 0$.

The economic content of this cross-derivative restriction is that cover bidders are more valuable in competitive markets: when HHI is low, genuine bidders exert downward pressure on prices, and cover bidders serve to create noise, satisfy minimum-bidder rules, and obscure the cartel's bid pattern from forensic screens. When HHI is high, the cartel can sustain prices through market power alone, reducing the marginal return to manufactured competition. $\square$

B Structural Model: Identification, Estimation, and Specification Tests

B.1 Identification of Structural Parameters

Genuine-bid parameters $(\boldsymbol{\beta}, \sigma_g^2)$. These are identified from within-item–year variation in non-FL losing bids. The exclusion restriction is that FL status does not enter the genuine-bid distribution F_{genuine}: conditional on observables $\mathbf{x}_j$ and fixed effects, a genuine bidder's cost draw and markup are independent of whether FL firms participate in the same tender. This permits estimation on 23.2 million non-FL losing bids without selection correction.

Cover-bid parameters (δ or σ_c, ϵ). Conditional on the winning bid b_t^* (observed in each tender), the cover-bid distribution is identified from 189,381 FL losing bids. Under Regime 1, δ is identified by the range of FL bids above b_t^*. Under Regime 2, (ϵ, σ_c) are identified by the mean and variance of $\log b_{\ell t} - \log b_t^*$ among FL bids. Model selection between regimes uses BIC.

Detection probability θ_k. Not point-identified from bid data alone. We bound it using cross-PBU variation in oversight capacity, proxied by procurement volume. The comparative static $\partial m^*/\partial \theta_k < 0$ (Proposition 1) generates a testable implication evaluated in the oversight heterogeneity analysis (Section 9).

Cartel markup. The markup $b^* - c_{j^*}$ is not separately identified: we observe b^* but not c_{j^*}. The structural model characterizes the cover-bid mechanism, not price effects.

B.2 Structural Parameter Estimates

B.3 Specification Tests

Three diagnostic implications can be assessed on moments not used in estimation; all are consistent with the framework. *First*, the corner-vs-interior solution: mean FL count is 10.7 when $n_{\text{genuine}} = 0$ (constraint forces $m \geq 3$) and ≈ 2 for $n \geq 2$ (interior solution)—consistent with the framework (Proposition 1). *Second*, within-tender bid dispersion: in 29,398 tenders with both FL and non-FL bidders, FL bid CV is 0.57 vs. non-FL 1.65; FL bids are less dispersed in 82.6% of tenders (paired $t = -104$, $p < 0.001$), consistent with Regime 2 at the tender level. *Third*, the m–θ comparative static: mean FL per tender decreases monotonically from Q1 (0.41) to Q4 (0.31) by PBU size.

B.4 Regime 2 QQ-Plot

Figure B.1 compares the empirical quantiles of FL bid residuals against Regime 2 (truncated normal) theoretical quantiles. The fit is close in the interior of the distribution, with deviations in the tails consistent with a mixture of approximately 90% truncated normal and 10% uniform draws.

C Core Supporting Tables and Figures

Table C.1 reports conditional descriptive statistics comparing FL-present and FL-absent tenders before and after absorbing item, year, and PBU fixed effects.

Table B.1: Structural Parameter Estimates: Cover-Bid Mixture Model

Parameter	Estimate	Bootstrap SE	95% CI
Panel A: Genuine Bid Distribution			
$\hat{\mu}_g$ (mean log bid)	4.059	—	—
$\hat{\sigma}_g$ (residual SD)	1.642	—	—
First-stage R^2	0.772	—	—
N (genuine losing bids)	23,177,905		
Panel B: Regime 1 — Complementary Cover Bidding (Uniform)			
$\hat{\delta}$ (99th-pctile log-spread)	6.397	0.097	[6.213, 6.587]
$\tilde{\delta}$ (median log-spread)	0.499	—	—
Log-likelihood	$-320,273$	—	—
BIC	640,559	—	—
KS distance	0.640	—	—
Panel C: Regime 2 — Coordinated Cover Bidding (Log-Normal)			
$\hat{\epsilon}$ (mean log-spread)	0.831	0.005	[0.822, 0.843]
$\hat{\sigma}_c$ (cover-bid SD)	1.187	0.011	[1.167, 1.212]
Log-likelihood	$-274,531$	—	—
BIC	549,086	—	—
KS distance	0.242	—	—
Panel D: Model Selection			
Selected regime	Regime 2 (Coordinated)		
ΔBIC (R2 − R1)	$-91,473$		
$\hat{\sigma}_c/\hat{\sigma}_g$	0.723		
N (FL losing bids)	189,381		
Panel E: Implied Markup (Reduced-Form, Model-Disciplined)			
$\hat{\mu}$ (n-conditional, item/year/PBU FE)	0.062	0.019	—
Implied markup (%)	6.4%		
OLS baseline (%, collapsed data)	6.6%		
Cross-fit average (%)	3.6%		
IV upper bound (%)	21.4%		

