Online Appendix for

The Price of Exclusion: SME Set-Asides in Public Procurement

Darcio Genicolo-Martins

This appendix documents the diagnostics behind the paper's identification and policy claims. The checks do not prove the maintained auction model. They discipline specific threats: institutional timing, reduced-form imbalance and placebo timing, willingness-to-supply recovery, protected-pool composition, bidder dependence and coordination, reference-price evolution, and static welfare sensitivity. The appendix also records where the interpretation is stable and where it is conditional, especially in pharmaceutical procurement. Appendix OA-A documents the institutional setting; OA-B reports reduced-form validation; OA-C covers willingness-to-supply recovery and structural estimation; OA-D reports mechanism robustness and conduct threats; and OA-E gives static welfare and policy-benchmark details.

A Institutional Details

A.1 Legal timeline

Brazil's federal SME procurement framework begins with LC 123/2006, which defined the SME categories used in procurement policy. The statute distinguishes MicroEmpresas, with annual revenue up to R$360,000, from Empresas de Pequeno Porte, with annual revenue up to R$4,800,000 after the LC 155/2016 update. LC 147/2014 made SME-only procurement mandatory for eligible purchases below R$80,000, but implementation depended on how procurement authorities interpreted the threshold.

In São Paulo, the relevant interpretation changed from a purchase-notice view to an item-by-item view. Under the restrictive view, a purchase notice whose total value exceeded the statutory threshold was not treated as SME-only even when individual items were below the threshold. The item-level interpretation made many more items eligible for SME-only treatment. BEC communicated the technical implementation through Comunicados 02/2017 and 03/2017, while operational take-up in Group 65 occurs around March 2018. I therefore treat March 2018 as the empirical cutoff: it is the mass-adoption date, not the first legal moment at which SME preferences existed.

Table OA-1: Institutional timeline

Instrument	Date	Role in the empirical design
LC 123/2006	2006	Federal SME statute; defines SME procurement treatment and preference tools.
13.122/2008	2008	São Paulo state SME procurement law.
LC 147/2014	2014	Makes SME-only procurement mandatory for purchases below R$80,000.
TCE-SP eTC 5509.989.15	2015	Illustrates the restrictive interpretation of the threshold.
PGE-SP Parecer 151/2017	2017	Supports the item-level interpretation used in BEC implementation.
BEC Comunicados 02/2017–03/2017	2017	Platform communication enabling SME-only and mixed-exclusivity functionality.
March 2018	2018	Mass take-up of SME-only treatment in Group 65; empirical cutoff.
TCE-SP eTC 9589.989.18	2018	Later ruling consistent with the expansive item-level interpretation.

The table separates legal authority from empirical timing. The design uses the observed mass take-up in Group 65 rather than treating any single legal document as the treatment date.

A.2 Platform, products, and sample

The empirical setting is the BEC electronic procurement platform, created under Decrees 45.085/2000 and 45.695/2001 and later used for Pregão auctions under Decree 49.722/2005. The appendix uses three nested empirical objects: the platform data universe, the reduced-form benchmark panel, and the structural Pregão sample. Sections OA-B, OA-C, and OA-D draw on different combinations of the three; this subsection defines them.

The platform data universe covers 36 months, from September 2016 through August 2019, and contains 4.8 million item-level transactions and 860,000 purchase offers. Group 65 accounts for 27 percent of platform transactions in this window, which is why the item-level reinterpretation has substantial empirical bite there. The reduced-form benchmark in Section OA-B uses this broader panel to compare Group 65 against the never-treated control groups (see Section OA-B for balance and placebo diagnostics).

The structural sample restricts attention to Group 65 Pregão auctions with at least 2 bidders and normalized bids b/p_t^{ref}, denoted c_ϵ, in $(0, 3]$. The upper-bound filter removes only 0.4 percent of bids, so the main results are not driven by discarding a large fraction of the auction data. The resulting sample contains 48,740 pre-period auctions and 49,253 post-period auctions; pharmaceutical Pregão auctions account for 59,997 of these. Each auction is classified by bidder type (SME versus non-SME from the BEC registry) and

product class (non-pharmaceutical versus the CMED-regulated pharmaceutical class). The classification is used throughout OA-C and OA-D.

B Reduced-Form Validation

The reduced-form evidence validates timing, sign, and approximate scale; it is not the source of the structural magnitude. The diagnostics below document pre-treatment imbalance that prevents a reduced-form-only design, placebo cutoffs that are smaller than the real cutoff but not all zero, and modern DiD estimators that attenuate the point estimate relative to TWFE. Together they support the role assigned in the main paper: external validation of the institutional episode rather than identification of the structural price magnitude.

