Racing for Rents: Allocation by Arrival and the Organization of Trade*

Mayara Felix Santiago Fernández-Chacón Jorge Florez Diego Jimenez-Hernandez

September 2026
Preliminary and in progress

Abstract

We study how firms compete for scarce preferential tariff rates allocated first come, first served, using Colombian customs declarations around the U.S.–Colombia trade agreement. Allocation by arrival order turned the corn tariff-rate quota's saving into a 30–95 million dollar annual prize: a regression discontinuity at the exhaustion timestamps shows the tariff jumping ten percentage points for declarations accepted minutes late. Firms reorganized shipping to win this prize—median vessel cargo rose from 5,300 to 13,500 tons—but only where allocation was chronological and binding, not in auctioned or unconstrained products. Consistent with a model of an arrival race under a minimum efficient vessel scale, mid-sized importers pooled cargo on early vessels, the largest shipped alone, small egg-and-poultry producers formed a seven-firm alliance, and coordination outlived the prize. Because arrival order only reallocates a fixed duty-free quantity, resources spent racing are dissipated—advancing cargo 68–81 days costs 17–63 percent of the prize—waste an auction would avoid.

Keywords: tariff-rate quotas, quota administration, rent seeking, coordination.
JEL codes: D72, F13, F14, L91.

*Felix: Yale University and NBER, mayara.felix@yale.edu; Fernández-Chacón: Yale University, santiago.fernandezchacon@yale.edu; Florez: Banco de la República de Colombia, jhflorez@gmail.com; Jimenez-Hernandez: Federal Reserve Bank of Chicago, diego.j.jimenez.h@gmail.com.
First draft: September 2026. All estimates use the universe of Colombian import declarations (DIAN). For useful comments at early-stage presentations of this project we thank Andy Bernard, Lorenzo Caliendo, Rafael Dix-Carneiro, Dave Donaldson, Teresa Fort, Penny Goldberg, Andrew Greenland, Amit Khandelwal, Sam Kortum, Giovanni Maggi, Meredith Startz, Felix Tintelnot, Diana Van Patten, and Daniel Xu, as well as audience members at Yale Trade Day 2023, CESifo 2024, and the Devils and Wolves Conference 2026. For invaluable interviews conducted in 2023—on importers' experience with the U.S.–Colombia TPA, how the TRQ system changed firm logistics, Colombia's poultry industry, and port logistics—we thank Agricol's CEO, FENAVI, and the operations team at the Port of Cartagena. We thank Isabella Caro, Juan Quimbaya-Javela, and Isabella Valencia for exceptional research assistance. Anthropic's Claude Opus 5 and Fable 5.0 were used for streamlining code, iterating on and checking model proofs, and improving clarity and grammar. All errors are our own. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Banco de la República, its Board of Directors, the Federal Reserve Bank of Chicago, or the Federal Reserve System.

1 Introduction

When governments create valuable rights—spectrum, fishing quotas, emission permits, import licenses—the allocation rule determines not only who receives the rent but also what is spent contesting it. In trade policy one such rule operates at the border every working day and has gone almost unstudied: *first-come, first-served* (FCFS) assignment of preferential tariff rates. Under a tariff-rate quota (TRQ), a limited quantity enters at a low tariff and later imports pay a higher rate; FCFS administration awards the low tariff in the order declarations clear customs—no licenses, no discretion—and it is the only method in use that assigns the rent on the basis of *when* cargo arrives.[1] Allocation by arrival opens a margin of response that license systems close: firms can change when, how, and with whom they ship. This paper studies the race this rule sets off—who runs it, alone or together, and at what resource cost—in a setting where the rule is enforced to the second and shipping choices are observable shipment by shipment.

Racing to arrive first is a problem of scale, and importers of different sizes must run the race differently. The setting is Colombia, where the U.S.–Colombia Trade Promotion Agreement (TPA; in force May 2012) placed nineteen sensitive agricultural product groups under TRQs with zero within-quota tariff, seventeen of them first come, first served—among them yellow corn, the largest, whose quota has bound year after year. In-quota treatment goes to the cargo that arrives before the quota fills, so a firm raises the expected value of a shipment by moving the same tons earlier in the year. But bulk grain travels in vessels whose minimum efficient scale exceeds all but the largest importers' cargo. The largest importers can race alone, committing their cargo to early charters of their own. A mid-sized importer cannot fill, or time, a vessel by itself: to arrive early, it must place its cargo on a chartered vessel alongside its competitors' cargo. A first-come, first-served quota under an indivisible shipping technology therefore sets off a race on two margins—the largest firms rushing alone, everyone else coordinating to rush together.

The race should run on the quota calendar. The incentive exists only while in-quota access is scarce: it should appear in years when the quota binds, build as exhaustion approaches, and vanish

[1] Four to eleven percent of the TRQs notified to the WTO are administered first come, first served: 11% in Skully (2001, Table 1) and 4% in the WTO's 2018 count as reported by Beckman et al. (2021). The largest single category in both tabulations—47% and 35% respectively—is "applied tariffs," under which the quota is not enforced and nothing is rationed; among the methods that do ration, licenses on demand are the most common (25% and 22%), followed by FCFS, historical allocation, and auctions. The 2013 Bali Ministerial Decision (WTO, 2013) addresses persistent underfill rather than the choice of method.

once the quota is spent; just before the January reset, when a few weeks' wait restores duty-free access, imports should stop altogether. And it should be absent where the same agreement assigns access by auction—before any ship sails—and in products whose quotas never bind. Racing—alone and together—timed to the quota calendar and specific to arrival-order products is the signature of the rule.

We formalize the race in a partial-equilibrium model of shipping choices under a FCFS quota allocation system. Firms take prices, volumes, and the quota's fill path as given; the only choice is how to ship, and the only thing shipping changes is the *likelihood* that a lot is accepted before the quota exhausts.

The model reduces the race to three key objects, each with a measured counterpart. The first is the *prize per ton*, βr: the scheduled tariff wedge r, times the share β importers retain, is the expected quota rent that arrival order confers on a shipment—identically zero under an auction, where access is assigned before any ship sails. The second is the *dispatch threshold*, V: the scale at which cargo sails on its own schedule, estimated at 16,280 tons from solo bulk charters. The third is the *coordination coefficient*, θ: the per-unit value of pooling, which a binding quota raises above its ordinary-sourcing level by the prize term. The value of coordinating a cargo of size x is $\theta\, x(V - x)$—too little cargo and there is nothing worth organizing, at the threshold no partners are needed, in between firms coordinate—while above the threshold the same prize pulls residual cargo off shared service and into the firm's own early charters.

The model's central prediction is a *sign reversal at vessel scale*: a rising prize makes mid-sized importers pool their cargo on shared vessels and makes the largest importers pull theirs off shared vessels to race alone. The same prize term generates the quota-year cycle described above, and, once forming a shipping relationship carries a sunk cost, predicts that links formed to race the quota persist after the prize is gone. Because the quota admits a fixed quantity, timing gains are zero-sum: the resources spent improving queue position buy no additional access in aggregate. These are signatures that ordinary explanations of bigger, fuller vessels—trade growth, denser service, scale economies—cannot generate.

We test these predictions in the universe of Colombian import declarations. The declarations carry second-level acceptance timestamps—the exact margin on which the quota is allocated—and manifest numbers that tie each firm's cargo to a physical vessel, so we observe who shares a vessel

with whom. Because the agreement runs first-come, first-served and auctioned quotas over the same period and ports, and because some quotas bind while others never do, the allocation rule and the prize vary while everything else is held fixed (Appendix Table A2). The rent assigned by arrival order in yellow corn is 30–95 million dollars a year, in a market whose entire import bill is a small multiple of that.

The first step is to establish, at the timestamp level, that the saving is in fact assigned by arrival order. Accumulating declarations to the scheduled quota reproduces the exhaustion timestamps implied by recorded tariff treatment—to the second in three of the six crossing years, within sixteen minutes in two more, and seven days apart only in 2014, the first quota year, where carryover accounting intervenes. A regression discontinuity at those timestamps shows the tariff jumping ten percentage points, on average, for filings accepted minutes late, with the prices of the goods themselves smooth through the cutoff. Arrivals exceed the quota by 19–75% in every binding year, so the rent is strictly positive. To our knowledge this is the first shipment-level verification that a major trade preference is allocated by a race—and therefore the first study of the race itself.

When the quota binds, the race reorganizes who ships with whom. Three patterns emerge at the level where the race operates—pairs of importing firms on vessels. Pairs whose combined cargo fills one vessel increase their co-shipping, and the increase reverses on either side of vessel scale. The largest importers avoid shared vessels in the opening months, when their own early charters secure their position. And at every scale, co-shipping recomposes through a common exporter, a margin that survives holding the exporter composition at its pre-agreement distribution.

Within the year, the race is organized around the actual exhaustion dates rather than the calendar. Exact inference attributes much of the net within-year movement to when, and at what scale, vessels sail. The within-year evidence therefore rests on the same-exporter margin—elevated in every phase of binding years and in neither slack year—and on the cross-regime contrasts. A market-wide doubling of vessel scale accompanied these responses, and appears only where allocation is chronological and the quota binds—not under auction, slack quotas, or no quota.

The sign reversal appears where the model puts it. In race windows relative to post-exhaustion windows of the same year, importing groups just below the vessel threshold cut their shared-vessel participation by nine percentage points and charter alone instead. Against a randomization benchmark that reassigns the same cargo to the same vessels, pairs involving an above-threshold

firm co-occur 14 percent *less* than chance in race windows—and at chance in the same product's no-race windows—while mid-sized pairs co-occur modestly more, an excess that disappears after exhaustion. Firms leaving the cheapest per-ton shipping technology exactly when arrival order carries a prize, with solo charters sailing below capacity most common in race windows, is difficult to reconcile with economies of scale, which push the largest firms *toward* shared bulk service. A timing motive explains leaving it. The net increase in co-shipping is concentrated where the theory places it—pairs near vessel scale—and is correspondingly modest averaged over the whole year: the race moves the firms the model says it should move, and no others.

Three further results tie the reorganization to the race. A bloc detector that sees only joint-booking statistics—no names, no sectors—recovers all seven members of an egg-and-poultry buying alliance that one member independently described in an interview, an external validation of the detection; the alliance is lawful under Colombian competition law, organized to secure access before larger firms exhausted the quota.

The race left a permanent mark on who ships with whom. New co-shipping relationships form at several times their usual rate when the quotas begin to bind, and existing ones persist through 2023, after the quota became unlimited, in line with evidence on the durability and value of trading relationships (Macchiavello and Morjaria, 2015). And imports stop before each January reset, most completely in the highest-wedge products: crude soybean oil, which used to land a third of its volume in November and December, essentially stops importing at year-end once its quota binds. Sections 5.3–5.5 present these results.

The race is costly, and the accounting needs no functional form, because the duty-free quantity and the tariff wedge are fixed by the agreement's schedule: racing changes when the quota is claimed, by whom, and at what resource cost, never the size of the transfer. In-quota cargo arrives at a seasonal center of gravity 68–81 days earlier than before the agreement, and at conventional carrying costs that advancement alone costs 9–34 million dollars a year—17–63 percent of the mean annual prize—while measured deadfreight adds little. Two-thirds of the retained rent accrues to just six importing groups.

Coordination itself can save real resources, and we measure the sourcing economies that do: small solo lots pay a landed-cost penalty of nearly a fifth that disappears on shared service. Prize-induced racing saves none: it spends resources to reallocate a fixed quantity of access. An auction of the same

quota would deliver the same scarcity with none of the race, and the comparison is not hypothetical: Colombia runs both systems side by side, and the auctioned products show none of these margins.

Overall, the analysis offers three take-aways that travel beyond corn: (i) the administration rule is not an implementation detail—two quotas of identical size, one auctioned and one raced, produce different coordination among importers and different resource costs, margins that tariff-equivalent accounting cannot detect; (ii) allocation by waiting acquires new economics when a queue position can be improved only by cooperating with competitors—the contest reorganizes who cooperates with whom rather than only burning resources; and (iii) sunk relationship costs convert transient administrative rules into permanent institutions—the coalitions formed to race the corn quota were still booking jointly a year after the quota ceased to exist.

Our first contribution is to the economics of quantitative restrictions, which has concentrated on who holds the right rather than on how the right is assigned. The assignment of quota rights governs the incidence of the rents (Anderson and Neary, 1992), quotas change how firms compete (Krishna, 1989), and discretionary licensing misallocates them (Khandelwal et al., 2013); the administration rule itself is treated as a detail of implementation, and this paper shows it is not one.

We also contribute to the economics of rent seeking and rationing by waiting, where the contest input has always been private. That literature prices contests for policy-created rents (Krueger, 1974; Posner, 1975; Tullock, 1980), with queue costs dissipating rents when allocation is by waiting (Barzel, 1974; Deacon and Sonstelie, 1985), and we borrow from it the term *prize*. But in all of these each agent queues, or spends, alone, whereas here a firm can improve its queue position only by coordinating with competitors—so the contest reorganizes who cooperates with whom.