Notes: Two-stage ML estimation of the cover-bid mixture model (Section 5.3). Panel A: genuine bid distribution from non-FL losing bids with item/year FE. Panels B–C: cover-bid parameters from FL bids conditional on winning price. $\tilde{\delta} = 0.50$: typical FL bid exceeds winner by ∼65%. Panel D: BIC selects Regime 2 (ΔBIC $= -91{,}473$); $\hat{\sigma}_c/\hat{\sigma}_g = 0.72$ (cover bids less dispersed). Panel E: implied markup from n-conditional regression with item/year/PBU FE. Bootstrap SE from 500 tender-level replications.

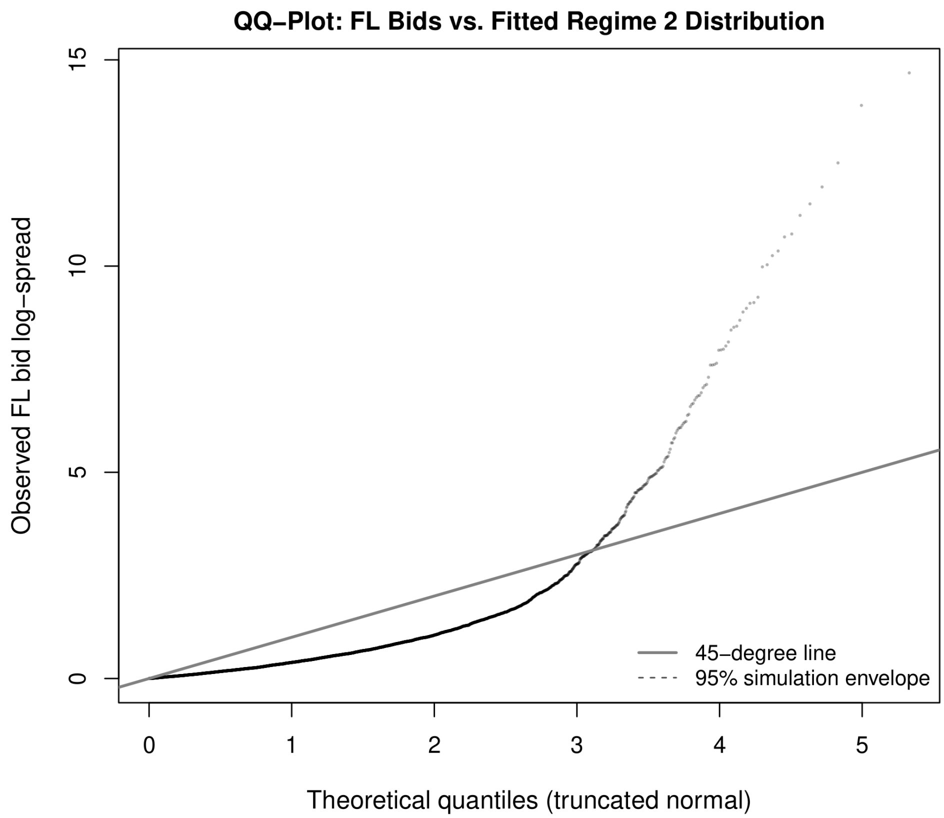

Figure B.1: QQ-plot: FL bid residuals vs. Regime 2 (truncated normal) theoretical quantiles. Dashed line: 45-degree reference.

