

Online Appendix

Sourcing under Sanctions: Judicial Urgency and Pharmaceutical Procurement Costs

Darcio Genicolo-Martins^a, Paulo Furquim de Azevedo^a

^a*Inspere Institute of Education and Research, São Paulo, Brazil*

Supplementary material for the short paper. The appendix reports only the classifier and sample checks, selection bounds, diagnostic evidence, mechanism diagnostics, and fiscal procurement-cost calculation needed to support the main text. All empirical values are imported through `values.tex` or generated tables.

A. Classifier and Sample Construction

The regime classifier operates at the purchase-order/tender-notice level, before classified orders are linked to purchase-offer-item (POI) records on BEC. This is the institutional level at which legal status is observed; prices, quantities, winners, and bidder participation are then measured at the POI level.

Table A.1: Purchase-regime classifier validation.

Validation object	N	Agree.	Prec.	Recall	F1	Source
Exact agreement	179,148	98.6%	–	–	–	Ground truth
Judicial class	–	–	0.87	1.00	0.93	Ground truth
Administrative class	–	–	0.92	0.99	0.96	Ground truth
Urgent-class macro-F1	–	–	–	–	0.94	Urgent classes
Judicial class	34,191	–	0.88	0.99	0.93	ML hold-out
Administrative class	34,191	–	0.91	0.98	0.94	ML hold-out

Notes: The classifier operates at the purchase-order/tender-notice level. The table is generated from the production classification report and classifier code. Exact agreement counts 176,713 matched ground-truth labels. The empirical POI-level data link these classified regimes to item-level procurement records.

Validation against 179,148 hand-coded purchase orders gives 98.6% exact agreement. The urgent-class F1 scores are 0.93 for judicial purchases and 0.96 for administrative purchases. The main remaining concern is differential error across urgent regimes among high-cost items; the validation table reduces, but cannot eliminate, that concern.

Table A.2 reconciles the upstream classification universe with the downstream empirical samples. The ordinary-versus-urgent sample supports the procurement effect of externally imposed legal urgency. The urgent panel supports the selection-bounded sanction margin. The firm-buyer-item sample is narrower by design because it tests same-firm pricing, not total procurement costs.

Table A.2: Sample construction and empirical units.

Object or sample	Size	Unit and use
Classifier universe	764,362	Purchase orders/tender notices classified into procurement regimes.
Full BEC pharmaceutical file	479,330	Purchase-offer-item observations after linking BEC item records to regime classifications.
Analysis sample	226,305	POI observations satisfying core item and variable restrictions.
Winners-only price sample	196,883	Accepted winning bids used in negotiated-price regressions.
Urgent panel	56,803	Administrative and litigated urgent winning bids used for the under-the-gun comparison.
Both-regime urgent cells	5,581	Item-by-year-month cells containing both administrative and litigated urgent purchases.
Singleton urgent cells	11,972	Comparable urgent cells containing only one urgent regime.
Firm-buyer-item triples	1,206 triples; 4,573 observations	Same supplier, buyer, and item comparisons used to isolate within-firm pricing.

Notes: This table separates the classifier unit from the empirical regression units. POI denotes purchase-offer-item. Price regressions use accepted winning bids, while classifier validation operates at the purchase-order/tender-notice level.

B. Selection Bounds and Inference

Administrative requests are screened before purchase, so the administrative-versus-litigated coefficient is bounded rather than point-identified. The main paper uses Lee trimming within item \times year \times PBU strata. Trimming affects 12,038 of 44,654 strata, with mean rate 26.9% and maximum 100.0%. Coefficients are estimated as administrative minus litigated; tables report the reader-facing litigated-over-administrative gap.

Table B.1: Lee bounds under alternative trimming strata.

Strata	Admin coef.	Gap (%)	Trim (%)
item \times year	[-0.326, -0.202]	[22.4, 38.6]	23.1
item \times year \times PBU	[-0.192, -0.148]	[15.9, 21.1]	26.9
item \times year-month \times PBU	[-0.202, -0.202]	[22.4, 22.4]	26.8

Notes: The outcome and second-stage fixed effects are held fixed across rows: log negotiated price with item, year, and PBU fixed effects. Rows vary only the strata used to compute the administrative trimming share. The item \times year \times PBU row is the preferred specification in the main analysis. Coefficients are administrative minus litigated log prices; percentage gaps are reported as litigated-over-administrative prices. In the finest item \times year-month \times PBU strata, cells contain too few administrative observations for top-tail and bottom-tail trimming to remove different observations, so the lower and upper bounds coincide; this row is reported for completeness, and the preferred bounds use the coarser item \times year \times PBU strata.

Table B.1 varies the trimming partition. The positive bounded gap is not a point-identified sanction effect, but it is stable enough to support the selection-bounded sanction-margin interpretation used in the main text.

Table B.2: Wild cluster bootstrap inference on the UTG coefficient.