Finally, we bring shipment timing and vessel sharing into the literature on lumpy trade and the organization of shipping. That literature explains shipment size and frequency under fixed shipment costs, delivery lags, and inventories (Alessandria et al., 2010; Kropf and Sauré, 2014; Hornok and Koren, 2015; Coşar and Demir, 2018) and has begun to endogenize the transport technology itself (Brancaccio et al., 2020), but it has not studied shipment timing as a policy-created asset, nor who shares a vessel with whom—largely for want of data linking cargo to voyages.

The coordination we document has a counterpart outside trade: Bassi et al. (2022) show that small Ugandan manufacturers reach the scale of an indivisible machine by renting its capacity from one another—achieving scale collectively. Vessel pooling is the same arrangement on an

international margin, with one difference that matters for welfare: there the shared scale produces output, while under a FCFS quota it buys position in a queue for a fixed rent.

Two limitations bound what the tests establish. The pairwise design is powered only for yellow corn among the binding products, and purchase contracts are not observed, so bargaining-motivated pooling is bounded indirectly, by its timing and by the price schedule (Section 5.2). Appendix C collects the limitations and the data decisions a careful reader should know.

The paper proceeds as follows. Section 2 describes the agreement and the data and shows that the tariff saving is assigned by arrival order. Section 3 states the model and its three main results; the full theory and all proofs are in Appendix A. Section 4 translates it into an empirical design. Section 5 documents the reorganization of shipping, reports the tests and the cost moments that discipline the value function, and measures the incidence of the saving. Section 6 presents the cost accounting, and Section 7 concludes.

2 Setting, data and preliminary evidence

2.1 The U.S.–Colombia Trade Promotion Agreement

The U.S.–Colombia Trade Promotion Agreement (TPA) was signed on November 22, 2006, and—after a five-year ratification delay in the U.S. Congress—entered into force on May 15, 2012, so its schedules were public years before they operated. The agreement is comprehensive. On the import side studied here, Colombia granted immediate duty-free access to 81.8 percent of U.S. products upon entry into force—most of it raw materials and capital goods not produced domestically—with the remaining tariffs phasing out on product-specific schedules; virtually all of Colombia's industrial exports obtained immediate access to the U.S. market in return; and chapters on services, investment, intellectual property, government procurement, competition, labor, and the environment complete the treaty.

This paper concerns the agreement's treatment of sensitive agricultural products. For nineteen product groups the TPA created tariff-rate quotas: a scheduled duty-free quantity $\overline{Q}_{hy}$ for product h in year y, growing at a contracted rate, with out-of-quota imports taxed at a tariff that declines linearly to zero over a phase-out horizon. For yellow corn—the leading case throughout, because it

is transported in bulk, its quota bound repeatedly, and its shipments carry economically meaningful multi-importer structure—the initial quota was 2.1 million tons, growing 5% per year, with an out-of-quota tariff of 25%—the average of Colombia's price-band tariffs over 2001–2004—phasing to zero in 2023; the white corn quota was 130,000 tons on the same growth rate. In world-price terms the scheduled wedge was 40 USD per ton in 2014, declining to 11 in 2021 and zero from 2023 on.

What matters for this paper is not the quota but its *administration*. Colombia allocates the duty-free quantity in the chronological order in which eligible import declarations (Form 500) are presented to and accepted by the customs authority, DIAN; anticipated declarations do not qualify. No license is required and no quantity is reserved: the right to the low tariff attaches to a shipment's position in the acceptance queue in the year of arrival.

Other TPA products are administered differently. Rice and chicken leg quarters are allocated by *auction* through export trading companies: access is scarce and valuable, but it is assigned at least three months before shipment, so arrival order is irrelevant. Still other covered products have quotas that never bind, and close substitutes in the bulk-grain trade (soybean meal, soybeans, wheat) have no quota at all. The same agreement, period, and ports under different allocation rules is the variation that isolates the rule itself. Before the TPA, Colombian corn imports came mostly from Mercosur under the *Andean Price Band System*; the agreement shifted origin decisively to the United States, a confound the design must and does address (Sections 4 and 5.1).

2.2 The data

The data are the universe of Colombian import declarations (DIAN) for 2004–2023, covering yellow corn and every comparison product under the same agreement; a declaration is one importer's customs filing on one shipment, recording the importer, the cargo manifest, the exporter of record, product, origin, port, tonnage and values, the tariff paid, and—for 2013–2021, to the second—the timestamp at which DIAN accepted the filing. The manifest number ties filings to physical vessels (in a purchased sample of AIS ship-tracking voyages matched to the declarations, each physical voyage corresponds to exactly one manifest), so we observe *who shares a vessel with whom*—the outcome coordination operates on, measured at the *lot* level, one ownership group's cargo on one

manifest. Acceptance timestamps are the exact margin on which the quota is allocated, so we observe the race at the resolution at which it is run. And coverage of every product and origin under one agreement supplies the cross-regime contrasts—auctioned, slack, no-quota, non-U.S.—on which identification rests. Importers are aggregated into ownership groups; firm sizes are fixed from pre-TPA (2004–2010) data alone, so no size cut can respond to the treatment it is used to study; and the dispatch threshold at which a cargo can sail on its own schedule is estimated at V = 16,280 tons (Section 4.1). The clean yellow-corn panel has 42,574 declaration lines on 6,979 manifests; after the ownership-group match, the lot panel the coordination tests use has 17,891 lots on 5,990 manifests. Sources and vintages, construction, harmonization, the arrival-date hierarchy, and validation statistics are in Appendix B.

2.3 The tariff saving is assigned by arrival order

We next establish two facts: in-quota treatment is valuable, and it is assigned by acceptance order.

Table A3 reports the *cutoff ledger*. For each quota year we order all U.S. yellow-corn declarations by their second-level acceptance timestamps and accumulate net weight until the scheduled quota is reached; the timestamp of the crossing declaration is the reconstructed exhaustion time. We compare this tonnage-based reconstruction with a second, independent construction: the *treatment-based* timestamp, the acceptance time of the last declaration that received in-quota tariff treatment, computed from the recorded tariff on each declaration. DIAN publishes no exhaustion time, so agreement between the two independent constructions—one using only cumulative tonnage, the other only recorded tariff treatment—is the validation. In 2015, 2017, and 2019 the two match *exactly*; in 2016 they differ by one second and in 2018 by sixteen minutes. The only material discrepancy, in 2014, has the tonnage-based crossing seven days ahead of the treatment-based one, consistent with in-quota carryover accounting in the first full quota year; we use the treatment-based timestamps, which encode the allocation DIAN actually applied, wherever available.

Fill ratios—annual U.S. arrivals over the scheduled quota—are 1.19–1.75 in every crossing year, so the out-of-quota margin is always active: the marginal ton pays the high tariff, and the rent per in-quota ton equals the statutory wedge. This over-filled regime is also what keeps demand and supply elasticities out of the rent. With out-of-quota imports flowing, arbitrage pins the domestic

price at the marginal import cost $\bar{p}_y(1+\tau_y)$: in-quota and out-of-quota corn sell at the same domestic price, so the value of in-quota treatment on a ton the firm imports anyway is exactly the tariff it did not pay. Elasticities determine the *quantities*—how much corn crosses at the high rate, and hence the fill ratio—but not the per-ton rent; they would price the rent only in the intermediate TRQ regime where the quota binds without out-of-quota imports and the domestic price settles between the two tariff-inclusive costs (Skully, 2001), a regime never observed here in a crossing year. Two maintained assumptions complete the argument: Colombia's imports are small relative to the world corn market, so the world price $\bar{p}_y$ does not respond to the quota (the transaction-price evidence of Section 5.7 supports this—exporters capture at most a modest share of the wedge), and annual import volumes are taken as given, so the rent is evaluated on inframarginal tons. In 2013 the quota never crossed (fill ratio 0.23); the race regime correctly begins in 2014.

2.4 Enforcement at the timestamp: a regression discontinuity

Figure 1 and Table 1 show that the tariff a declaration pays is determined by its position in the acceptance queue, declaration by declaration. The running variable is the (fractional) number of days between a declaration's acceptance timestamp and its year's treatment-based exhaustion timestamp T_y^*; the sample is all TPA yellow-corn declarations within 210 days of the cutoff in the crossing years 2014–2019, with local-linear estimation, importer and year fixed effects, clustering by importer, and a donut around the thin carryover band at the crossing. Declarations accepted just after T_y^* pay the out-of-quota tariff: the effective tariff proportion jumps by 0.695 (s.e. 0.035) from essentially zero, and the tariff rate itself by 10.0 percentage points (s.e. 0.64)—two filings minutes apart face tariffs ten points apart, with no license, no appeal, and no discretion at the margin.

An allocation this sharp closes every avenue of influence *at* the border, so the contest for the prize must be fought *earlier*, in chartering, pooling, and filing decisions made weeks to months in advance. That displacement of competition from the cutoff into the organization of shipping is what the rest of the paper measures.

Two companion facts complete the picture. FOB unit values are smooth through the cutoff (-0.001 log points, s.e. 0.011)—the jump is in the tariff, not in the goods, a first local piece of the incidence result—and at the January reset the mirror image obtains: in-quota treatment is fully

restored (the effective tariff proportion falls by 0.997) as bulk racing lots return, so each year is a self-contained race and the reset a magnet for deferred cargo (Section 5.5). Quantity and price movements *at* the cutoff are treated throughout as description, because sorting into the queue is the phenomenon under study; the remaining discontinuities (declaration size and its selection reading), the reset RD, and robustness are in Appendix B.1. First come, first served describes this quota's administration exactly, to the second.

Two features of this validation matter for what follows. It justifies treating the within-year exhaustion time T_y^* as data: the *regime calendar* of Section 4 partitions every product-year at frozen, externally validated timestamps rather than at estimated breaks. And the stakes are large: multiplying scheduled wedges by scheduled quotas, the rent assigned by arrival order was roughly 94 million USD in 2014, declining with the phase-down to about 35 million in 2021—an annual prize competed for by a few dozen importing groups.

3 A model of coordination under first-come, first-serve quotas

This section introduces a partial-equilibrium model of shipping choices that provides the relationship between the prize and the organization of trade. Prices, aggregate volumes, and the exhaustion date T^* are taken as given, at the values the data record;[2] the model solves for the object the empirics measure: which firms coordinate their shipping, with whom, and when in the quota year. The discussion is organized around three key objects: the prize per ton βr, the dispatch threshold V, and the coordination coefficient θ—and three results that trace their implications across the size distribution, across the quota year, and over time. The full development—the environment in detail, the remaining results, and all proofs—is in Appendix A.

Environment. In a quota year, tariff t_L applies to the first $\overline{Q}$ tons accepted and $t_H > t_L$ thereafter; with out-of-quota imports in every binding year, the domestic price is pinned by the high-tariff

[2]Taking prices as given abstracts from bilateral price formation between importers and exporters. Alviarez et al. (2026) model firm-to-firm trade as a bargaining process in which oligopoly and oligopsony power set relationship-specific markups and pass-through, and Alviarez et al. (2023) study the resulting link between two-sided concentration and markups in the same Colombian import records we use. In that framework a pooled purchase could also move the negotiated price; we hold prices fixed to isolate the race among importers for a rent whose size is set by the tariff schedule, measure the split of that rent from transaction prices (Section 5.7), and bound bargaining-motivated pooling directly (Section 5.2).

marginal import cost, so the value of in-quota treatment is the full wedge $r = t_H - t_L$ per ton, free of demand or supply elasticities (Section 2.3); the importer retains a share $\beta \in (0, 1]$ of it (measured, not modeled). Because the world corn price and downstream prices are taken as given, the race operates through a single channel: shipping choices change the *probability* that a lot is accepted before the exhaustion date. Advancing a shipment by one unit of time raises that probability by ρ, the reciprocal of shipping-time dispersion—when sailing is predictable, a one-day advance is very likely to change the allocation, and the race is steep (Appendix Proposition A1).

The shipping technology has one kink, and the analysis builds on it. Bulk carriage becomes worthwhile only once a committed load reaches the dispatch threshold V: cargo of size $x < V$ using regular consolidation waits for the market to assemble the missing tonnage, delaying departure by $\eta(V - x)$, with η the assembly time per ton, while an organized shared vessel that commits a dispatch load departs immediately. Pooling therefore buys time, at carrying cost c per ton-day on the advanced cargo, and it also carries ordinary sourcing economies $s \geq 0$—freight and procurement savings from shared bulk service; s also absorbs any bargaining transfer a pooled purchase extracts from the exporter, a component indistinguishable from real economies in the participation decision, and one Section 5.2 bounds directly.

Organizing costs each member F in its first season, of which a relationship-formation component ζ is sunk, leaving a smaller continuation cost $F^m = F - \zeta$. An importer's gain from coordinating a lot of size $x \in (0, V]$ is

$$g(x) = \theta\, x(V - x) - F, \quad \text{with} \quad \theta \equiv s + \eta(\beta r \rho - c). \tag{1}$$

Equation (1) encapsulates the model's economics: the value of pooling scales with the cargo a firm has to move, x, times the scale it is missing, $V - x$, at a rate that adds the sourcing economy to the net value of the time pooling buys. Because the quota admits a fixed quantity, the timing gains the prize term buys are zero-sum across firms: one lot's improved position displaces another's one for one, so the resources spent producing position buy no additional access in aggregate (Appendix Lemma A3)—the accounting Section 6 rests on.