B.1 Balance and scope of the comparison

The reduced form is a timing, sign, and scale check for the structural exercise, not the source of the decomposition. This distinction matters because Group 65 is an exposed product group rather than a randomized treated unit. In pre-period balance checks, Group 65 contributes 30718 items and the pooled controls contribute 92837 items across 76 never-treated groups. The largest absolute standardized difference is 0.599; 7 of 9 variables exceed 0.10 in absolute value and 4 exceed 0.25. These diagnostics justify the main text's conservative use of the DiD estimates. Table OA-2 reports the covariates behind this summary.

The imbalances in Table OA-2 are taken seriously. They do not by themselves test the identifying assumption. The relevant question is whether Group 65 had differential pre-trends or anticipatory breaks relative to the comparison groups, which is why the placebo and alternative-DiD diagnostics below are reported explicitly.

B.2 Placebo timing and alternative estimators

Placebo cutoffs before the reform are smaller in magnitude than the real-cutoff estimates: -0.013, -0.030, and -0.034. They are not zero in every case, so they do not support a strong reduced-form-only interpretation. They do support the narrower claim used in the main text: the largest movement occurs around the institutional take-up date and has the sign predicted by the set-aside.

Table OA-2: Pre-treatment covariate balance: Group 65 versus never-treated controls

	Group 65	**Control-group mean distribution**				**Std. diff.**
Variable	Mean	P25	Median	P75	Mean	Imbens–Rubin
Log reference price	4.036	4.057	5.385	6.874	5.657	-0.219
Bids per item	7.396	8.160	9.982	13.748	11.486	-0.176
Firms per item	2.784	3.203	3.681	4.347	3.765	-0.599
Share Pregão	0.645	0.358	0.471	0.805	0.559	0.445
Share items below R$80k	0.966	0.886	0.964	0.992	0.907	-0.037
Share SME bids	0.000	0.021	0.075	0.104	0.068	-0.567
Share São Paulo-state bidders	0.704	0.719	0.794	0.835	0.768	-0.331
Log bidder-buyer distance	4.126	3.579	3.858	4.197	3.853	0.160
Final/reference price	0.683	0.595	0.626	0.661	0.641	0.010

Sample: pre-period items in the DiD panel, restricted to auctions with at least two bidders. Group 65 contributes 30718 items; the 76 never-treated control groups contribute 92837 items. Standardized differences are Imbens–Rubin normalized differences (Imbens and Rubin, 2015). The table is not a parallel-trends test; it documents cross-group composition that the fixed-effects design must absorb. The differential-pre-trend question is addressed by the placebo-cutoff and alternative-DiD estimators in the next subsection (Table OA-4 and Table OA-3).

Table OA-3: Reduced-form diagnostics with reported standard errors

Diagnostic	Estimate	Standard error	Sample / notes
TWFE, headline estimator	-0.1080	0.0073	Item-clustered; 18-month window
BJS imputation estimator	-0.0560	0.0293	Robust imputation; post-treatment ATT
Callaway–Sant'Anna estimator	-0.0165	0.0932	Never-treated controls; simple post ATT
Phased-adoption check, full sample	-0.109	0.012	Platform enablement period retained
Phased-adoption check, excluding enablement months	-0.087	0.012	Months 690–697 excluded

Estimates use the paper's sign convention. Negative coefficients mean the treated Group 65 price/reference discount narrows after SME-only adoption, so prices rise relative to the pre-policy benchmark. The headline TWFE row uses item fixed effects and item-clustered standard errors. The modern DiD rows are reported because they discipline the reduced-form validation rather than because the paper relies on their magnitude for the structural decomposition.

Table OA-4: Placebo timing checks on log prices

Fake cutoff	Estimate	Standard error	Observations
September 2017	-0.013	0.0182	311,883
March 2017	-0.030	0.0132	215,611
June 2017	-0.034	0.0128	237,625
Real-cutoff benchmark range	-0.108 to -0.142	–	–

Placebo regressions use pre-treatment data only, item and month fixed effects, sealed-bid and log-quantity controls, and item-level clustering. The placebo estimates are not all zero, so they limit the force of a reduced-form-only design. Their magnitudes are smaller than the real-cutoff benchmark, which is why the main text uses the DiD as validation for timing and sign rather than as the source of the structural price effect.

Modern DiD estimators reduce precision and attenuate the point estimate. The BJS imputation estimate is -0.0560 and the Callaway–Sant'Anna estimate is -0.0165, compared with a panel-aggregated TWFE estimate of -0.1080. The reduced form is therefore not asked to carry the structural price magnitude or a reduced-form welfare calculation. The decomposition uses the auction model and treats the DiD as external validation for timing and sign.

B.3 Within-Group-65 falsification: medications versus other supplies

A standing concern with the reduced-form coefficient is that it could be driven by pharmaceutical-specific price dynamics rather than by the loss of competitive discipline. The medications subclass (CADMAT 6531) is subject to the CMED pharmaceutical price-regulation regime, which periodically adjusts ceilings for regulated medicines. If the reduced-form reading were contaminated by a CMED ceiling shift or by differential pharma inflation, the price coefficient should appear primarily in the medications cell and attenuate sharply outside it.