Table C.1: Conditional Descriptive Statistics: Residualized Prices by FL Status

	FL-Present (losers = 1)	FL-Absent (losers = 0)
Panel A: Unconditional		
Mean log(price)	4.9241	2.4938
Difference	2.4303	
Panel B: Conditional on Item + Year + PBU FE		
Mean residualized price	0.0472	-0.0024
SD residualized price	1.5734	0.8704
Difference (gap)	0.0495	
Overlap coefficient	0.978	
t-statistic	8.81	
p-value	0.000000	
Cohen's d	0.0390	
Observations	79,452	1,574,949

Notes: Residuals from regressing log(negotiated price) on item, year, and purchasing unit (PBU) fixed effects. The gap represents the mean difference in residualized prices between FL-present and FL-absent tenders, after absorbing common item-level, temporal, and PBU-level variation. Overlap coefficient = $2\Phi(-|\Delta\mu|/\sqrt{\sigma_1^2 + \sigma_0^2})$ measures the proportion of the two residual distributions that overlap; values near 1 indicate near-identical distributions. Cohen's d measures the standardized effect size.

Table C.2: CADE Validation: Co-Participation and Permutation Test

Panel A: Co-participation with CADE Cartelists		
CADE procurement cartel convictions (2009–2019)	65	
CADE firms matched to BEC	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-convicted cartelists. 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 firms.

Table C.3: Price Regressions Excluding CADE-Involved Markets

	(1) General+PBU	(2) Pregão
FL presence	0.0616***	0.0901***
	(0.0213)	(0.0256)
Observations	1,622,954	534,956
CADE tenders dropped	12,514	

Notes: All tenders involving CADE-convicted firms excluded. SE clustered at item level. *** p<0.01, ** p<0.05, * p<0.1.

Table D.1: 2SLS Estimates: Effect of FL Presence on Tender Outcomes

	Log Price	Log Firms	Log Bids	Log Firms (excl.)
	(1)	(2)	(3)	(4)
Panel A: 2SLS (General + PBU FE)				
FL presence (instrumented)	0.1944**	0.5961***	0.5689***	0.4037***
	(0.0758)	(0.0268)	(0.0328)	(0.0278)
Observations	1,654,401	1,654,447	1,654,447	1,653,156
Panel B: 2SLS (Pregão only)				
FL presence (instrumented)	0.2210***	0.4258***	0.2666***	0.1564***
	(0.0819)	(0.0389)	(0.0613)	(0.0448)
Observations	546,549	546,559	546,559	545,269
Panel C: 2SLS (Item-group × Year FE)				
FL presence (instrumented)	3.6356***	0.7752***	0.7966***	0.5894***
	(0.1968)	(0.0334)	(0.0358)	(0.0356)
Observations	1,654,401	1,654,447	1,654,447	1,653,156
Item + Year + PBU FE	YES	YES	YES	YES
Instrument	FL supply at other PBUs (LOO)			

Notes: 2SLS using leave-one-out FL supply as instrument. FL supply (LOO) = count of FL firms active at other PBUs in the same item group and year. SE clustered at item level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

D Key Robustness Results

D.1 IV Results

D.2 Bajari–Ye Corrected Results

Table D.2: Bajari-Ye Tests: Corrected First Stage (No *n_bids*)

	Statistic	*p*-value
Panel A: First-Stage Auxiliary Regression		
R-squared	0.7701	
Observations	27,857,633	
Covariates: firm size (porte)	YES	
Covariates: firm age	YES	
Covariates: CNAE sector	YES	
Covariates: *n_bids*	NO (excluded)	
Panel B: Exchangeability (KS Test)		
KS statistic	0.1537	0.000000
Panel C: Conditional Independence		
FL mean pairwise product	4.2835	0.000000
(SE)	(0.0529)	
FL tenders with ≥2 bidders	29,636	

Notes: First-stage residuals from auxiliary regression of log(bid) on firm covariates with item and year fixed effects, estimated on losing bids only. *n_bids* is deliberately excluded as it is endogenous to collusion. Panel B tests whether FL and non-FL bid residual distributions are exchangeable. Panel C reports the mean pairwise product of residuals among FL firms within the same tender; a positive value indicates bid coordination (rejection of conditional independence).