Key objects. Three objects carry the model to the data, and each has a measured counterpart. The first is the *prize per ton*, βr: the scheduled wedge r is written into the agreement (10–40 USD/t; Section 2.3) and the retained share β is measured from transaction prices (0.71–0.96; Section 5.7). The second is the *dispatch threshold* V, estimated from post-TPA solo bulk charters at 16,280 t (Section 4.1); it fixes where every firm and every pair stands in the size distribution, x/V and $(x_a + x_b)/V$. The third is the *coordination coefficient* θ of equation (1), the single sufficient statistic through which the prize reaches behavior: in no-prize states—$r = 0$ or a slack quota—it is $\theta_0 = s - c\eta$; a binding FCFS quota raises it to $\theta_1 = \theta_0 + \beta r \rho \eta$. These two values are the contrast that drives every empirical finding: each design in Sections 4–5 compares a binding-FCFS state against a no-prize state—race against post-exhaustion windows within a year, binding against slack years, FCFS against auctioned and no-quota products—and the timing primitives ρ and η never need to be separated, because behavior depends on them only through the product $\beta r \rho \eta$ (Section 5.7). The participation cutoffs (x_-, x_+) of Proposition 1 below are functions of (θ, V, F); they are the objects the co-shipping tests read off the data.

The model's first result concerns which firms coordinate, and in what configurations.

Proposition 1 (Coordination and unbundling: the sign reversal at vessel scale)**.** *Let delays be uniform and* $\theta V^2 > 4F$. (i) *A lot coordinates if and only if its size lies in the interval* (x_-, x_+),

$$x_\pm = \frac{V \pm \sqrt{V^2 - 4F/\theta}}{2}, \tag{2}$$

with $0 < x_- < V/2 < x_+ < V$*; a firm with* $x = V$ *departs alone. The same cutoffs apply to a coalition that pools its cargo into one lot* $X_B = \sum_{i \in B} x_i$ *at joint cost* F_B*. For a pair, the prediction therefore attaches to the combined size* $x_a + x_b$*: a pair pools when its combined cargo lies in* (x_-, x_+)*, and not when it jointly falls far short of* x_- *or overshoots* V. (ii) *A firm with* $x = nV + \xi$*,* $n \geq 1$*, ships its whole loads on its own early charters; its only margin is the residual* ξ*. The firm* internalizes *the residual—stretches it across its own early charters at deadfreight cost* $\phi(\xi)$*—rather than consolidating it cheaply but late if and only if*

$$(\beta r \rho - c)\, \eta\, \xi (V - \xi) \;>\; \phi(\xi) + s^c\, \xi (V - \xi), \tag{3}$$

where $s^c \geq 0$ is the economy consolidation passes to the residual. The internalized set is empty at $r = 0$ and expands in r, β, and ρ. A rising prize therefore raises co-loading for sizes in (x_-, x_+) and, where internalization is the residual's preferred mode, lowers *it above the dispatch threshold.*

The reversal is the model's central observable, and it is what separates the timing motive from every ordinary explanation of bigger, fuller vessels: trade growth, denser service, and bulk discounts push co-loading in the same direction at all sizes, while a prize on arrival order moves the two sides of the size distribution in opposite directions. Part (i) is why the co-shipping tests cut pairs by combined size $(x_a+x_b)/V$; part (ii) is why they watch the largest firms leave shared vessels exactly when the prize turns on. A residual can also be organized like any lot of size ξ—part (i) applied to ξ—and Proposition 2 confronts that channel with its alternative. The interval does not depend on uniform delays or on the quadratic form—Appendix Proposition A11 derives it under a general delay distribution—and the sourcing term's single-peaked shape is not assumed but disciplined by cost moments measured in no-prize states (Section 5.6).

The second result concerns when in the quota year they coordinate.

Proposition 2 (The coordination cycle)**.** *Let $r > 0$, let $T^* < 1$ in the normalized year, let waiting for the January reset cost $c_w > 0$ per ton per unit time, and let firms with $n_i \geq 1$ commit charters at $t = 0$. The year partitions into four phases. (i)* Opening*: little cargo is at risk of missing the quota; pooling among sub-threshold pairs is close to its ordinary level, and above-threshold firms—whose early charters are their race position—avoid shared service the most. (ii)* Stockpiling *and (iii)* depletion*: the at-risk mass rises, and with it the return to pooling for pairs whose combined cargo fills a vessel; the coordination margin peaks in depletion, when a vessel's arrival order is most likely to determine tariff treatment. (iv)* Post-exhaustion*: the prize is spent and coordination reverts to its ordinary-sourcing level θ_0. In the terminal* dead zone *$(t_w, 1]$, $t_w = \max\{T^*, 1 - r p_E^+/c_w\}$ with p_E^+ next season's early-access probability, no shipments arrive and all demand postpones to the reset; the dead-zone width increases in the tariff wedge and is zero at $r = 0$.*

One margin the proposition deliberately leaves unsigned: for above-threshold firms, two channels can raise mid-season co-shipping. The residual margin is Proposition 1(i) applied to the residual lot ξ, and its value is graded in $\xi(V - \xi)$; the platform channel of Appendix Proposition A6 concentrates purchases on vessel-capable traders and carries no ξ-gradient. The ξ-gradient is how Section 5.2

tells them apart.

The proposition dictates the empirical calendar twice over. Every within-year test must partition product-years at the exhaustion timestamps, because pooling regimes attenuate every coordination margin toward windows where the model predicts nothing. And the phase profile itself is a prediction: net pooling among dispatch-scale pairs should build with the fill state and peak in depletion, above-threshold avoidance should be front-loaded, and both should vanish after exhaustion. The dead-zone width carries a dose-response in the predetermined scheduled wedge—identifying variation immune to the origin-switch confound.

The third result concerns whether the coordination outlives the prize.

Proposition 3 (Persistence: formation responds to the prize, dissolution does not)**.** *Let the formation cost have a sunk component* $\zeta = F - F^m > 0$ *and let the state* θ_t *be drawn each season. Then the set of states in which an existing group persists strictly contains the set in which a new group forms: there is a band of states in which no new group forms but every existing group continues. If the prize falls permanently to zero, an incumbent persists whenever* $(s - c\eta)x(V - x) \geq F^m$*, while entry requires* $(s - c\eta)x(V - x) \geq F - \delta\zeta$*; the band of inaction has width* $(1 - \delta)\zeta > 0$.

Three measurement implications follow. The *flow* of new co-shipping links, not the *stock* of continuing ones, is the prize-responsive outcome; the model does not predict that participation falls promptly when a quota becomes unlimited—it predicts formation reverts to its ordinary level while links persist; and the asymmetry is itself testable—a symmetric response of formation and dissolution would *reject* the sunk-cost structure. The 2023 transition of the corn quota to unlimited access provides the direct test.

The rest of the theory. Appendix A develops the environment in full and states the remaining results, each tied to a test in Sections 4–5: the probability channel, the value of scarcity, and the zero-sum lemma (Proposition A1, Corollary A2, Lemma A3); the interval under general delay technologies and the comparative statics (Propositions A4 and A11, Corollary A5); the exporter as coordination platform—why prize-induced pooling recomposes co-shipping through common exporters (Proposition A6); partner rotation on the anonymous consolidated segment against standing blocs, and why stable pairs identify blocs alone (Propositions A7–A8 and Corollary A9); and the

randomization benchmark that makes "more than chance" an exact statement (Proposition A10). Appendix Table A1 collects the model's predictions and where each is tested.

4 Empirical design

Four principles organize the estimation. First, *regime partition*: every within-year object respects the calendar of Proposition 2. Second, *benchmark before inference*: co-occurrence is evidence of coordination only relative to an explicit null for how often the same cargo would share vessels anyway (Appendix Proposition A10).

Third, *no pooling across products*: products differ in vessel technology and quota administration, so every specification is estimated separately by product and labeled with its allocation regime; cross-product comparisons are made across separately estimated results, never inside one estimation. Fourth, *predetermined treatment*: dose-response designs use only the scheduled wedge and quota path fixed by the agreement, never realized exhaustion.

4.1 Measurement

Each object the tests use is the measured counterpart of one of the model's key objects, constructed once and read identically by every specification—so no empirical choice can drift from test to test. The *lot*—one ownership group's cargo on one manifest—is the unit on which coordination operates, the model's cargo x on a vessel: tons summed over all of the group's declaration lines on the manifest regardless of acceptance date, with modal exporter, tonnage-dominant origin, and the manifest's arrival date (manifest date where available, else earliest non-anticipated acceptance; Appendix B) defining its year and 14-day arrival window. The *technology* is the dispatch threshold and where each firm stands relative to it: $V = 16{,}280$ t—the median size of a vessel a single firm charters on its own account once the race is on (post-TPA solo bulk manifests; alternatives that bracket it are in Appendix B)—with predetermined sizes binned in x/V (0–2%, 2–6%, 6–15%, 15–35%, 35–70%, >70%). Because predetermined size is a *monthly mean* while V is a per-vessel scale, monthly size understates per-shipment cargo; the empirical counterpart of the model's above-threshold firms is therefore the top two bins ($x/V \geq 0.35$)—a group at $0.35V$ per month accumulates a dispatch load within a quarter—with $[0.35, 0.70)$ described as "just below vessel scale" and >0.70 as "at vessel

scale" throughout.

The *calendar* is the prize path the model takes as given: the frozen exhaustion timestamps of Table A3 partition each product-year into the race window [Jan 1, T_y^*], post-exhaustion (T_y^*, $t_{w,y}$], and dead zone ($t_{w,y}$, Dec 31], with $t_{w,y}$ the start of the terminal arrival gap; slack years, 2023, auctioned, and no-quota products are entirely "no race," and a ±7-day donut around T_y^* applies in any regression using a window assignment, because the crossing is partly behavior-determined.

The *dose* is the expected prize of Appendix Corollary A2, built from policy variables alone: $\widehat{\mathcal{P}}_{hy} = \Delta t_{hy} \times \widehat{\Pr}_0\{\text{exhaustion within season} \mid \overline{Q}_{hy}, \text{pre-TPA seasonality}\}$, with the scheduled wedge Δt_{hy}—fixed by the agreement—serving as the dose within the post period.

4.2 The randomization benchmark

Appendix Proposition A10 states the model's null in full. Fix a cell—product, origin, arrival port, non-overlapping arrival window—and condition on the lots present with their sizes and on the realized manifest opportunity set. If there is no timing prize, no organized shared service, and consolidation queueing is exchangeable across identities and sizes, then randomly reallocating lots to feasible manifests generates the *exact* null distribution of any assignment statistic, including pairwise co-occurrence; with identity-only exchangeability, exactness holds conditional on the realized size-to-manifest allocation. Under a positive prize the equilibrium departs from this null in the directions of Proposition 1: excess co-occurrence of lots whose sizes lie in the coordination interval, concentrated before T^*; depressed co-occurrence involving the residuals of above-threshold firms; and both departures growing with the expected prize and vanishing in no-prize states.

Three consequences follow. The conditioning set—origin, port, window, lot sizes, realized capacities—is what defines which cargo *could* have shared a voyage; the benchmark must *not* additionally condition on sector, partner identity, or manifest composition, which are the outcomes under test. The statistic to report is the observed-minus-simulated deviation *by size cell and regime*, not a single co-loading rate. And because ordinary organized sharing also violates exchangeability, a deviation alone does not implicate FCFS: attribution requires the regime and origin contrasts—race versus post-exhaustion within year, U.S. versus non-U.S. within product, FCFS versus auction versus no quota—applied to the deviation. In practice we implement two nulls: a *slot* permutation (manifest

labels permuted among lots within cell) and a stricter *capacity-feasible* null, whose comparison is itself informative (Section 5.2).

How the benchmark is constructed. Both nulls sample reallocations of cargo to vessels that respect the physical opportunity set, and differ only in what they hold fixed. The slot null permutes manifest labels among the lots in a cell, preserving each manifest's number of lots but not its tonnage. The capacity-feasible null—the one the co-shipping difference-in-differences uses—is stricter: it draws, uniformly, from the set of lot-to-manifest assignments in which every manifest's total tonnage stays within a tolerance of what it actually carried, so a reassignment counts only if the resulting vessel loads are physically feasible. That feasible set cannot be enumerated, so we sample it uniformly with a swap-MCMC; each draw is a complete counterfactual assignment of cargo to vessels, on which the co-occurrence statistic is recomputed exactly as in the data. Appendix D gives the cell definition, tolerance, proposal, and mixing diagnostics.

4.3 Outcomes and inference

The outcome families, each tied to Table A1 are: participation and composition (multi-group share, solo-full-vessel share, groups per manifest, vessel scale); benchmark deviations by pair category and regime; link formation, re-linking, and dissolution flows among the top-50 pre-TPA groups; bloc detection among sub-threshold groups (universe including entrants via the post-entry proxy: below the participation cutoff, organized early service is available only through a bloc, Appendix Proposition A8 (ii), so excluding entrants would exclude exactly the firms blocs recruit); dead-zone and reset outcomes by product-year; cost-shape moments constructed only in no-prize states with unit values residualized on importer–exporter, product×origin×month, and customs-office effects and the world price; and transaction-price gaps for incidence.