Table OA-5 reports the falsification. The price effect is present and statistically significant in *both* sub-cells. In the medications subclass (Panel A, CADMAT 6531) the coefficient is -0.166 without and -0.174 with PBU fixed effects; in the non-medication subclass (Panel B, hospital equipment and dental/consumable supplies) the coefficient is -0.099 without and -0.097 with PBU fixed effects. All four estimates clear the one-percent significance threshold with item-clustered standard errors.

The reading is that the medications subclass is larger in magnitude (-0.17 versus -0.10 for non-medications, a 75 percent gap) but not the sole locus of the effect. CMED regulation tends to *cap* pharmaceutical prices rather than push them up, so any contamination from CMED-side dynamics would attenuate the medications coefficient relative to its competition-channel counterpart, not amplify it. The fact that medications still exhibits the larger magnitude is consistent with deeper non-SME bidder pools in pharmaceutical procurement and with the larger structural exclusion share documented for the pharmaceutical class in the main paper's boundary-case discussion. The auxiliary table in this subsection does not deliver a structural identification claim; it rules out the specific reduced-form alternative that pharmaceutical inflation, rather than the loss of competitive discipline, is responsible for the headline price movement.

Table OA-5: Within-Group-65 falsification: medications (CADMAT 6531) versus other medical supplies

	Log prices		Log firms	
	Base	PBU FE	Base	PBU FE
Panel A: Medications (CADMAT 6531)				
$g65_{med} \times$ Pre	-0.1658^{***}	-0.1744^{***}	0.1516^{***}	0.1605^{***}
	(0.0174)	(0.0172)	(0.0100)	(0.0101)
Observations	573,080	573,080	678,547	678,547
Panel B: Other medical supplies (non-6531)				
$g65_{other} \times$ Pre	-0.0992^{***}	-0.0973^{***}	0.0444^{***}	0.0504^{***}
	(0.0092)	(0.0086)	(0.0057)	(0.0057)
Observations	593,920	593,920	705,616	705,616

18-month window; non-Group-65 items as the comparison in both panels. Item fixed effects in every column; PBU fixed effects added in the second column of each outcome block. Standard errors clustered at the item level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Panel A restricts Group 65 to the CMED-regulated medications subclass; Panel B restricts Group 65 to non-medication medical supplies (hospital equipment, dental products, consumable supplies). The fact that the price coefficient is statistically significant and economically large in Panel B is the load-bearing falsification: a confounder operating only through CMED-regulated medicines would not move the non-medication cell.

B.4 RAIS validation and the firm-size composition channel

The Group 65 set-aside operates on the BEC self-reported SME flag (*fornec_ enquad*). Two complementary questions discipline the reliance on that flag: whether the price effect survives restricting the sample to winners that are SMEs by an administratively-recorded firm-size measure orthogonal to BEC behavior, and whether the distance-widening channel is concentrated in non-SME winners as the geographic-catchment account predicts. Both questions are addressed by linking BEC winners to the RAIS 2017 establishment registry via CNPJ raiz (82.6 percent match rate among Group 65 winners).

Table OA-6 reports the DiD price and distance coefficients across five winner subsamples ordered by RAIS employment links: the full sample; winners matched to RAIS; winners with at most forty-nine employment links (the federal SME threshold by employment); the micro-firm restriction at at most nine employment links; and the not-large restriction below one hundred employment links. The price coefficient hovers between -0.09 and -0.12 across the five columns. The distance coefficient, by contrast, collapses from $+14.25$ kilometers in the full sample to a null -0.28 in the RAIS-SME column and remains statistically indistinguishable from zero in the micro-firm column.

Table OA-6: RAIS-validated winner subsamples (18-month window, item fixed effects)

	(1) Full	(2) Winner in RAIS	(3) SME (≤49)	(4) Micro (≤9)	(5) Not large (<100)
Panel A: Log price					
$g65 \times$ Pre	-0.1087^{***}	-0.1020^{***}	-0.1209^{***}	-0.0890^{***}	-0.1175^{***}
	(0.0120)	(0.0122)	(0.0100)	(0.0131)	(0.0100)
Observations	649,714	599,020	532,538	437,975	544,757
Panel B: Distance (km)					
$g65 \times$ Pre	$+14.25^{***}$	$+17.10^{***}$	-0.28	$+3.06$	$+6.68^{***}$
	(2.36)	(2.37)	(2.44)	(2.81)	(2.48)
Observations	649,714	599,020	532,538	437,975	544,757