D.3 Threshold Sensitivity

D.4 Regime Test

Reconciling pooled and individual bid dispersion. Table D.4 shows that FL presence raises total within-tender bid dispersion (coefficient 0.47–0.55 on log bid SD). This is *not* inconsistent with the Regime 2 finding that individual FL bids have lower dispersion than genuine bids ($\hat{\sigma}_c/\hat{\sigma}_g =$

Table D.3: Price Coefficient Across IQR Multiplier Thresholds

Multiplier	Threshold	FL Firms	Coef.	SE	Observations
1.0 × IQR	10	3,442	0.0791***	(0.0227)	1,654,401
1.5 × IQR	14	2,735	0.0636***	(0.0215)	1,654,401
2.0 × IQR	17	2,093	0.0598***	(0.0221)	1,654,401
2.5 × IQR	20	1,778	0.0543**	(0.0228)	1,654,401
3.0 × IQR	24	1,456	0.0501**	(0.0237)	1,654,401

Notes: DV: log negotiated price. Each row re-estimates the baseline specification (item + year + PBU fixed effects) using a different IQR multiplier to define the FL threshold. The threshold equals median $+k\times$ IQR of the tenders-count distribution among always-losers. The baseline (1.5× IQR) is highlighted. SE clustered at item level. *** p<0.01, ** p<0.05, * p<0.1.

Table D.4: Regime Test: Bid Dispersion

	(1) General	(2) General	(3) Pregão	(4) Convite
FL presence	0.5535***	0.4683***	0.3439***	0.5601***
	(0.0286)	(0.0265)	(0.0312)	(0.0256)
DV: Log bid SD	Total bid dispersion			

Notes: Positive coefficient: FL presence increases bid dispersion (consistent with Regime 1, complementary cover bidding).

0.72, Table B.1). The two statistics measure different objects: Table D.4 captures the pooled distribution of all bids in a tender, while $\hat{\sigma}_c/\hat{\sigma}_g$ compares individual-bid distributions conditional on the winning price. When FL firms enter a tender, they add bids clustered near $b^* + \epsilon$ but above the winning bid, mechanically widening the pooled distribution even though each FL bid individually exhibits lower variance.

D.5 Within-Band AUC Decomposition

Table D.5: Within-Band AUC: FL Screen Performance by Participation Volume

Decile	Tenders range	Firms	CADE-linked	AUC
1	1–1	1,684	14	0.500
2	1–2	1,684	10	0.613
3	2–3	1,684	19	0.551
4	3–5	1,685	14	0.563
5	5–8	1,684	18	0.467
6	8–13	1,684	20	0.510
7	13–19	1,685	38	0.493
8	19–30	1,684	40	0.559
9	30–57	1,684	55	0.500
10	58–1,126	1,685	147	0.633
Overall AUC (unconditional)				0.738
Pooled within-band AUC				0.539
Residualized AUC (demeaned)				0.505

Notes: AUC computed using participation count as the screening score and CADE co-participation as the binary outcome, within each decile of the always-loser participation distribution. Within-band AUC measures discrimination after controlling for volume. Residualized AUC uses participation count demeaned within deciles. Total always-losers: 16,843 (CADE-linked: 375). The ROC analysis in Table F.1 varies the IQR *multiplier* rather than the raw count, producing a higher overall AUC (0.94) because threshold variation spans a wider range of the participation distribution.

Table D.6: Detection Performance: FL Screen vs. Imhof-Style Bid-Level Screens

Screen	AUC	Data required	DeLong p
FL screen (this paper)	0.940	Participation only	—
Imhof CV only	0.758	Bid values	
Imhof spread only	0.789	Bid values	
Imhof composite	0.788	Bid values	0.0000
Combined (FL + Imhof)	0.608	Both	0.0000
Random classifier	0.500	None	—

Notes: AUC computed on 16,692 always-loser firms (98 CADE co-bidders, 16,594 non-CADE). Ground truth: co-participation with CADE-convicted cartelists. FL screen score: tender participation count (higher = more suspicious). Imhof composite: standardized mean of within-tender CV, excess kurtosis, skewness, and normalized bid spread across all tenders the firm participates in. DeLong p-value tests the null that the AUC difference from FL screen equals zero.