Inference: randomization p-values for all benchmark statistics; group-clustered standard errors and firm-level bootstrap for dyadic outcomes; wild-cluster or randomization inference where clusters are few; counts reported wherever denominators are thin. All randomization uses fixed seeds, and two independent end-to-end re-estimations reproduce every table.

5 The race in the data

This section first documents the market-wide reorganization that accompanied the race, then reports the tests: the within-year contrast in vessel sharing and the benchmark deviations, the pairwise co-shipping design and its exporter decomposition, the bloc detector and its validation, link formation and dissolution, and the quota cycle, closing with the cost moments, the incidence of the saving, and the limitations.

5.1 The reorganization of shipping

The race did not operate on a static market. Between the pre-agreement period and the quota years, yellow-corn shipping changed scale—the whole distribution of manifest tonnage shifts from boat scale to vessel scale (Figure 2)— the median U.S.-origin manifest more than doubled within two years of the first binding year and vessels came to carry more importing groups each (Appendix Figures A1 and A3), and by the post-TPA period the median manifest carried roughly 130 times the median importing group's weekly volume—no participant could fill, or time, a vessel alone.

The change concentrates where allocation is chronological and the quota binds (Appendix Table A5). White corn, the other binding FCFS grain, scales up. Auctioned rice stays at boat scale, and the never-binding animal feeds do not scale. No-quota soybean meal carries the same origin switch and comparable volume growth, yet scales far less. The firms doing the racing are, for the most part, the final users of the grain: feed producers, poultry and egg companies, and processors account for eighteen of the twenty largest importers, and the top ten ownership groups for 85% of volume (Appendix Table A6). The sellers they face are the world grain traders. Seventeen of the twenty largest exporter entities on U.S.-origin shipments are merchant traders—Seaboard (23% of U.S.-origin volume), Bunge (20% across entities), Cargill (16%, mostly through CAI Trading), and ADM (11%)—the only producer-owned entities are the cooperatives Zen-Noh and CHS, and no U.S. farming operation appears at all (Appendix Table A7). Cargill sits on both sides after acquiring the poultry importers Pollos El Bucanero (2017) and Campollo (2018). Two features of this structure matter later. The race is a contest *among concentrated direct consumers*, fought with vessels that a handful of traders charter and fill—which is why the exporter emerges below as the platform on which coordination is organized (Appendix Proposition A6, T7). And concentration on both

sides is the environment in which bilateral bargaining sets prices (Alviarez et al., 2023, 2026), the reading Section 5.7 adopts when it measures how the tariff saving is split. Appendix Figure A1 and Appendix Tables A5, A6, and A7 document these facts; because they concern market-level volumes, which the model of this paper takes as given.

The remainder of this section asks what share of this reorganization is coordination induced by the race, using designs that hold the year, the product, and the physical opportunity set fixed.

5.2 Non-linearity of coordination incentives at vessel scale

Raw regime contrast (T2). Within crossing years, we compare each group's shipping organization in race windows to post-exhaustion windows of the same year:

$$y_{gwt} = \alpha_g + \lambda_t + \sum_b \beta_b \, \text{Race}_{wt} \times \mathbf{1}\{b(g) = b\} + \varepsilon_{gwt}, \tag{4}$$

where g indexes ownership groups, w arrival windows within year t, α_g and λ_t are group and year fixed effects, Race_{wt} marks race windows, and $b(g)$ is g's pre-TPA size bin; standard errors are clustered by group. The coefficients β_b trace the race-window change in shipping organization across the size distribution (Figure 3, Table 2).

The above-threshold side is where Proposition 1(ii) puts it: in race windows, groups just below vessel scale withdraw from shared vessels—their multi-group share falls by 8.9 percentage points (-0.089, $p = .002$) and charter alone instead (solo full-vessel share up 0.6 points, 0.006, $p = .095$); the handful of groups already at vessel scale, with the least to gain from unbundling on the model's own logic, do not move at all (-0.0004, $p = .99$).

The sub-threshold side is positive but imprecise in raw multi-share (the 6–15% bin is +4.3 points, $p = .19$), for a mechanical reason: with 60% of post-TPA manifests already shared, binary co-loading is close to saturated, and mid-firm pooling appears sharply only against the benchmark. The benchmark deviations below bear this out.

The estimate survives moving the threshold 20% up (-0.093, $p < .001$) and attenuates when moved 20% down (-0.063, $p = .34$), where the bin absorbs groups that jointly cannot fill a vessel—robust to where vessel scale is placed, not to erasing the distinction—and a Benjamini–Hochberg adjustment across all twelve coefficients in Table 2 leaves the $[0.35, 0.70)$ multi-share effect as the

sole discovery at $q = .05$—where the theory places its central prediction (Appendix F).

Benchmark deviations (T3). The second test asks whether the co-occurrence patterns exceed what the physical opportunity set alone would produce. Table A9 and Appendix Figure A2 report observed minus simulated co-occurring pairs under the two nulls of Appendix Proposition A10, pooled over 2014–2021 and split by regime. In race windows under the slot null, the two sides of the size distribution move in opposite directions, as the model requires: pairs involving an above-threshold group co-occur 14 percent *less* than chance (−137 pairs on a simulated base of 989, significant in seven of eight race years), while both-mid pairs co-occur *more* than chance (an excess of 24 pairs, significant in three race years and never significantly negative). The top deficit shrinks after exhaustion (−84.4) and the mid excess disappears (−4.7; the raw within-year exhaustion split is Appendix Figure A9).

Under the capacity-feasible null the top deficit compresses to −21.5: most of the top-firm avoidance is a *scale* choice—chartering whole vessels of a size no tonnage-preserving swap can replicate—which is what unbundling means physically; the mid excess turns negative (−28.6), consistent with mid-pair pooling operating partly through lot-size adjustment, which the capacity null holds fixed by construction.

The placebo samples pin down what the deviation means. Top-pair deficits also appear in non-U.S. yellow corn (−68.0) and in no-quota U.S. soybean meal (−41.4): as Appendix Corollary A5 implies, large firms unbundle for pure scale reasons in any bulk product, and the prize term only strengthens it. The discriminating fact is the within-product contrast: in U.S. yellow corn's own no-race windows (pre-TPA, 2011–12, the slack 2013, post-2022), top pairs co-occur statistically at chance—against 14 percent below chance in race windows of the same product. The same firms that co-loaded at benchmark rates when there was no prize avoid one another in race windows; in 2022, the first post-prize year, the deficit is essentially gone. The mid margin shows the mirror-image timing: its excess is concentrated in race windows and vanishes after exhaustion within the same years.

Running the same benchmark on every digitized product, grouped by how its quota is administered, places the signature where the allocation rule is (Appendix Table A10). The reversal—a top-pair deficit against mid-pair pooling—appears only under binding FCFS yellow corn. No-quota soybean

meal carries the top-pair deficit alone, at nine percent below chance against yellow corn's fourteen, but shows no mid-pair pooling; the auctioned product and the slack FCFS quota that never binds show neither margin. The other two binding-FCFS products, white corn and crude soybean oil, are too thin at the top for the co-occurrence test to have power—no above-threshold pair co-ships in their race windows at all—so their reorganization registers in vessel scale (Section 5.1) and dead-zone timing rather than in the pairwise benchmark. The informative contrasts are yellow corn against the auctioned and no-quota bulk products, and there the mid-pair pooling that distinguishes racing from ordinary scale economies is present only where allocation is chronological and binding.

Because realized exhaustion is an equilibrium object and the dyadic inference is nonstandard, three further designs discipline these deviations; Appendix F reports each in full. Pooled randomization inference puts the race-window top deficit at $z = -10.3$ against $z = 1.5$ for the mid excess, so the top margin is the powered one and the mid excess is supportive rather than load-bearing. A dose design in the *scheduled* out-of-quota wedge—fixed by the agreement, immune to realized-exhaustion endogeneity—gives −1.75 pairs of top-pair avoidance per percentage point of wedge (exact $p = .005$); its own placebo on no-quota soybean meal gives −0.91, so the conservative reading is the differential dose, −0.84 pairs per point (Table A21). And restricting both regimes to a common support of manifest tonnage leaves the top deficit large and significant in both regimes while the mid excess keeps its race-only sign (Table A11); the trim conditions on the chosen fleet—it removes the large solo charters the race created—so the matched-fleet contrast is a lower bound on the race effect, not the preferred estimate.

Buying together versus shipping together. A buying alliance could be about bargaining rather than timing—pooled purchase volume as leverage against the exporter, the tariff race incidental. Two facts bound that motive to first order. First, joint negotiation does not require a shared hull, so bargaining-motivated pooling should appear as common-exporter purchasing *off* shared vessels: same exporter, same month, different manifests. Run against the swap benchmark like every other margin, that signal moves against the race calendar: off-manifest common-exporter purchasing sits *below* its null in every binding phase (pooled −0.011 to −0.025 across the size bands, upper-tail $p \geq .95$) and above it only in slack years (+0.024, exact $p = .016$ at dispatch scale; +0.042, $p = .002$ for the largest pairs; Appendix Table A20 notes). When the prize is on, common-counterparty

purchasing collapses onto shared vessels; when it is off, it appears off them.

Second, the price schedule leaves the bargaining motive little to buy: the volume discount in residualized unit values is exhausted by roughly 500 tons and essentially flat from there through 33,000 (Figure 8), and the Agricol club's race-window premia are no lower than unorganized small buyers' (3.2–4.0 USD/t, within noise; Table A29). What pooled cargo buys at dispatch scale is a vessel's departure, not a better price; how coordination might shift the division of the saving is the companion paper's object, and the split used here is the equilibrium one, measured under the realized coordination.

The exporter as coordination platform (T7). The last test of the mechanism is a pairwise difference-in-differences on the probability that two importers ship together, cut by whether the two firms *together* can fill a vessel. The unit of observation is an ordered pair of importing groups (a, b), both active in calendar arrival-month t; $P_{abt} = 1$ if any of the two firms' cargo travels on a common manifest that month. Each firm's predetermined size x_i is its mean monthly yellow-corn tonnage over the months before the agreement took effect (through April 2012[3]. Pairs are grouped by *combined* size $(x_a + x_b)/V$ relative to the dispatch threshold into six bands: $<0.25V$, 0.25–$0.5V$, 0.5–$0.75V$, 0.75–$1.25V$, 1.25–$2V$, and $\geq 2V$. The theory's central prediction attaches to pairs whose pooled cargo fills one vessel with a combined size of 0.75–$1.25V$.

Separately by band, stacking the observed allocation over the swap-MCMC benchmark, we estimate

$$P_{abt} = \sum_{\varphi} \beta_{\varphi} \left(\text{Real} \times \text{Phase}_t^{\varphi}\right) + \gamma \,\text{Real} + \alpha_a + \alpha_b + \alpha_{y(t)} + \alpha_{m(t)} + \varepsilon_{abt}, \tag{5}$$

where Real marks the observed allocation against the benchmark; α_a, α_b are importer fixed effects for the two pair members; $\alpha_{y(t)}$ and $\alpha_{m(t)}$ are year and calendar-month fixed effects, the latter absorbing the seasonality of grain arrivals. Phase_t^{φ} classifies each month from May 2012 on by the fraction of that year's quota already arrived by month-end: *opening* (0–33%), *stockpiling* (33–66%), *depletion* (66–100%), *post-exhaustion* (>100%), with non-binding years labelled *slack*. The omitted category is the pre-TPA period ending April 2012, so each β_{φ} measures the excess of observed over

[3]Volume by post-TPA entrants represents only 3.7% of post-period tonnage

benchmark co-shipping in phase φ relative to the pre-TPA baseline. A pooled variant replaces the phase indicators with a single $\text{Post}_t = \mathbf{1}\{t \geq \text{May 2012}\}$. Observations are weighted by the two firms' combined import volume in month t, so the coefficients reflect the coordination of the cargo that actually moves.

Two firms can co-ship either because they buy from the same exporter who bundles their cargos or because the cargo they bought from different exporters is booked into the same vessel. We split the co-shipping probability by whether the two firms' cargo comes from the same exporter,

$$P^{\text{any}}_{abt} = P^{\text{same}}_{abt} + P^{\text{diff}}_{abt}, \tag{6}$$

where P^{same} is the probability of sharing a manifest sourced from a *common* exporter and P^{diff} of sharing one sourced from *different* exporters; (5) is run on each. Because the dyadic cluster variance is often undefined here, inference is by randomization against the benchmark draws, and the whiskers we report are randomization standard errors (Appendix E). We estimate 6 for U.S.-origin shipments, where the first-come prize operates and for non-U.S. shipments as a spillover check.

Figure 4 reports the net (any co-shipping) phase coefficients in three panels—largest pairs ($\geq 1.25V$), mid-sized pairs at dispatch scale (0.5–$1.25V$), and smaller pairs ($<0.5V$); Figure 5 splits each group into its same-exporter and different-exporter components; the six finer sub-bands are in Table 3. Three results follow. First, the pooled coefficients recompose co-shipping through the exporter at every scale: same-exporter co-shipping rises above the benchmark in every group (and in all six sub-bands, +0.02 to +0.06, all at the randomization floor) while different-exporter co-shipping falls below it everywhere (-0.03 to -0.08). The margin is economically consequential because the two kinds of voyage split the market almost evenly: 47% of post-TPA U.S. shared cargo travels on single-exporter manifests and 53% on multi-exporter manifests, while across all manifests single-exporter voyages carry 61% of volume.