18-month window; completed items. Item fixed effects in all columns; standard errors clustered at the item level. RAIS link via CNPJ raiz; employment cutoffs are 2017 RAIS establishment links. Column (1) reproduces the full-sample reduced form; column (2) restricts to winners matched to RAIS; columns (3)–(5) further restrict to SME-, micro-, and not-large-firm winners. The collapse of the distance coefficient between column (2) and column (3) is the falsification reported in the text: the geographic-catchment widening is carried by non-SME winners whose operational footprint spans larger radii. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Two readings emerge. First, the reduced-form price coefficient is not mechanically a composition effect on the SME-winner indicator: it remains present when the sample is

restricted to winners with at most forty-nine RAIS employment links. Composition matters, but so does within-SME discipline. Second, the distance widening is decomposed cleanly. The +14.25 kilometer full-sample effect vanishes when the sample is restricted to RAIS-validated SME winners, and remains null in the micro-firm restriction. Non-SME winners win at a wider geographic radius because their operational footprint absorbs longer-distance bids; RAIS-validated SMEs win locally regardless of regime. The geographic channel is consistent with transport-cost amortization on contract value and the firm-size composition shift it implies, not with a SME-only-driven shrinkage of the buyer's catchment.

C Willingness-to-Supply Recovery and Structural Estimation

The recovery diagnostics do not prove IPV. They ask the narrower question of whether the exclusion-dominant decomposition is mechanically generated by winner censoring, common auction-level scale heterogeneity, or cross-format inconsistency in willingness-to-supply recovery. The checks below pursue this question on four fronts: alternative treatments of the winner observation, a multiplicative auction-level heterogeneity correction, a cross-modality comparison between Pregão drop-outs and Convite GPV recovery, and KS distances that locate where the recovery is most credible and where policy rankings must be treated as conditional.

C.1 Drop-out bids and winner censoring

In a Pregão reverse auction, losing firms exit as the price clock falls. Under the maintained IPV clock interpretation, those exits reveal willingness to supply at the auction scale. The winner is censored because the winning firm does not need to reveal the lowest price at which it would still supply. The baseline estimator treats the winning price as a tight upper-bound observation; the appendix checks a losers-only estimator and a Turnbull NPMLE (Turnbull, 1976) that treats the winning observation as left-censored.

The rejections in Table OA-9 are not treated as failed robustness checks. They are the empirical reason for using the observed post-policy SME pool in the baseline and reporting strict primitive invariance as a bounding exercise. The policy ranking is therefore emphasized where it survives that choice, especially in non-pharmaceutical markets.

Table OA-7: Winner-censoring and willingness-to-supply recovery checks

Estimator	Net price effect		Deviation from baseline	
	Non-pharma	Pharma	Non-pharma	Pharma
All bidders, baseline	0.259	0.308	–	–
Losers only	0.275	0.347	6%	13%
Turnbull NPMLE	0.246	0.357	5%	16%

Net price effects are $p_{S_3} - p_{S_1}$, normalized by the buyer reference price. The largest Turnbull deviation from the baseline is 16 percent; in non-pharmaceuticals it is 5 percent.

The conclusion is not that censoring is irrelevant. It is that the central price-formation ranking does not hinge on the baseline treatment of the winner. The Turnbull estimator still gives exclusion shares of 74.0 percent in non-pharmaceuticals and 82.0 percent in pharmaceuticals.

Table OA-8: Willingness-to-supply quantiles under alternative winner treatments

			Median cost $c_{0.5}$			Upper-quartile cost $c_{0.75}$		
Class	Type	Period	Losers	All	Turnbull	Losers	All	Turnbull
Non-pharma	SME	Post	0.922	0.861	0.840	1.215	1.097	1.117
Non-pharma	SME	Pre	0.947	0.881	0.862	1.314	1.193	1.212
Non-pharma	Non-SME	Post	0.802	0.762	0.726	1.068	0.964	0.971
Non-pharma	Non-SME	Pre	0.768	0.719	0.684	1.028	0.918	0.932
Pharma	SME	Post	0.860	0.799	0.781	1.078	0.966	1.002
Pharma	SME	Pre	0.863	0.800	0.710	1.441	1.188	1.245
Pharma	Non-SME	Post	0.674	0.657	0.633	0.814	0.774	0.765
Pharma	Non-SME	Pre	0.619	0.604	0.574	0.762	0.725	0.708

Losers-only uses only drop-out bids and is upward-biased because it omits the lowest-willingness-to-supply bidder under the maintained IPV interpretation in each auction. All-bidders treats the winner's final price as a point observation and is downward-biased if the final price exceeds the underlying model cost. Turnbull NPMLE treats the winner as left-censored at the final price.

C.2 Auction-level heterogeneity and cross-modality comparison

The baseline removes auction-level scale heterogeneity before simulating counterfactual bidder pools. The Pregão intraclass correlations range from 0.36 for non-pharmaceutical SMEs to 0.59 for pharmaceutical non-SMEs. These magnitudes are comparable to settings where auction-level unobserved heterogeneity is important: Krasnokutskaya (2011) reports an ICC benchmark of 0.66 in highway procurement.