Table D.7: Horse-Race Regression: FL Screen vs. Imhof CV Flag

	(1) FL only	(2) Imhof only	(3) Both
FL presence	0.064***		0.084***
	(0.022)		(0.023)
Imhof CV flag		0.031***	0.021***
		(0.008)	(0.008)
R-squared	0.8860	0.8861	0.8862
Item + Year + PBU FE	YES	YES	YES
Observations	1,654,401	1,654,401	1,654,401
Correlation(FL, Imhof)		0.060	

Notes: DV: log negotiated price. Imhof CV flag equals one if the within-item-group coefficient of variation exceeds the median. Both screens contribute marginally to R-squared given the dominance of high-dimensional fixed effects; the coefficient changes document their independent information content. SE clustered at item level. *** p<0.01, ** p<0.05, * p<0.1.

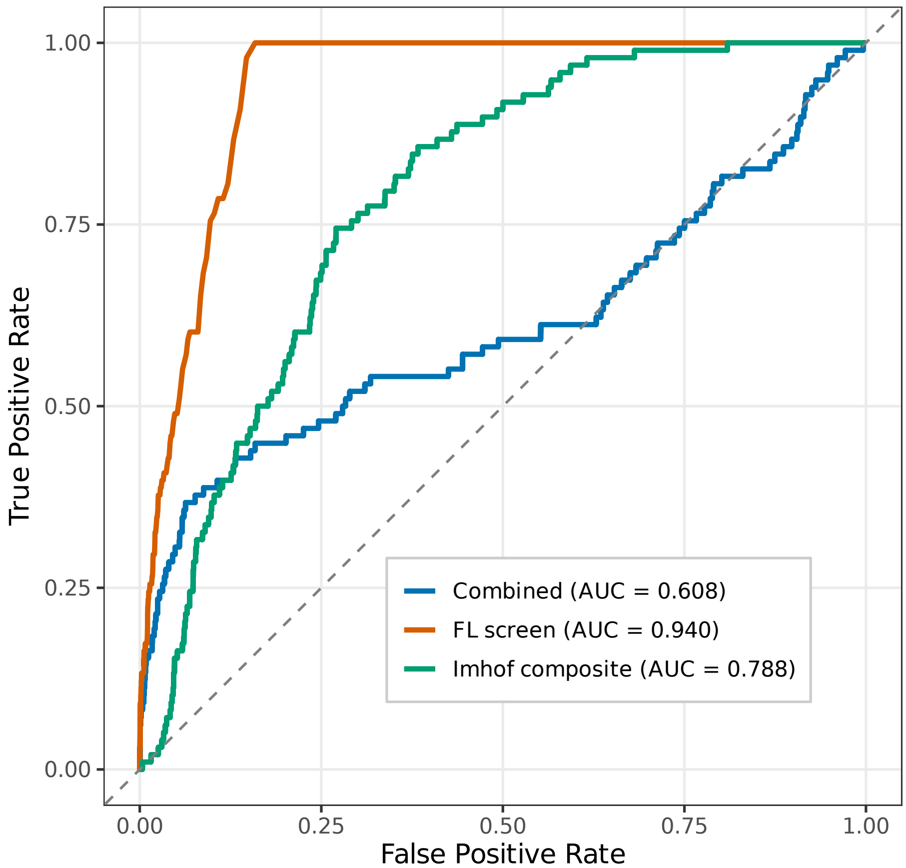

Figure D.1: ROC curves: FL screen vs. Imhof-style bid-level screens. Ground truth: co-participation with CADE-convicted cartelists among always-loser firms. The FL screen (AUC = 0.94) dominates the Imhof composite (AUC = 0.79; DeLong $p < 0.001$).

D.6 Imhof Comparison

D.7 Horse-Race: FL vs. Imhof Screens

E Staggered Difference-in-Differences

We construct a market-level panel (item × PBU × year) where treatment is first FL entry. The sample is restricted to markets with at least three pre- and three post-treatment observations, yielding 144,168 market-year observations across 19,777 markets (1,511 treated, 18,266 never-treated).