Second, the *net* effect has the shape the dispatch threshold implies: pooled net co-shipping rises only in the dispatch-scale group (+0.012), and within it only where the pair's combined cargo fills roughly one vessel (+0.031 in the 0.75–$1.25V$ sub-band), turning *negative* on either side (-0.011 among smaller pairs, -0.031 at 1.25–$2V$). Two firms pool into a net increase in shared shipping when together they reach a dispatch load, and the effect reverses once they would overshoot it.

Third, coordination tracks the quota's fill within the year. In the dispatch-scale panel, net co-shipping is absent in the opening phase (−0.008), appears as the quota fills (+0.022 in stockpiling), peaks in depletion (+0.052; same-exporter +0.078)—the window in which a vessel's arrival order actually determines tariff treatment—and vanishes after exhaustion (−0.001).

The largest pairs run the cycle in reverse: their net co-shipping is most negative in the opening months (−0.048, with different-exporter co-shipping at −0.106), when a large firm's own early charters are its race position, and the avoidance is essentially gone once the quota is spent (−0.006). Mid-cycle their net co-shipping turns briefly positive (+0.044 in stockpiling, same-exporter +0.104). A natural reading is the residual margin—whole loads committed, leftover cargo riding organized service, its value graded in $\xi(V-\xi)$—but the model's own sharper test rejects it: sorting these pairs by the residual value $\xi(V-\xi)$ of their largest member, the mid-cycle elevation shows no ξ-gradient (Appendix Table A20). Judged against null-draw means the net uptick vanishes for both halves, while same-exporter co-shipping is elevated at the randomization floor in both—if anything more strongly where residuals are worth least. The mid-cycle cell is therefore the platform channel of Appendix Proposition A6(ii)—large firms' purchases concentrating on the same vessel-capable traders—not residual placement; the model leaves this mid-cycle sign unsigned precisely because both channels operate there, and this test tells them apart (Section 3).

The smaller pairs recompose—same-exporter up, different-exporter down—without ever pooling into a net increase, consistent with pairs that cannot reach a dispatch load even jointly (their margin is the bloc, Section 5.3). The within-year profile is the model's quota cycle (Proposition 2) operating on the coordination margin itself.

Exact inference tempers the *net* profile, in a way that is itself informative. The swap draws preserve each month's realized manifests and capacities, so a phase coefficient's null distribution need not be centered at zero: a typical draw delivers −0.043 of the largest pairs' −0.048 opening coefficient and +0.029 of the dispatch-scale +0.052 depletion peak, leaving assignment-margin residuals of −0.005 ($p = .64$) and +0.023 (one-sided $p = .08$). The within-year *net* profile therefore operates largely through when, and at what scale, vessels sail—structure the null carries, and what unbundling means physically—with a marginal assignment residual on top. The *same-exporter* phase coefficients, by contrast, exceed their null-draw means at the randomization floor in every binding phase and in no slack year, and it is on that margin, and on the cross-regime contrasts, that

the within-year evidence rests.

Two problems could manufacture this profile: the realized phases are equilibrium objects—racing itself shifts the fill path—and co-shipping was trending upward before the TPA. Two falsifications address them (Appendix Table A15; details in Appendix F). Reassigning phases from the *scheduled* fill path flattens the dispatch-scale net profile and attenuates the largest pairs' front-loaded avoidance: the within-year concentration is organized around the actual exhaustion dates, which move from year to year, not around the calendar.

Estimating the identical design on 2004–2011 data with a pseudo-treatment in 2008 splits the verdict: the net margin passes—dispatch-scale net co-shipping shows no positive pre-period effect—but the same/different-exporter recomposition was already drifting in the treated directions at roughly half the treated magnitudes. The drift is not a shrinking-exporter-set artifact and it reversed before the treatment (the 2012 event-study coefficient sits below baseline, and the slack years 2012–13 show no same-exporter elevation), so the binding-year step is not the continuation of a straight line; but it does mean the pooled recomposition cannot carry a prize-based interpretation by itself. We therefore rest the exporter-platform reading on the margins the falsifications leave intact—the binding-year-specific elevation absent in slack years, the depletion-phase structure, the cross-regime contrasts—and on the check the drift makes decisive: holding the exporter composition fixed.

The reweighting holds the monthly exporter composition at its pre-TPA distribution—re-weighting every same-exporter co-shipping event by the exporter's pre-TPA volume share over its share in the event's year, identically on the real allocation and on every null draw, so the test remains exact—and is estimated at both the harmonized-entity and the parent-trader level (Appendix Table A16). At the parent level the margin survives: +0.034 pooled in the dispatch-scale group at the randomization floor, positive in every binding phase, and robust to dropping the three largest binding-year traders outright. At the finest entity level it disappears: those events are carried by entities with no pre-TPA share—incumbent parents' dedicated vehicles and genuine entrants—which is what concentration toward vessel-capable counterparties looks like, but which a strict reweighting cannot separate from composition change. Appendix F reports the construction and a slack-cell caution; binding-year specificity continues to rest on the unweighted margin and the cross-regime contrasts.

The same equation on non-U.S. shipments from Argentina and Brazil is a spillover test, not a placebo: the first-come right does not attach to those routes, but the organization built to race the U.S. quota—the exporter relationships, the chartering arrangements—costs nothing to redeploy there. The data show the spillover and nothing more. Same-exporter co-shipping is elevated among smaller non-U.S. pairs (+0.030), consistent with the coordinating organization being reused on Mercosur purchases. Every prize-specific margin is absent: different-exporter co-shipping shows no withdrawal on non-U.S. routes (flat to positive, against −0.03 to −0.08 for U.S.), the dispatch-scale group shows no recomposition, and there is no quota-phase structure—non-U.S. arrivals concentrate after the U.S. quota is spent, and their coefficients show no depletion-phase peak (Appendix Figure A5). What spills across routes is the organization; what does not is the race.

Three consistency checks complete the design, reported in Appendix F: the event-study time path (drift within the pre-period, return to baseline by 2012–13, then a discrete binding-year step; Figure 6); a pair-quarter aggregation and unweighted estimates, both qualitatively identical to the monthly volume-weighted design; and an external replication at a different atom and clock—bill-of-lading parcels, rank tiers—that agrees on the load-bearing margins (the same-exporter rise, the different-exporter fall, and the top-tier recomposition), with sign disagreements confined to the noisier any-margin tiers (Table A14).

5.3 Anonymous rotation, standing blocs, and an external validation

Appendix Proposition A7 predicts that on the anonymous consolidated segment partners *rotate*: even with pervasive sharing, pair-level overlap should show no stable matching. The data bear this out—pair-level Jaccard difference-in-differences are null everywhere and repeated-overlap rates are flat—and under the model the null confirms the mechanism rather than contradicting it. Stable pairs belong to blocs alone (Appendix Corollary A9).

The bloc detector implements Appendix Proposition A8. The universe is sub-threshold groups ($x/V \leq 0.35$) in U.S. yellow-corn race windows 2014–2021, *including* post-TPA entrants under the post-entry size proxy, since the sunk-cost logic says blocs recruit the firms with no pre-period. The edge statistic is a pair's count of distinct joint manifests—booking *intensity*, because binary co-occurrence saturates. The null permutes manifest labels among race-window lots within year ×

port × 14-day windows; conditioning on window co-presence, congestion cannot manufacture excess same-manifest assignment, so the surviving margin is joint booking. An edge requires recurrence in at least four race years at $p < .05$, and blocs are network communities of at least three firms.

Joint booking among small importers is pervasive and statistically unambiguous. Of 379 recurrent sub-threshold pairs, nearly half book jointly at rates chance cannot explain—164 significant at the 5% level against roughly 19 false positives expected, most surviving a false-discovery-rate adjustment, as do all eight within-club edges below.

Two falsifications establish what the detection means (Appendix Table A18). Run unchanged on no-quota soybean meal, the detector still flags a quarter of recurrent pairs (104 of 366), against more than two-fifths in yellow-corn race windows (164 of 379) and one in twenty by chance. Run unchanged on auctioned rice—the product where access is assigned before shipment, so arrival order carries nothing—it finds almost nothing to test: nine recurrent sub-threshold pairs in total, of which three flag. Rice's U.S. lot universe is smaller (649 lots against 4,239), but the thinness is not mechanical: per thousand lots, yellow corn and soybean meal have ninety recurrent pairs each, rice fourteen. Joint booking above chance is therefore a feature of organized bulk consolidation in general, not of the race alone; the attribution of the yellow-corn blocs to the prize rests on the excess over the no-quota rate, on the near-absence of any standing pair landscape under auction, on formation dated to quota onset, and on the external validation that follows.

That validation comes from outside the customs data. In a 2023 interview, the poultry producer Agricol described organizing six other poultry and egg producers to secure in-quota access before larger firms exhausted the quota—chartering ships, tracking them, and coordinating unloading and trucking. The detector, which sees only booking statistics—no names, no sectors—attaches significant joint-booking edges to all seven of the named members. A four-firm core—Agrícola Colombiana, Incubadora Santander, Nutrientes Avícolas, and Villegas Mejía—forms a complete graph, booking jointly at 1.7–2.1 times the null rate in all eight race years; a three-firm periphery connects to it through Inversiones Santagueda. At the strictest threshold the club resolves into this core and periphery rather than a single component (Table A24; Appendix Figure A7 shows the raw co-shipping pattern the statistics formalize). Avícola Pilena, a post-TPA entrant invisible under any pre-TPA-size universe, is the recruited-entrant case the sunk-cost logic anticipates.

The club's economics match Appendix Proposition A8 on every dimension. Each member sits

far below the participation cutoff ($x/V \leq 0.15$), yet the members' pooled cargo, roughly $0.46V$ per month, is one synthetic mid-size importer. Every manifest carrying three or more members enters through a single port; membership is sector-homophilous; and the core's joint booking recurs in every race year from 2014 on. The reported motive—avoiding displacement by the firms that would otherwise exhaust the quota first—is the defensive formation the model predicts (Appendix A.5). Two facts bound the interpretation: the club's vessels are large organized services that also carry top-decile firms rather than exclusive charters, and they arrive no earlier than the multi-importer average.

Coordination of this kind is legal in Colombia, and openly so. Colombian competition law—anchored in Law 155 of 1959, Decree 2153 of 1992, and Law 1340 of 2009—and the guidelines of the Superintendency of Industry and Commerce expressly recognize joint purchasing as a legitimate form of horizontal cooperation; the Superintendency's own illustration of lawful conduct is small firms pooling demand to negotiate inputs against larger rivals, which is what the Agricol club does. The law draws its line at price fixing, market allocation, and the exclusion of suppliers—none of which the arrangements we document involve. The alliances are a lawful organizational response to the allocation rule, not covert collusion, which is why the firms describe them without hesitation.

At scale, the significant-edge graph resolves into three communities whose pooled sizes (X_B/V from 0.23 to 1.4) bracket the dispatch threshold from below and above (Table A23). The poultry-plurality community has 17 members, four of them Agricol-club firms, and 13 remain active in 2023 with 67 joint manifests in 2022–23—persistence after the prize, to which we now turn. One caveat travels with these results: community detection at this edge density is coarser than the seven-firm resolution, because large traders bridge producer groups.

5.4 Formation without dissolution: persistence after the prize

Proposition 3 separates the prize-responsive flow from the persistent stock (T1). Table A8 reports, for the top-50 pre-TPA groups, raw—not benchmark-adjusted—rates: the race-window formation rate (new links among at-risk pairs, 3-year lookback) and the re-link rate (co-shipping again among previously linked active pairs). Formation spikes at the onset of binding quotas—nearly half of all at-risk pairs form a new link in 2014 (0.46, 146 of 320)—then decays to a fraction of that rate once

the standing structures are in place (0.09–0.25 from 2017 on), and does not revive when the wedge reaches zero. The re-link rate, by contrast, never moves: established partners keep co-shipping at a steady rate in every year from 2014 through 2023 (0.46–0.63), with no trend and no post-2022 decline. No-quota soybean meal shows no comparable onset break and no post-2022 change.

Formation tracks the prize; established links do not dissolve with it—the asymmetry that identifies sunk relationship costs, and the microstructure behind the market-level persistence in Appendix Figure A1: in 2023, with the quota unlimited, participation rose, vessels reached their largest median size, and all three detected communities were still booking jointly. The 2013–14 spike partly reflects the origin switch, so the benchmark-adjusted formation rate is the primary estimate going forward; the asymmetry and the post-2022 flow reversion are robust to that adjustment because they compare flows to flows.

5.5 The quota cycle: dead zone and reset

Proposition 2 predicts that imports stop before each January reset, more so the larger the wedge (T5). We measure the late-year share (tonnage accepted November 16–December 31) per product-year and regress it on the scheduled out-of-quota tariff, separately by product, in crossing years (the dead zone exists conditional on realized exhaustion). The wedge coefficient carries the predicted negative sign in all three binding FCFS products, though each is individually imprecise at these sample sizes of four to nine product-years (Table A22, Appendix Figure A6).