The cross-modality check compares Pregão drop-out recovery with a GPV first-price recovery from Convite auctions (Guerre et al., 2000). The operational Kolmogorov–Smirnov distance threshold is 0.05. The cleanest falsification cell, pharmaceutical non-SMEs in Convite, has distance 0.032. Non-pharmaceutical Pregão is close to the threshold (0.0486), while pharmaceutical Pregão before the heterogeneity correction is farther away (0.0722). This pattern supports the maintained correction but also motivates the main text's cautious treatment of IPV.

Table OA-9: Primitive-stability and cross-modality diagnostics

Diagnostic cell	KS distance	Threshold	Interpretation
Non-pharma Pregão, raw drop-outs	0.0486	0.05	Close to the operational threshold.
Pharma Pregão, raw drop-outs	0.0722	0.05	Fails the threshold before common-scale cleaning.
Non-pharma Pregão, UH-clean	0.0613	0.05	Moves above the threshold after removing common scale.
Pharma Pregão, UH-clean	0.1407	0.05	Remains far from strict primitive invariance.
Non-pharma Convite GPV	0.0514	0.05	Borderline cross-modality comparison.
Pharma Convite GPV, non-SME	0.032	0.05	Cleanest validation cell; passes the threshold.
Convite pharma non-SME, post-UH	0.0225	0.05	Continues to pass after the heterogeneity correction.

The table reports Kolmogorov–Smirnov distances for diagnostics used to discipline the willingness-to-supply recovery step. The goal is not to prove primitive invariance in every cell; the pharmaceutical Pregão cells visibly reject that strong restriction. The point is to locate where the recovery is most credible and where the paper must treat policy rankings as conditional.

D Mechanism Robustness and Conduct Threats

The mechanism checks ask whether the exclusion-dominant decomposition survives alternative assumptions about protected-pool composition, bidder dependence, coordination, and reference-price evolution. Composition is not made irrelevant, conduct is not declared competitive, and reference-price changes are not assumed away; what the checks establish is narrower. The price-formation ranking is not mechanically generated by any one of those channels. Welfare-ranking sensitivity is the more conditional object, and it is also the locus of the pharmaceutical boundary case.

D.1 Strict invariance and protected-pool composition

The decomposition separates the effect of lost non-SME competition from the protected-pool offset. The weakest interpretive link is the offset, because $S_3 - S_2$ combines addi-

tional SME participation with changes in the active SME willingness-to-supply distribution, interpreted as a cost distribution under the maintained model. The strict-invariance benchmark therefore imposes the pre-policy SME willingness-to-supply distribution on the post-policy SME pool.

Table OA-10: Mechanism robustness

Specification	Non-pharma (main result)		Pharma (boundary case)	
	Net effect	Excl. share	Net effect	Excl. share
Baseline	0.227	72.0%	0.309	68.8%
Strict invariance	0.29	85%	0.47	79%
Turnbull recovery	0.246	74.0%	0.357	82.0%
18-month window	0.272	79.7%	0.292	67.0%
12-month window	0.267	75.4%	0.310	73.0%
6-month window	0.269	76.4%	0.338	73.6%

Net effects are normalized by the reference price. Exclusion share is the absolute share of the decomposition attributed to removing non-SMEs from the price-forming pool. Rows combine different diagnostic families: strict invariance disciplines the protected-pool composition margin, Turnbull disciplines winner censoring, and window rows discipline timing sensitivity. The pharmaceutical boundary case is the scope condition identified in main-text §4.4; welfare-ranking inversion under strict invariance is reported in Table OA-16.

The non-pharmaceutical result is stable: the exclusion component remains the dominant force under baseline, strict invariance, Turnbull censoring, and alternative windows. The pharmaceutical result is informative but more conditional. Post-policy SME turnover is larger in pharmaceuticals: 61.9 percent of post-policy SME firms are new relative to the pre-period pool and these firms account for 37.8 percent of post-policy SME bids. In non-pharmaceuticals, new firms account for 23.0 percent of post-policy SME bids. Pharmaceuticals are therefore a boundary case for policy ranking rather than the headline market.

D.2 Entry counts, filters, and dependence

The baseline summarizes entry using class-by-period arrival rates. Window checks show little drift in the non-pharmaceutical exclusion share: 4.3 percentage points across the 18-, 12-, and 6-month windows. The pharmaceutical drift is larger, at 6.6 points, again matching the interpretation that pharma is composition-sensitive.

Cost-filter checks also preserve the qualitative result. Tightening the filter reduces the non-pharmaceutical net effect by 15 percent and the pharmaceutical net effect by 8

percent; loosening it raises the effects by at most 1 percent and 6 percent, respectively. Allowing within-auction cost correlation up to $\rho_c = 0.3$ moves the exclusion share by at most 5 percentage points and the total effect by at most 10 percent.