Specification	$\hat{\beta}$	Asy. SE	t	p_{boot}	95% CI (log-points)
Preferred (item + year + PBU)	-0.259	0.092	-2.80	0.008	[-0.469, -0.039]
Tightest (item×year-month + PBU)	-0.283	0.116	-2.44	0.039	[-0.549, -0.001]

Notes: Rademacher wild cluster bootstrap on the Admin coefficient with $B = 999$ replicates and restricted-null residuals; clustering at the PBU level. Negative coefficients mean litigated purchases are more expensive than administrative ones, in log-points. Source: `analysis/44_wild_bootstrap.R`.

PBU clusters are few and uneven. Table B.2 reports Rademacher wild-cluster inference for the preferred and tightest under-the-gun specifications. The preferred specification has $p = 0.0080$; the item-by-year-month specification has $p = 0.0390$.

C. Identification Diagnostics

The first diagnostic uses items never observed in litigation. If the urgent-price pattern were a generic feature of BEC dynamics or administrative urgency, the same pattern should appear in that sample. It does not: the negotiated-price placebo coefficient is -0.020 with standard error 0.032.

Table C.1: Placebo Test on Never-Litigated Items

	Negotiated price		Reference price	
	Never-litigated	Main sample	Never-litigated	Main sample
Urgent purchase	-0.020 (0.032)	0.051*** (0.015)	-0.031 (0.045)	0.031** (0.014)
Observations	39,283	196,883	46,874	226,278
R^2	0.834	0.868	0.811	0.852
Item FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
PBU FE	Yes	Yes	Yes	Yes

Notes: The never-litigated sample contains items with zero litigated purchases during the sample period and variation between ordinary and administrative urgent purchases. The main sample contains items observed in both ordinary and litigated procurement. All specifications include item, year, and PBU fixed effects. Standard errors, in parentheses, are clustered by PBU. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The second diagnostic summarizes the dynamic design around first item-level exposure to litigation. The BJS estimates rise after exposure, but the Honest-DiD sensitivity calculation does not survive deviations at the observed maximum pre-period scale. The event-study evidence is therefore used as timing evidence, not as the primary causal design.

Table C.2: Dynamic-design sensitivity summary.

Diagnostic quantity	Value	Interpretation
BJS coefficient at first post period	0.052	First post-exposure timing estimate
Standard error	0.018	BJS period-specific standard error
Maximum absolute pre-period coefficient	0.041	Observed pre-period deviation scale
Breakdown linear-extrapolation M	0.018	Smallest linear-extrapolation allowance that reaches zero
Survives observed pre-period maximum	no	Diagnostic robustness indicator

Notes: The BJS event study is used to assess timing patterns, not as the primary identifying design. The final row records whether the first post-period estimate remains different from zero when the allowed post-period violation equals the observed maximum absolute pre-period coefficient. The sensitivity calculation uses the saved BJS period-specific standard errors; the off-diagonal covariance is unavailable in the source BJS event-study output.

D. Mechanism Diagnostics

The main mechanism claim combines three pieces of evidence. First, the same-firm pricing test conditions on firm-by-item-by-buyer triples. The one-sided upper 95% bound rules out broad same-firm markups above about 10.8% in this repeated urgent-market sample. At a 10% equivalence margin, the TOST p -value is 0.070; at a 5% margin, it is 0.364. The minimum detectable effect at 80% power is 12.2%, and power against a 10% markup is 0.64. These diagnostics justify the limited wording that the data do not show a broad same-firm markup in deep repeated urgent markets.

Second, quantity is a mechanism, not a control for a preferred direct effect. Administrative urgent orders are larger, and the main figure uses the bulk-discount elasticity to show how scale loss contributes to the administrative-minus-litigated price gap. Third, winner-set evidence shows supplier-set reallocation. Randomly splitting purchases within the same regime gives a baseline Jaccard similarity of 0.378, while the cross-regime Jaccard is 0.268, a gap of 0.109. The cross-regime difference is therefore larger than ordinary turnover within a regime.

The heterogeneity evidence in the main table is interpreted as mechanism evidence. The same-firm margin is weak in deeper repeated urgent markets, while supplier leverage can reappear in thinner or earlier markets. The mechanism claim is therefore conditional: the core cost margin is fragmented sourcing, with same-firm pricing more plausible where demand is small, irregular, or institutionally less routinized.

E. Fiscal Procurement-Cost Implication

The fiscal calculation translates the Lee-bounded sanction margin into a procurement-cost magnitude. It applies the Lee midpoint 18.5% to an admissibility calibration of 50.0% and annual litigated spending of \$300 M, yielding \$27.8 M per year (Lee-bound range \$23.9 M–\$31.7 M). This is not a full social valuation: it excludes health benefits, search costs, compliance costs, and other relevant components.

Table E.1: Fiscal procurement-cost implication of the bounded UTG gap.

Component	Value
Lee midpoint	18.5%
Lee lower bound	15.9%
Lee upper bound	21.1%
Admissibility calibration	50.0%
Annual litigated spending	\$300 M
Annual procurement-cost implication, midpoint	\$27.8 M
Annual procurement-cost implication, lower	\$23.9 M
Annual procurement-cost implication, upper	\$31.7 M

Notes: This calculation is not a full welfare estimate. It applies the bounded price gap to a calibrated admissible share of litigated spending and excludes patient health benefits, search costs, and other welfare components. It should be read together with the main-paper sourcing decomposition.