Timing is measured on declaration *acceptance* dates, the margin the quota allocates on; a vessel arriving in late December and clearing in January is deferral, not a late arrival. The ship-tracking match makes this deferral physically visible: among AIS-matched voyages, the last declaration on a December-arriving vessel is accepted a median of 25 days after the ship's physical arrival—against 3 days for January–September arrivals—and 28% of December vessels have even their *first* acceptance pushed into January or later: the corn is in port; the filings wait for the reset (Table A4).

The level shifts are what carry the economics: the late-year import share falls modestly in yellow corn (9.2 to 8.0 percent), nearly halves in white corn (12.7 to 6.7), and collapses in crude soybean oil—a near-total shutdown of a product that used to land a third of its annual volume in those weeks (33.3 to 2.0 percent). Figure 7 shows the raw daily pattern: in binding years, acceptances nearly

stop in the week before January 1 and surge more than tenfold immediately after (from roughly 2 to 34 kt per day)—a trough and spike entirely absent in pre-TPA years, whose profile is flat through the boundary.

The January-mass counterpart is significant only for yellow corn (-0.0049, $p = .006$): reset bunching mixes deferred demand with race advancement, and the decomposition awaits the auction control. One accounting note: the yellow-corn extract ends at mid-2023, so 2022–23 late-year outcomes are coded missing, not zero.

5.6 Cost shape: disciplining the value function

The interval logic requires a single-peaked net sourcing term with zero endpoints, not sub-additivity (which the standard vessel cost satisfies while generating *no* interior interval). Four moments, all constructed in no-prize states with residualized unit values, pin the shape (Tables A25–A26, Figure 8). (1) The unit-value premium over the world price is U-shaped in lot size—a fifteen-percent premium on the smallest lots, a flat minimum over the cost-efficient range of roughly 550–4,000 t, and a mild upturn in the largest—rejecting ever-increasing bulk discounts and locating the interior cost-efficient scale the single-peaked value function needs. (2) Sailing small and alone is expensive: solo lots under 500 t pay a landed-cost penalty of nearly a fifth relative to shared lots of the same size (ratios of 1.205 against 1.025), and the penalty disappears entirely at scale—which signs the ordinary sourcing economy $\theta_0 = s - c\eta > 0$ at small scale and explains why a consolidation market exists at all.

(3) Underfilled near-V solo charters (80–100% of V, below-median fill) are more frequent in race windows than pre-TPA (6.1% vs. 3.8%): measured deadfreight, the physical resource cost of unbundling that enters the cost accounting. (4) The post-TPA median manifest is ≈130 times the median group's weekly volume, sustaining the indivisibility primitive itself.

5.7 Who keeps the rent

Throughout, the race operates on a single margin: coordination changes the *likelihood* that a firm's cargo receives in-quota treatment, not the price at which the corn is bought or sold. The pre-TPA transaction-price gaps below, centered on zero, and the small in-quota premia are the evidence

for that reading: what moves across firms is the split of a fixed tariff saving, not the commodity price. This section's only job is to validate that reading and bound the importer share β that the cost accounting takes as a primitive; the international incidence of the rent—why the split is what it is, how it depends on the administration rule, and what it implies for the design of the agreement—is the subject of the companion paper, whose within-match estimates are consistent with the bounds here.

The rent the race allocates is then pinned down by the agreement itself. With out-of-quota imports in every binding year, the value of in-quota treatment on one ton in year y is the scheduled wedge

$$r_y = \tau_y \, \bar{p}_y, \tag{7}$$

the out-of-quota ad valorem rate τ_y applied to the benchmark price $\bar{p}_y$, and the annual rent pool is the wedge on the quota,

$$R_y = r_y \, \overline{Q}_y. \tag{8}$$

Both inputs are scheduled, not estimated: the tariff phase-down and the quota growth path are written into the agreement. Equation (8) gives $R_{2014} = 40.4 \times 2.32$ Mt ≈ 94 million USD, declining with the phase-down to 31–35 million in 2020–21; over the binding years 2014–2021 the pool averages 54 million per year and cumulates to roughly 434 million USD (nominal, as are all dollar figures).

Who captures it is an empirical question about prices. If exporters priced away the whole saving, in-quota transactions would carry an FOB premium equal to r_y; if importers kept it all, FOB prices would not move. The exporter share is therefore

$$1 - \beta_y = \frac{g_y}{r_y}, \qquad g_y = \text{med}\big(p_{iy}^{\text{FOB}} - \bar{p}_y\big) \text{ over race-window declarations}, \tag{9}$$

the median transaction-price gap over the world benchmark, expressed as a fraction of the wedge. Before the TPA the gaps straddle zero (annual medians within −6 to +6 USD/t), so the benchmark is centered. In the high-wedge years 2014–2017, with r_y between 20 and 40 USD/t, the race-window gaps run +1.8 to +5.8 USD/t: exporters capture $1 - \beta_y = 4$–29% of the saving, importers retain $\beta_y = 71$–96% (Tables A27–A28). Applied to the pool, Colombian importers kept on the order of 66–90 of the 94 million USD assigned in 2014, and roughly 310–420 of the cumulative 434 million

over 2014–2021.

Within importers, the distribution is as concentrated as the size structure implies (Appendix Table A19): the six above-threshold groups take 66% of in-quota tonnage—195–264 million USD of the cumulative retained rent—mid-sized groups 28%, and the interview-verified Agricol club 6.7%, a share that falls from 7.3% to 4.3% along the phase-down while post-TPA entrants grow from 2% to 7%. The rent the rule keeps at home accrues mostly to the largest racers; what the defensive alliances secured was access at small scale, not a large slice of the pool—consistent with defensive coordination buying continuity of supply rather than rents.

The estimate degrades where the design says it must. From 2018 on the scheduled wedge falls below 18 USD/t and the annual medians (−2.7 to +6.8) imply shares from −18% to +68%: once the wedge is small relative to transaction-price noise—quality mix, fixture timing against a monthly benchmark—the ratio in (9) is uninformative, and 2022 is excluded outright for a recorded-value anomaly. The conclusions we use are therefore the ones the high-wedge years support: β is high, roughly 0.7 to 0.96, and does not swing violently across those years. Within race windows, buyer-type medians order as bargaining power predicts—above-threshold solo racers pay the smallest premium (3.2 USD/t), unorganized sub-threshold buyers 3.8, the Agricol club 4.0 (Table A29)—though these differences are within noise.

Two limits on quantification. First, the price data identify the *split* β_y and, with the schedule, the retained rent $\beta_y R_y$; they do not separately identify the model's race primitives ρ (the probability gain per day of advancement) and η (assembly time per ton), which enter behavior only through the product $\beta r \rho \eta$. Recovering ρ alone would require the distribution of arrival shocks around the exhaustion date, which the paper's design treats as the benchmark's randomness rather than an estimand. Second, the resource-cost side of the ledger—what firms spend to win $\beta_y r_y$ per ton—is bounded rather than point-identified, and Section 6 reports those bounds from the carrying-cost and deadfreight moments.

Three limitations bound what the tests establish—the pairwise design is powered only for yellow corn among the binding products; the transaction-price split is measured credibly only in the high-wedge years; and the same/different-exporter recomposition margin is read jointly with its binding-year specificity because of its pre-period drift. Appendix C collects them, together with the data decisions a careful reader should know and the margins the model deliberately leaves outside.

6 What the race costs

The estimates above measure how firms reorganized to compete for the quota; this section asks what that competition cost. Table 4 collects the ledger. The accounting rests on an asymmetry the schedule creates: the race is dynamic, but the rent it contests is a scheduled stock. The quota admits exactly $\overline{Q}_y$ tons at the low tariff whatever firms do, and the wedge is written into the agreement, so the aggregate transfer R_y (equation (8)) is invariant to the racing itself—the contest determines when the quota is claimed, by whom, and at what resource cost, never how much rent there is. Holding fixed what the estimates condition on—the world price, the resale price, the year's fill path—timing gains are therefore zero-sum (Appendix Lemma A3): the real resources firms spend improving queue position—what the rent-seeking literature calls dissipation—buy no additional access in aggregate. We measure those resource costs and who bears them, and stop there; ranking allocation rules would require taking a stand on whose rents and whose costs count, which the paper does not do. The estimates sort the reorganization into three bins.

Transfers. The rent itself is a transfer from Colombian tariff revenue to quota recipients—94 million USD in 2014, declining with the scheduled phase-down to 31–35 million by 2020–21 and averaging 54 million over the binding years—and the incidence evidence of Section 5.7 says most of it ($\beta = 0.71$–0.96) stays with Colombian importers. The resource costs below do not depend on that split. Stockpiling ahead of the January reset moves consumption across time at carrying cost but responds to a real intertemporal price—a margin any tariff of the same size would create, race or no race.

Coordination that saves resources. Shared vessels economize on real sourcing costs where $s > c\eta$. The cost moments say these economies are real and concentrated at small scale: the 18-point landed-cost penalty on small solo lots vanishes on shared service (Figure 8). Coordination by firms who would gain even at $r = 0$—much of the small-firm consolidation, and plausibly the standing blocs' governance once formed—lowers real sourcing costs. One caveat: coordination value that reflects bargaining transfers from exporters is a transfer, not a resource saving—a component the flat price schedule bounds tightly (Section 5.2)—and the racing-cost bin does not depend on it. The race *accelerated* the arrival of this organization and, through the sunk relationship costs, made it permanent; to the extent the sourcing economies are real, that acceleration brought the savings

forward—an entry that belongs in the ledger next to the race costs below rather than assumed away.

Resource costs of racing. Three components are resources spent improving queue position—spending that reallocates the fixed transfer without enlarging it; two are measured, both through constructions rather than direct observation, and the magnitudes are correspondingly ranges. *Acceleration*: in-quota cargo is accepted 68–81 days earlier than the pre-TPA season—81 across all origins, 68 within U.S. origin, which nets out the harvest calendar by construction (Appendix Figure A11)—and at conventional carrying costs of 5–15 cents per ton-day this costs 9–34 million USD per year, 17–63 percent of the mean annual prize; an upper bound on the race-attributable part, since even the within-origin shift includes any secular re-timing of U.S. sourcing.

Prize-induced unbundling: the excess of underfilled near-V solo charters in race windows (6.1 vs. 3.8 percent of solo manifests) monetizes to under 0.2 million USD per year at bulk freight rates—the measured cost travels through when cargo moves, not empty holds. *Excess organization*: the relationship and governance costs of coordination that exists only under the binding prize, identified by the size cells that participate in none of the no-prize states, are real but not monetized.

What is measured, and what is not: shared service consumes more resources than it saves when its acceleration and organization costs exceed the ordinary sourcing savings, yet is privately chosen whenever that gap is smaller than the prize term. The prize term is measured—the pool and its split (Section 5.7)—the resource costs only partially, so the ledger reports bounds and measured components rather than a single net number.

Two facts about the rules follow without further structure. An *auction* of the same quota delivers the same scarcity rent with none of the race: no acceleration, no unbundling, no dead zone—and the auctioned products in the data display none of these margins. And because the FCFS right attaches to arrival rather than to a license, FCFS keeps the rent largely in the importing country, whereas exporter-held licenses assign it abroad—so the comparison an importing government faces is FCFS against a *domestic* auction, which would avoid the racing costs with the rent staying at home either way. Which rule to prefer is a design question—the answer depends on whose rents and whose costs enter the objective—and belongs to the companion paper. The one qualification persistence forces on any such comparison: the race built durable coordination capital, so the rules must be compared knowing whether that capital would have formed anyway—the no-quota bulk placebo, with its far smaller reorganization, suggests it would not have, at least not by 2023.

7 Conclusion

This paper studied how importers coordinate when a government allocates a valuable trade preference by arrival order at the border. In Colombia, attaching a 30–95 million dollar annual tariff saving to positions in a customs queue changed who ships with whom: pairs of importers came to ship together exactly where their combined cargo fills a vessel; the largest importers shipped alone in the opening months, when their own charters secured early arrival, and solo charters sailing below capacity became more common in race windows. Small poultry producers formed a standing seven-firm alliance that a detector using only booking statistics recovers in full; coordination at every scale came to run through a common exporter; imports stopped before every January reset, most completely in the highest-wedge products; and when the prize disappeared in 2023, the formation of new relationships reverted to its ordinary rate while the existing ones did not unwind.

Overall, the analysis offers three take-aways that travel beyond corn. First, the allocation rule shapes the organization of the market that lives under it: two quotas of identical size, one auctioned and one raced, produce different coordination among importers and different racing costs—margins that tariff-equivalent accounting cannot detect, and inputs that any design analysis of quota administration (the subject of the companion paper) must take from the data.

Second, allocation by waiting, the classic rationing mechanism of Barzel (1974), acquires new economics when a queue position can be improved only through cooperation with competitors: the contest then reorganizes who cooperates with whom—creating coordination among some competitors and removing it among others—rather than only burning resources, and the response is measurable in matched transaction data.