Replacing the Poisson bidder-count draws in the baseline simulation with sampling from the empirical class-period-type bidder-count distributions attenuates net effects but leaves the exclusion-dominant decomposition intact. Under empirical counts, the non-pharmaceutical net effect is 0.171 (versus baseline 0.227), with exclusion component 0.307 and protected-pool offset -0.136; in pharmaceuticals the net effect is 0.223 (versus baseline 0.309), with components 0.537 and -0.314. The corresponding exclusion shares are 69.4 percent in non-pharmaceuticals and 63.1 percent in pharmaceuticals (against baseline 72.0 and 68.8 percent). The Poisson specification therefore slightly overstates the dispersion of bidder counts, which inflates the simulated price spread; the ranking of the two channels is unaffected.

Table OA-11: Filter, window, and dependence sensitivities

Sensitivity	Non-pharma	Pharma	Interpretation
Tight c_ϵ filter: net effect	0.215	0.303	Smaller net effect after removing high normalized bids.
Baseline filter: net effect	0.254	0.330	Benchmark filter used in the robustness registry.
Tight-filter drop	15%	8%	Attenuation relative to the benchmark filter.
Loose-filter increase	1%	6%	Maximum increase under the looser filter.
Exclusion-share drift across windows	4.3 pp	6.6 pp	Difference across 18-, 12-, and 6-month windows.
Dependence grid	$\rho_c \in \{0, 0.1, 0.2, 0.3\}$		Total-effect drift at most 10%; exclusion-share drift at most 5 pp.

Net effects are $p_{S_3} - p_{S_1}$, normalized by the reference price. The dependence sensitivity allows within-auction cost correlation up to $\rho_c = 0.3$. The table does not require every specification to have the same level. It asks whether the exclusion-dominant decomposition survives the main filter, window, and dependence perturbations.

D.3 Coordination screens

Coordination is a first-order threat because repeated bidder pairs could make drop-out prices reflect conduct rather than willingness to supply. The screens detect baseline clustering; the identifying question for this paper is narrower: does clustering intensify after non-SMEs are removed? The Conley-style screen is stable in non-pharmaceuticals: realized shares are 16.9 percent before and 16.8 percent after the cutoff, compared with null means of 10.6 and 10.2. In pharmaceuticals, the realized share falls from 27.6 to 24.4 percent. Bajari–Ye-style ratios show the same directional pattern: the non-pharmaceutical ratio falls from 2.63 to 1.83, and the pharmaceutical ratio falls from 1.29

to 1.11.

These screens weaken the differential-coordination story, not the broader possibility of noncompetitive conduct. Residual baseline clustering remains a limitation. The diagnostics answer the narrower threat to the decomposition: the main post-policy price increase is not obviously driven by a new coordination shock among the surviving SME pool.

Table OA-12: Bidder-pair coordination screens

Stratum	Conley realized	Conley null mean	Bajari–Ye T_1 obs/null
Non-pharma Pre	16.9%	10.6%	2.63
Non-pharma Post	16.8%	10.2%	1.83
Pharma Pre	27.6%	16.3%	1.29
Pharma Post	24.4%	14.4%	1.11

The Conley-style statistic is the share of auctions in which the closest pair of UH-clean residual bids lies within a narrow reference-price band. The null resamples bid sequences within cell while preserving bidder counts. The Bajari–Ye statistic is the ratio of the observed maximum pair co-bidding count to its permutation-null mean among repeated bidder pairs. Both screens detect baseline clustering. The identifying question here is the pre/post direction: neither statistic rises after the set-aside expands.

D.4 Reference-price validation

All structural prices are normalized by buyer reference prices, so endogenous reference-price evolution at the cutoff could contaminate reduced-form interpretation and structural levels. The direct evidence below shows that p^{ref} moves weakly relative to p^{final}. A uniform p^{ref} shift would not mechanically generate the decomposition, which is constructed in p^{ref}-normalized units; the relevant remaining concern is heterogeneous reference-price movement correlated with the reform, and the quartile placebo addresses that channel directly.

The TWFE DiD on $\log p^{\text{ref}}$ under the main specification (Table OA-13) is -0.0273 (SE 0.0105) in non-pharmaceuticals and -0.0325 (SE 0.0124) in pharmaceuticals. Both are an order of magnitude smaller than the corresponding DiD on $\log p^{\text{final}}$ in each subset (-0.0741 and -0.1444). BJS imputation gives slightly larger magnitudes (-0.1048 NP, -0.0598 PH). Callaway-Sant'Anna estimates at group-level aggregation have wide standard errors (above 0.3 in each subset) and are statistically indistinguishable from zero.