Table E.1: Staggered DiD: Callaway & Sant'Anna (Restricted Sample)

	C&S (2021)		TWFE	
Outcome	ATT	SE	Pre-trend	Post
Log Price	0.0143	(0.0417)	-0.0131	-0.0079
Log Firms (excl. FL)	-0.0220	(0.0342)	-0.0088	0.0113
Panel B: Statistical Power				
MDE (5%, two-sided)	0.0818			
MDE (80% power)	0.1168			
OLS benchmark	0.064			
Panel D: Sharp Entry Subsample				
C&S ATT	0.0143	(0.0419)		
Sharp entry markets	1,511 treated, 18,266 control			
Markets	19,777 (1,511 treated, 18,266 control)			
Sample restriction	≥3 pre & ≥3 post observations			

Notes: Panel A: C&S = Callaway & Sant'Anna (2021) doubly-robust estimator. Panel B: MDE = minimum detectable effect at stated significance/power levels. Panel C: Rambachan & Roth (2023) sensitivity bounds under smoothness restriction $\bar{M}$ on maximum change in the slope of the trend. Panel D: restricts treated markets to those with zero FL presence before treatment year.

E.1 Stacked Regression (Cengiz et al.)

As a complement to the Callaway & Sant'Anna estimator, Table E.2 reports a stacked regression following Cengiz et al. (2019). Each cohort-year sub-experiment pairs the treated markets entering in year g with all never-treated markets within a symmetric ±5-year event window, and cohort-specific market and year fixed effects absorb level differences across sub-experiments. This design eliminates the "bad comparison" problem—already-treated units serving as controls—that can bias standard TWFE in staggered settings.

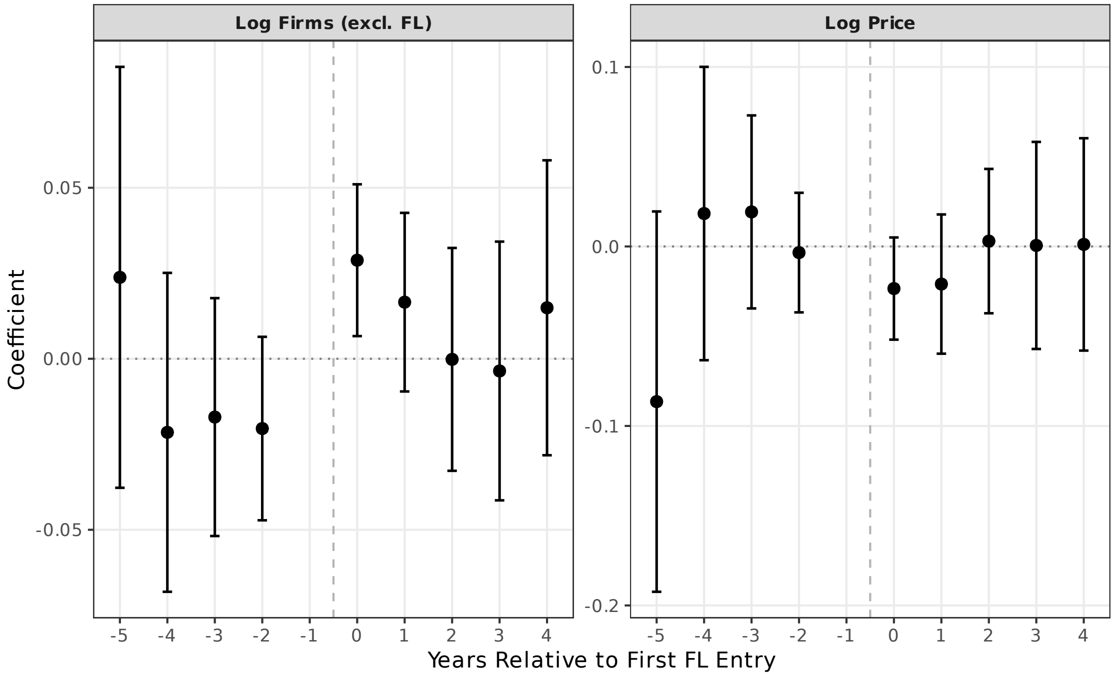

Figure E.1: TWFE event study: coefficients on years relative to first FL entry, by outcome variable.