Third, sunk relationship costs convert transient administrative rules into permanent institutions: the coalitions formed to race Colombia's corn quota were still booking jointly a year after the quota ceased to exist. Whether that inherited coordination among downstream competitors matters for conduct in the domestic market is a question this paper's data can pose but its model does not address; it is the natural next question.

References

Alessandria, G., J. P. Kaboski, and V. Midrigan (2010): "Inventories, Lumpy Trade, and Large Devaluations," *American Economic Review*, 100(5), 2304–2339.

Alviarez, V., M. Fioretti, K. Kikkawa, and M. Morlacco (2023): "Concentration and Markups in International Trade," Working paper.

Alviarez, V., M. Fioretti, K. Kikkawa, and M. Morlacco (2026): "Two-Sided Market Power in Firm-to-Firm Trade," NBER Working Paper 31253 (revised May 2026).

Anderson, J. E., and J. P. Neary (1992): "Trade Reform with Quotas, Partial Rent Retention, and Tariffs," *Econometrica*, 60(1), 57–76.

Barzel, Y. (1974): "A Theory of Rationing by Waiting," *Journal of Law and Economics*, 17(1), 73–95.

Beckman, J., F. Gale, and T. Lee (2021): *Agricultural Market Access Under Tariff-Rate Quotas*. USDA Economic Research Service, Economic Research Report 279.

Bassi, V., R. Muoio, T. Porzio, R. Sen, and E. Tugume (2022): "Achieving Scale Collectively," *Econometrica*, 90(6), 2937–2978.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020): "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*, 88(2), 657–691.

Coşar, A. K., and B. Demir (2018): "Shipping Inside the Box: Containerization and Trade," *Journal of International Economics*, 114, 331–345.

Deacon, R. T., and J. Sonstelie (1985): "Rationing by Waiting and the Value of Time: Results from a Natural Experiment," *Journal of Political Economy*, 93(4), 627–647.

Hornok, C., and M. Koren (2015): "Administrative Barriers to Trade," *Journal of International Economics*, 96(S1), S110–S122.

Khandelwal, A. K., P. K. Schott, and S.-J. Wei (2013): "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," *American Economic Review*, 103(6), 2169–2195.

Krishna, K. (1989): "Trade Restrictions as Facilitating Practices," *Journal of International Economics*, 26(3–4), 251–270.

Kropf, A., and P. Sauré (2014): "Fixed Costs per Shipment," *Journal of International Economics*, 92(1), 166–184.

Krueger, A. O. (1974): "The Political Economy of the Rent-Seeking Society," *American Economic Review*, 64(3), 291–303.

Macchiavello, R., and A. Morjaria (2015): "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *American Economic Review*, 105(9), 2911–2945.

MacKinnon, J. G., and M. D. Webb (2020): "Randomization Inference for Difference-in-Differences with Few Treated Clusters," *Journal of Econometrics*, 218(2), 435–450.

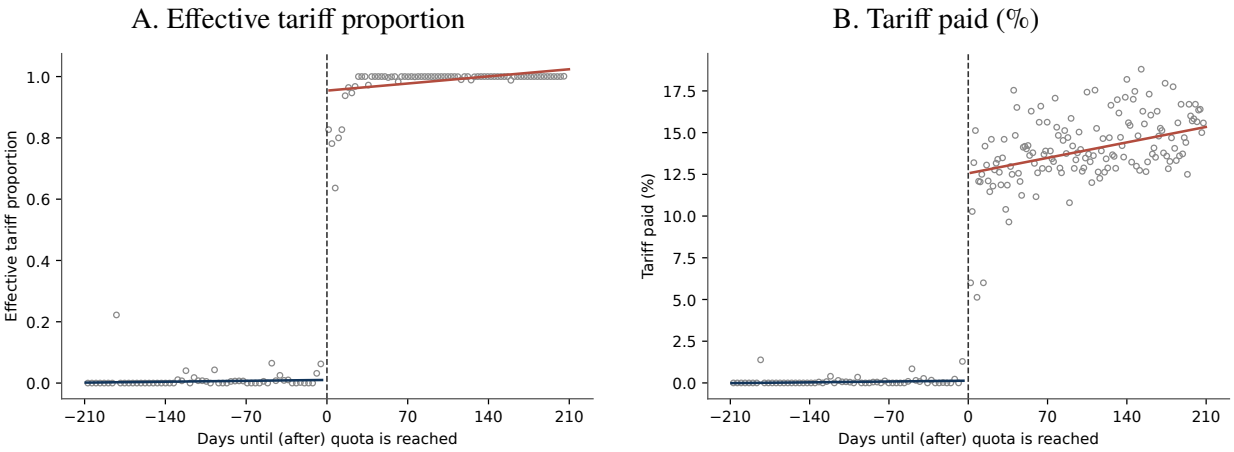
Posner, R. A. (1975): "The Social Costs of Monopoly and Regulation," *Journal of Political Economy*, 83(4), 807–827.

Skully, D. W. (2001): *Economics of Tariff-Rate Quota Administration*. USDA Economic Research Service, Technical Bulletin 1893.

Tullock, G. (1980): "Efficient Rent Seeking," in *Toward a Theory of the Rent-Seeking Society*, ed. by J. M. Buchanan, R. D. Tollison, and G. Tullock. College Station: Texas A&M University Press, 97–112.

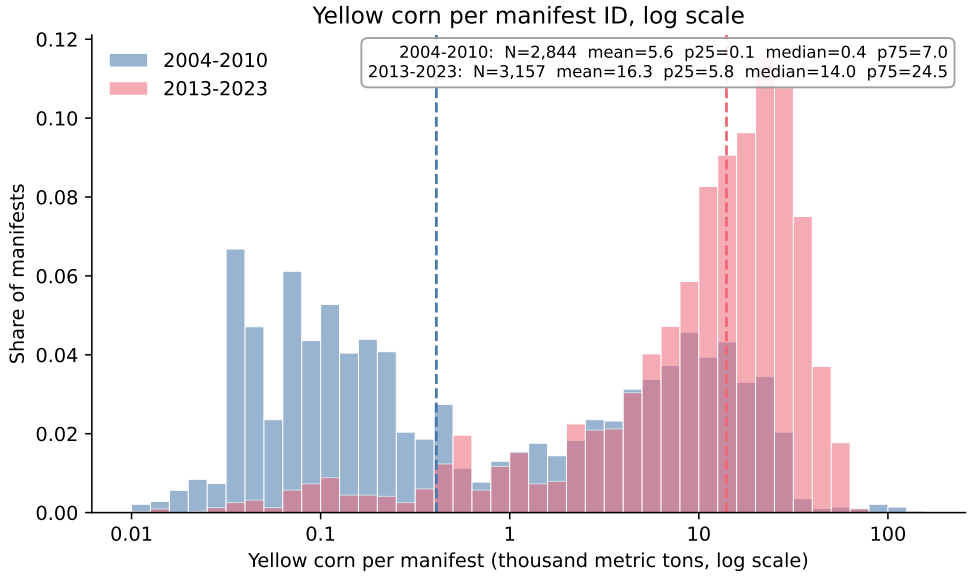
World Trade Organization (2013): "Understanding on Tariff Rate Quota Administration Provisions of Agricultural Products," Ministerial Decision WT/MIN(13)/39–WT/L/914.

Figure 1: FCFS enforcement at the exhaustion timestamp



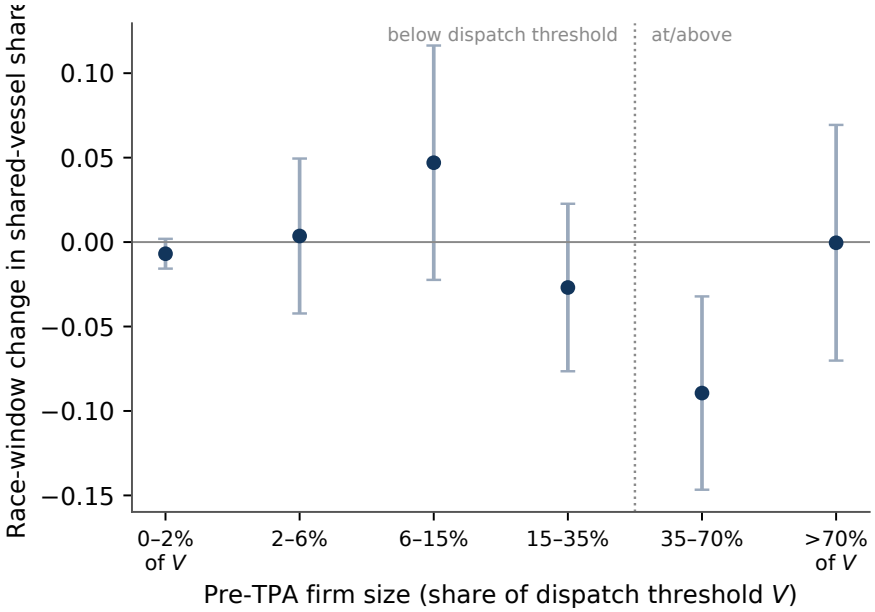
Notes: Binned means against days between a declaration's acceptance timestamp and the year's treatment-based quota-exhaustion timestamp; U.S.–Colombia TPA yellow-corn declarations, crossing years 2014–2019, donut spec of Table 1. Lines: linear fits on each side. Bias-corrected RD estimates with importer and year fixed effects, robust SEs clustered by importer: Panel A jump 0.695 (0.035); Panel B jump 9.99 (0.64).

Figure 2: Distribution of manifest tonnage, by period



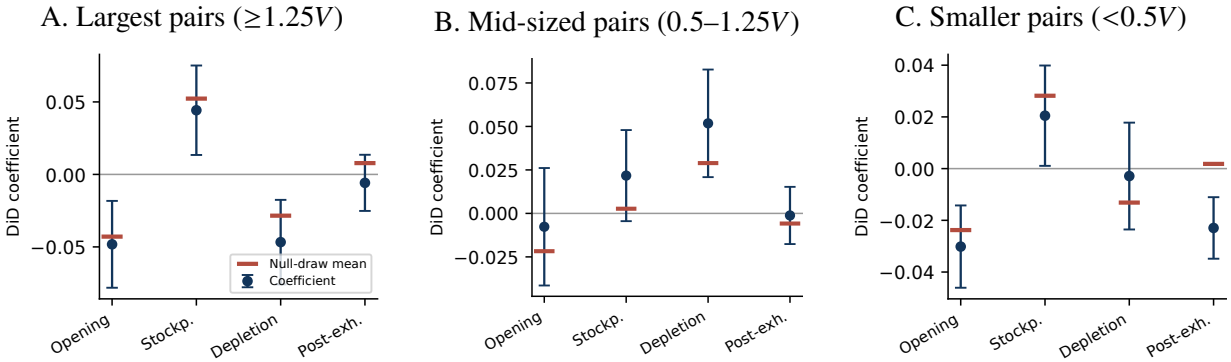
Notes: Tonnage per manifest (all origins, thousand MT), 2004–2010 against 2013–2023; bar heights are within-period shares. Manifest totals sum all declaration lines sharing the manifest ID.

Figure 3: In race windows, importers just below vessel scale withdraw from shared vessels



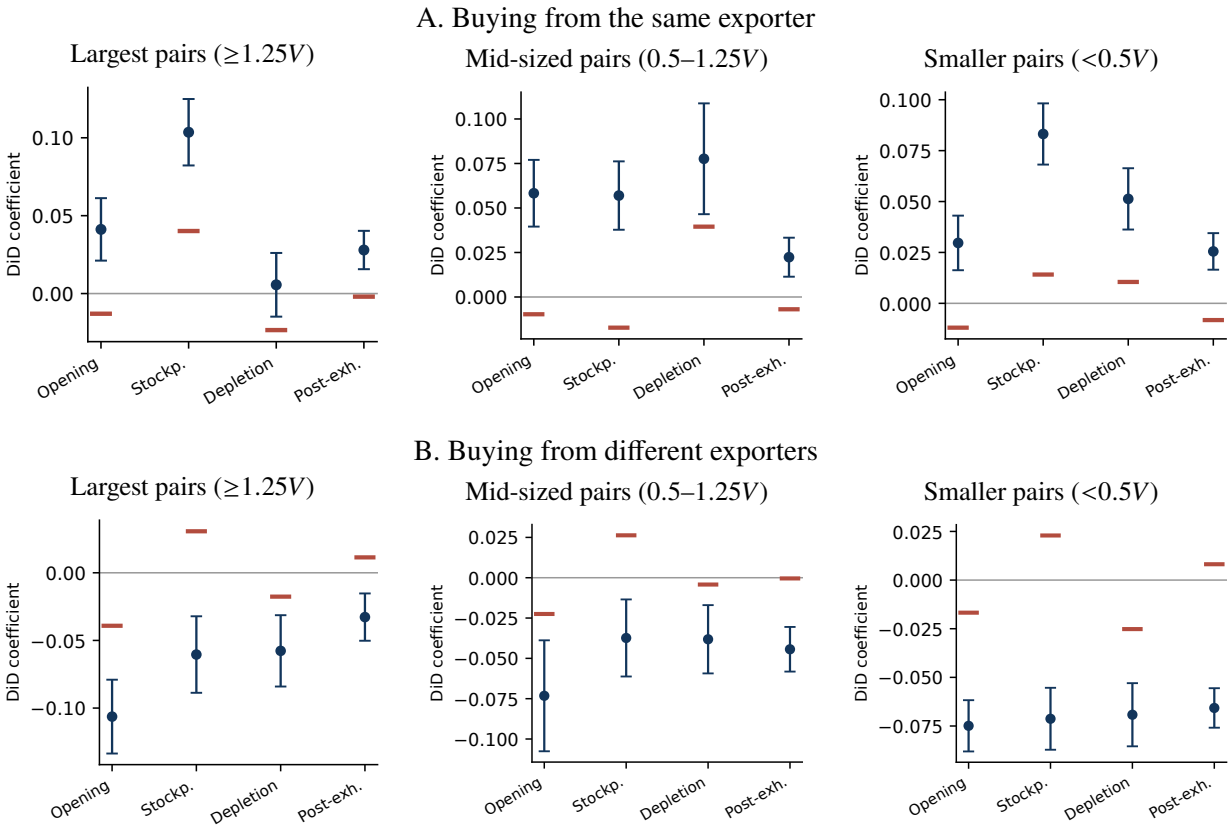
Notes: Race × size-bin coefficients on the multi-group share, race versus post-exhaustion windows within crossing years, U.S.-origin yellow corn; group and year fixed effects, SEs clustered by ownership group, $n = 572$ group-window-years; bars are 95% CIs (Table 2). Dotted line separates sub- from above-threshold bins. The across-period (pre/post) participation change and its non-U.S. placebo are Appendix Figure A8.