The internal placebo by quartile of $|\Delta \log p^{\text{ref}}|$ is constructed by splitting items into quartiles of the absolute pre-to-post change in $\log p^{\text{ref}}$ and re-running the main DiD on

Table OA-13: Reference-price evolution: DiD on $\log p^{\text{ref}}$ across estimators and classes

	Non-pharma			Pharma		
Estimator	$\hat{\delta}$	SE	N	$\hat{\delta}$	SE	N
TWFE	-0.0273	0.0105	800,128	-0.0325	0.0124	784,626
Callaway–Sant'Anna	0.1208	0.3129	—	-0.2057	0.3062	—
BJS imputation	-0.1048	0.0259	—	-0.0598	0.0201	—
Sanity: same spec on $\log p^{\text{final}}$ *(TWFE)*						
Coefficient	-0.0741	—	—	-0.1444	—	—
Heterogeneity by $\lvert\Delta \log p^{\text{ref}}\rvert$ *quartile (DV:* $\log p^{\text{final}}$*)*						
Q1 (low)	-0.0438	0.0102	—	-0.0923	0.0113	—
Q4 (high)	-0.1871	0.0367	—	-0.4538	0.1775	—

TWFE uses item and month fixed effects with item-clustered standard errors, mirroring the main DiD specification. Callaway-Sant'Anna aggregates to group level for tractability; the standard errors are wide. BJS uses item-level imputation. The sanity row reports the main DiD on $\log p^{\text{final}}$ in the same NP/PH sample. The quartile rows split items by absolute change in $\log p^{\text{ref}}$ pre-to-post and re-run the DiD on $\log p^{\text{final}}$ within each quartile.

$\log p^{\text{final}}$ within each quartile. Items in the lowest quartile, where the reference price was effectively stable across the cutoff, still display a price effect of -0.0438 in non-pharmaceuticals and -0.0923 in pharmaceuticals, about two-thirds of the class-specific DiD coefficient on $\log p^{\text{final}}$. Items in the highest quartile, where the reference price moved most, show larger effects (-0.1871 NP and -0.4538 PH). The headline price effect is therefore not concentrated only among items whose reference prices moved. A McCrary density test at the R$ 80,000 item-eligibility threshold detects no bunching (post-period $T = 1.06$, $p = 0.29$; pre-period $T = -1.43$, $p = 0.15$). The density evidence does not indicate systematic splitting or restructuring around the threshold. This does not rule out all buyer-side adaptation, but it rules against the canonical threshold-gaming channel.

Independent corroboration comes from the second-lowest bid in the iterative Pregão phase. The DiD on $\log b^{(2)}$ delivers -0.0390 (SE 0.0206) in non-pharmaceuticals and -0.1307 (SE 0.0351) in pharmaceuticals. The estimates are constructed from raw bid units, with no p^{ref} normalization, and corroborate the order-statistic mechanism directly. The structural net effects of 34 percent and 47 percent of the open-regime price exceed the reduced-form benchmark of 12 percent because the structural exercise simulates the order statistic under fixed bidder pools while the reduced form averages over take-up, controls, and product-time variation. That gap is a reason to interpret levels carefully,

not a reason to discard the mechanism.

E Static Welfare and Policy Benchmark Details

The welfare and preference grids are static benchmark exercises under recovered primitives. They are not implementation forecasts. The grids report welfare losses across λ and across SME price-preference rates, the indifference welfare weight $w_\star^{\text{SME}}$ at which the planner is indifferent between full exclusion and the 10 percent preference benchmark, and the strict-invariance sensitivity that delineates the pharmaceutical boundary case. The annual scaling block serves only to translate the static per-auction comparisons into a sense of magnitude.

E.1 Lambda grid

The main text reports welfare losses at $\lambda = 0.30$. The welfare loss is

$$L(\lambda) = DWL^{alloc} + \lambda \Delta G,$$

where DWL^{alloc} is the allocative loss and ΔG is the extra government outlay. The grid ranges from 0.15 to 0.45 in increments of 0.10. Moving λ by 0.10 changes the welfare loss by about 3 percentage points in non-pharmaceuticals and 5 points in pharmaceuticals.

Table OA-14: Welfare loss across marginal-cost-of-public-funds values

Class	$\lambda = 0.15$	$\lambda = 0.20$	$\lambda = 0.30$	$\lambda = 0.40$	$\lambda = 0.45$
Non-pharma	24.1%	25.7%	28.9%	32.2%	33.8%
Pharma	38.1%	40.3%	44.8%	49.3%	51.6%

Losses are reported as percentages of the open-regime simulated price.

E.2 Preference grid and welfare weights

The 10 percent SME price preference is a benchmark instrument. It keeps non-SMEs inside the auction and tilts allocation through the scoring rule. In the simulated preference grid from 0 to 30 percent, the 10 percent preference is the cheapest welfare-dominant preference threshold in the current grid. At 10 percent, the price shift is -0.0036 in non-pharmaceuticals and +0.0015 in pharmaceuticals, while SME win rates rise by 4.3 and 1.4 percentage points.