Table E.2: Stacked DiD vs. Callaway & Sant'Anna

	Stacked (Cengiz et al.)		C&S (2021)	
Outcome	ATT	SE	ATT	SE
Log Price	-0.0059	(0.0138)	0.0143	(0.0417)
Log Firms (excl. FL)	0.0162*	(0.0084)	-0.0220	(0.0342)
Stacked obs.	715,116			
Cohorts	6			
Event window	±5 years			
Cohort × Market FE	YES		—	
Cohort × Year FE	YES		—	
Clustering	Market level			

Notes: Stacked regression following Cengiz et al. (2019). Each cohort-year sub-experiment includes treated markets entering in year g and all never-treated markets, within a symmetric ±5-year event window. Cohort-specific market and year fixed effects absorb level differences across sub-experiments. SE clustered at market level. C&S = Callaway & Sant'Anna (2021) doubly-robust estimator from Table E.1. *** p<0.01, ** p<0.05, * p<0.1.

F Additional Print Tables and Figures

Table F.1: ROC Analysis: FL Screening Performance by IQR Multiplier

IQR Multiplier	Threshold	Flagged Firms	TP	FP	TPR (Sensitivity)	FPR (1-Specificity)	Precision	Youden's J
0.5	6	5,098	98	5,000	1.000	0.299	0.019	0.701
1.0	10	3,442	98	3,344	1.000	0.200	0.028	0.800
1.5†	14	2,735	98	2,637	1.000	0.157	0.036	0.843
2.0	17	2,093	79	2,014	0.806	0.120	0.038	0.686
2.5	20	1,778	75	1,703	0.765	0.102	0.042	0.664
3.0	24	1,456	64	1,392	0.653	0.083	0.044	0.570
3.5	28	1,258	59	1,199	0.602	0.072	0.047	0.530
4.0	31	1,077	55	1,022	0.561	0.061	0.051	0.500
5.0	38	819	47	772	0.480	0.046	0.057	0.433
AUC					0.9368			

Notes: ROC analysis varying the IQR multiplier for the FL threshold (median + multiplier × IQR of participation counts among always-losers). Ground truth: 193 FL firms that co-bid with CADE-convicted cartelists. Total always-losers: 16,843 (positives: 98, negatives: 16,745). † = baseline threshold used in the paper; ‡ = Youden's J-optimal threshold.

Table F.2: FL Classification: IQR Rule vs. Model-Based Alternatives

Classifier	AUC	Data required	DeLong p
IQR rule (this paper)	0.937	Participation count	—
Structural score	0.562	Bid values	0.0000
Random Forest (7 features)	0.904	Bid values + firm chars	0.0416
Combined (IQR + structural)	0.894	Both	
Random classifier	0.500	None	—

Notes: AUC computed on always-loser firms with valid bid-level features. Ground truth: co-participation with CADE-convicted cartelists. IQR rule: participation count (higher = more suspicious). Structural score: log-likelihood ratio of mean bid spread under Regime 2 cover-bid distribution ($\hat{\epsilon} = 0.83$, $\hat{\sigma}_c = 1.19$) vs. genuine-loser spread distribution. Random Forest: 500 trees trained on 7 firm-level features (participations, mean/SD/median spread, % above winner, bid CV). DeLong p-value: null that AUC difference from IQR rule equals zero.

Table F.3: Bajari-Ye Pairwise Products with Tender Fixed Effects

	Item + Year FE		Tender FE	
	Mean Product	p	Mean Product	p
FL pairs	5.1597	0.000000	0.3795	0.000000
Non-FL pairs	2.2082	0.000000	0.8617	0.000000
Fake groups	5.8029	0.0000	0.5087	0.0000

Notes: Mean pairwise product of first-stage bid residuals. Columns 1–2: residuals from item + year FE regression (baseline). Columns 3–4: residuals from tender × item FE regression, which absorbs all tender-level shocks (reference price, commodity prices, PBU effects). Under the null of competitive bidding, the mean product equals zero.