The preference is not a costless substitute for the set-aside. At 30 percent, the welfare loss remains 4.99 percent in non-pharmaceuticals and 1.92 percent in pharmaceuticals, but the instrument still delivers less redistribution than full exclusion. The relevant comparison is therefore a frontier: how much additional SME surplus, access, or dynamic capacity does the planner require to justify removing non-SMEs from the price-forming pool? In the main specification, the implicit SME welfare weights required to prefer the set-aside to the 10 percent preference are 2.42 in non-pharmaceuticals and 2.61 in pharmaceuticals. Under strict invariance, the non-pharmaceutical comparison continues to require a weight above one, while the pharmaceutical comparison falls to 0.7.

Table OA-15: Preference grid: price effects, SME win rates, and welfare loss

Class	Preference	$\bar{p}$	Δ vs. S_1	% of V0 Δ	SME win rate	Welfare loss
Non-pharma	0%	0.772	+0.0046	2.0	17.1%	0.14%
Non-pharma	5%	0.768	+0.0015	0.7	20.0%	0.00%
Non-pharma	10%	0.764	-0.0028	-1.2	21.4%	0.25%
Non-pharma	15%	0.766	-0.0013	-0.6	24.6%	0.00%
Non-pharma	20%	0.775	+0.0081	3.5	27.8%	1.40%
Non-pharma	25%	0.790	+0.0227	9.7	30.2%	2.43%
Non-pharma	30%	0.809	+0.0420	17.9	34.0%	4.99%
Non-pharma	Full set-aside	1.001	+0.2342	100.0	100.0%	28.9%
Pharma	0%	0.648	-0.0061	-2.1	16.2%	0.00%
Pharma	5%	0.670	+0.0161	5.5	16.5%	1.24%
Pharma	10%	0.648	-0.0061	-2.1	17.6%	0.00%
Pharma	15%	0.655	+0.0009	0.3	19.6%	0.00%
Pharma	20%	0.655	+0.0015	0.5	19.8%	0.00%
Pharma	25%	0.656	+0.0019	0.7	22.1%	0.00%
Pharma	30%	0.676	+0.0219	7.5	25.7%	1.92%
Pharma	Full set-aside	0.948	+0.2941	100.0	100.0%	44.8%

Bayes–Nash simulation with B = 3,000 Monte Carlo draws per cell. Preference rate k scores SME bids by a factor $1-k$ for winner selection while paying the actual winning bid. Δ is the simulated price effect relative to S_1, in reference-price units. Welfare loss is $(DWL^{alloc} + \lambda\Delta G)/p_{S_1}$ at $\lambda = 0.30$. Entries that are negative before rounding are reported as 0.00% because the underlying welfare effect is economically negligible and lies within the Monte Carlo simulation tolerance of the BNE solver. The preference-rate rows are from the preference-grid simulation; the full-set-aside welfare loss is reported as the headline welfare-arithmetic value (28.9% non-pharma, 44.8% pharma, from the generated value registry) so it matches the main text and the λ grid.

E.3 Annual scaling

Annual scaling translates the static per-auction welfare comparison into a sense of magnitude; it is not part of the identification exercise and is not used as a policy forecast. The translation is mechanical and inherits all the scope conditions of the structural exercise, including the non-pharmaceutical/pharmaceutical split.

Table OA-16: Set-aside versus preference ranking across λ

	Non-pharma		Pharma		
λ	Loss(V0)	Ranking	Loss(V0)	Ranking, main	Ranking, strict inv.
0.15	24.1%	V3 > V0	38.1%	V3 > V0	V0 > V3
0.20	25.7%	V3 > V0	40.3%	V3 > V0	V0 > V3
0.30	28.9%	V3 > V0	44.8%	V3 > V0	V0 > V3
0.40	32.2%	V3 > V0	49.3%	V3 > V0	V0 > V3
0.45	33.8%	V3 > V0	51.6%	V3 > V0	V0 > V3

$V0$ is the full SME-only set-aside and $V3$ is the 10 percent preference benchmark. The non-pharmaceutical ranking is stable across the λ grid. In pharmaceuticals, the ranking depends on how the post-policy SME pool is modeled, not mainly on the marginal cost of public funds.

For reference, Group 65 annual reference outlays are R$345 million in non-pharmaceuticals and R$363 million in pharmaceuticals. Applying the structural welfare losses and adherence rates between 30 and 70 percent gives an annual welfare-cost range of R$38–89 million, or US$11–25 million at an exchange rate of 3.50 R$/US$. This scaling is useful for magnitude, but it should not be read as a claim about all procurement categories, about dynamic long-run SME capacity, or about general-equilibrium adjustments outside the auction stage.

References for the Online Appendix

Guerre, E., I. Perrigne, and Q. Vuong (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica 68*(3), 525–574.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* New York: Cambridge University Press.

Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *Review of Economic Studies 78*(1), 293–327.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological) 38*(3), 290–295.