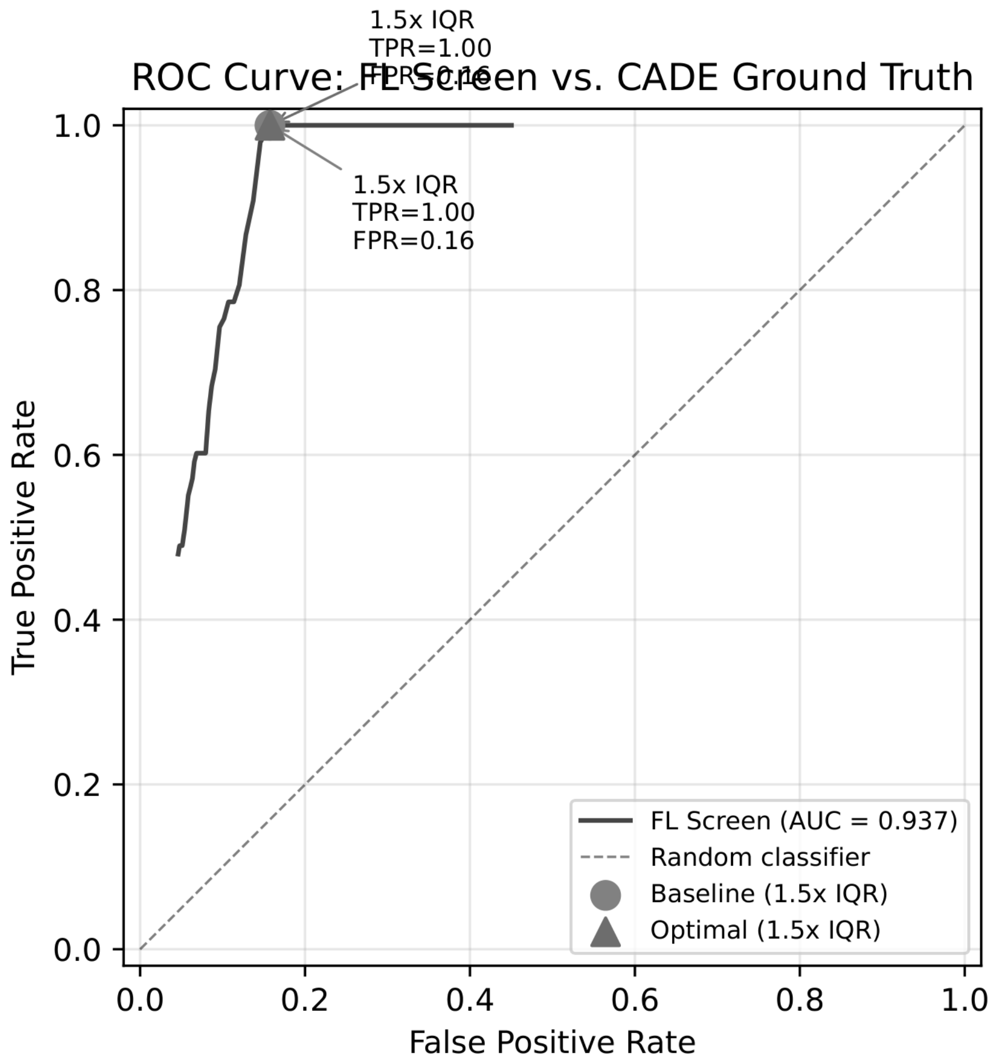

Figure F.1: ROC curve for the FL detection screen. Ground truth: co-participation with CADE-convicted cartelists among always-loser firms. AUC = 0.94 (95% CI: $[0.93, 0.95]$).

Table F.4: FL Price Effect by Oversight Proxy

Sub-sample	Coefficient	SE	N
By PBU size quartile:			
Q1 (lower oversight)	0.2138	(0.1670)	7,503
Q2 (lower oversight)	0.0111	(0.0513)	87,571
Q3 (higher oversight)	0.0660**	(0.0304)	302,258
Q4 (higher oversight)	0.0165	(0.0160)	1,257,069
By procedure type:			
Pregão (electronic, more transparent)	0.0933***	(0.0255)	546,549
Convite (sealed-bid, less transparent)	0.0382**	(0.0186)	1,107,852

Notes: If FL coefficient is smaller under higher oversight, this is consistent with the regime model where detection probability θ constrains cartel behavior. Item and year FE (PBU FE for procedure split). SE clustered at item level. *** p<0.01, ** p<0.05, * p<0.1.