Figure 4: Net co-shipping over the quota cycle, by combined pair size (U.S. origin)



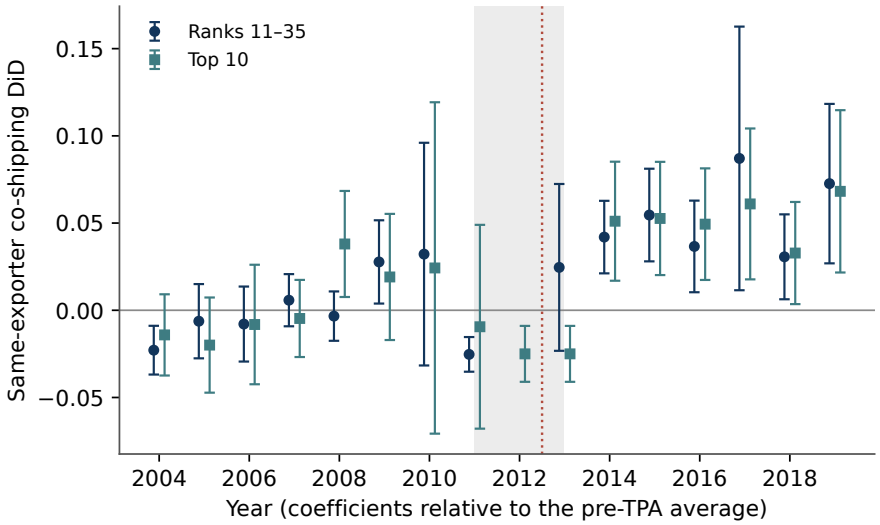
Notes: Estimates of β_φ in equation (5) for net co-shipping (P^{any}), U.S.-origin yellow corn, ≤2019, estimated separately by combined pair size $(x_a+x_b)/V$. The x-axis orders the quota-fill phases—opening (0–33% of the quota arrived by month-end), stockpiling (33–66%), depletion (66–100%), post-exhaustion—with the pre-TPA period (through April 2012) the omitted baseline; estimation is against the uniform swap-MCMC benchmark, weighted by the pair's combined monthly import volume; whiskers are ±1.96 randomization standard errors (Appendix E). Because the swap draws preserve each month's realized manifests and capacities, the exact null distribution of a phase coefficient need not be centered at zero: the rust dashes mark each coefficient's null-draw mean, so the gap between point and dash is the assignment-margin effect, and exact p-values judge each coefficient against its own null distribution. The raw profile of net co-shipping among dispatch-scale pairs builds through stockpiling and peaks in depletion, then vanishes after exhaustion, and the largest pairs avoid co-shipping most in the opening months—but the null-draw means track much of both paths, which is the vessel-structure component discussed in the text. The same-exporter/different-exporter decomposition is Figure 5; the six finer sub-bands are in Table 3; the non-U.S. spillover comparison is Appendix Figure A5.

Figure 5: The exporter decomposition of co-shipping over the quota cycle (U.S. origin)



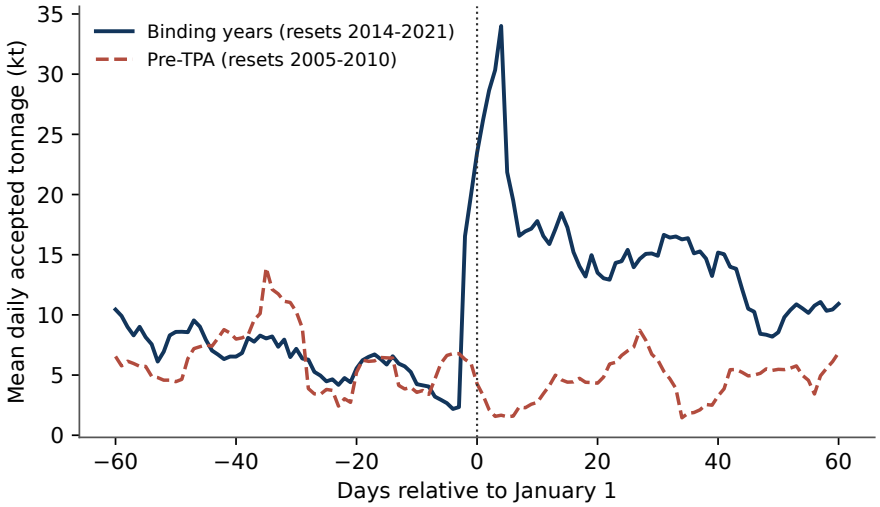
Notes: The decomposition $P^{\text{any}} = P^{\text{same}} + P^{\text{diff}}$ of Figure 4, same specification, sample, benchmark, weights, and inference; within each subfigure, the columns are the size groups of Figure 4, and rust dashes mark each coefficient's null-draw mean as there. Subfigure A is co-shipping through a *common* exporter, subfigure B through *different* exporters. The two kinds of voyage carry comparable cargo: 47% of post-TPA U.S. shared (multi-importer) volume travels on single-exporter manifests and 53% on multi-exporter manifests (of 28.2 million tons); across all manifests, single-exporter voyages carry 61% of post-TPA volume. Same-exporter co-shipping rises above the benchmark in every row while different-exporter co-shipping falls below it in every row: at all scales, coordination recomposes *through* the exporter. Appendix Figures A3–A4 show the underlying manifest composition (importers and exporters per manifest, pre vs. post).

Figure 6: Same-exporter co-shipping over time, dispatch-scale pairs



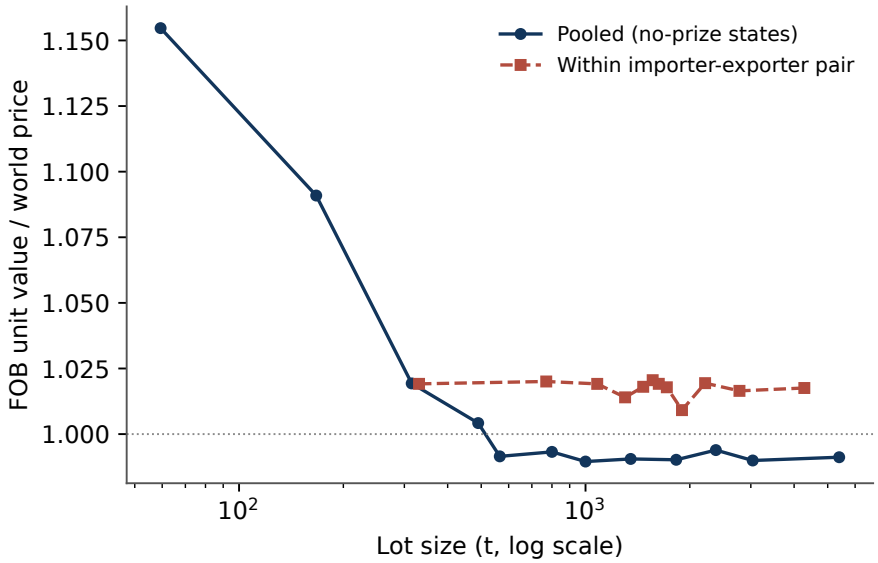
Notes: Year-by-year estimates of the same-exporter co-shipping DiD (equation (5) with the phase indicators replaced by year indicators), U.S.-origin pairs with combined size 0.5–0.75V and 0.75–1.25V, volume-weighted against the swap-MCMC benchmark and expressed relative to the pre-TPA (2004–April 2012) average. Points are DiD coefficients and whiskers are ±1.96 randomization standard errors (Appendix E). The shaded region marks the 2011–12 transition and the dotted line the agreement's entry into force (May 2012). Coordination hovers at its pre-TPA level before the agreement and steps up after it.

Figure 7: The quota reset in daily declarations



Notes: Mean daily U.S. yellow-corn tonnage accepted by DIAN, by day relative to January 1, averaged across the indicated reset boundaries and smoothed over seven days; declaration acceptance dates. Solid: binding years (resets 2014–2021); dashed: pre-TPA (resets 2005–2010).

Figure 8: Cost shape: unit-value premium by lot size



Notes: Residualized FOB unit value over the World Bank maize benchmark by lot-size bin, no-prize states only (Tables A25–A26). A *lot* is one importing group's cargo on one manifest, so lot size is the tonnage that importer moves on that vessel. Both series are decile binscatters of their own samples. The within-pair series conditions on importer–exporter pairs with at least 20 declarations, so its fixed effects are estimable; such repeat relationships almost never ship lots below roughly 300 t, which is why that series has no support at the small-lot end where the pooled solo-lot penalty is steepest.

Table 1: Enforcement RD at the exhaustion timestamp

	Effective tariff proportion	Tariff paid (%)	Log volume (MT)	Log customs value	Log FOB price
Panel A: donut RD (main)					
RD estimate	0.695***	9.986***	-1.590***	-1.552***	-0.001
	(0.035)	(0.639)	(0.277)	(0.278)	(0.011)
Bandwidth (days)	39.4	33.8	26.2	26.8	19.4
Effective obs.	1,675	1,416	1,028	1,043	642
Left-of-cutoff mean	0.006	0.090	7.324	12.598	5.184
Panel B: no donut					
RD estimate	0.246***	3.028***	-0.521**	-0.477**	0.007
	(0.075)	(1.147)	(0.223)	(0.222)	(0.006)

Notes: U.S. yellow-corn declarations within ±210 days of the exhaustion date. Local-linear (rdrobust) discontinuities at the treatment-based quota-exhaustion timestamp T_y^*. Running variable = Fractional days between a declaration's acceptance timestamp and T_y^* (Table A3); (n = 11,740). Bias-corrected estimates; triangular kernel, MSE-optimal bandwidths, importer and year fixed effects as covariates, robust SEs clustered by importer NIT in parentheses. Effective tariff proportion = tariff paid over the scheduled out-of-quota rate. Panel A excludes declarations between the 99th percentile of negative and the 1st percentile of positive running values. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 2: The size profile of the race response: regression estimates

	Multi-group share	Solo-full-vessel share
Race $\times$ x/V 0-.02	-0.0069	0.0003
	(0.0045)	(0.0004)
Race $\times$ x/V .02-.06	0.0036	-0.0001
	(0.0234)	(0.0003)
Race $\times$ x/V .06-.15	0.0427	-0.0002
	(0.0326)	(0.0005)
Race $\times$ x/V .15-.35	-0.0357	0.0126
	(0.0334)	(0.0105)
Race $\times$ x/V .35-.70	-0.0895***	0.0064*
	(0.0292)	(0.0039)
Race $\times$ x/V >.70	-0.0004	0.0111
	(0.0356)	(0.0497)
Group FE	Yes	Yes
Year FE	Yes	Yes
Observations	572	572

Notes: Estimates from equation (4). U.S. yellow corn imports, 2014–2021. Race vs. post-exhaustion windows. Coefficients are the race-window effects by pre-TPA size bin; SEs clustered by ownership group in parentheses. Outcomes are shares of the group's lots that travel in manifests carrying ≥ 2 groups, and the share on its own solo manifests $\geq 0.8V$. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 3: The co-shipping DiD by combined pair size: six sub-bands

Combined size	Pooled	Opening	Stockpiling	Depletion	Post-exh.
Panel A: same-exporter co-shipping (P^{same})					
$<0.25V$	$+0.037^{***}$	$+0.013^{***}$	$+0.077^{***}$	$+0.029^{***}$	$+0.034^{***}$
	(0.006)	(0.010)	(0.010)	(0.009)	(0.007)
	[+0.000]	[-0.019]	[+0.009]	[+0.003]	[+0.003]
$0.25–0.5V$	$+0.046^{***}$	$+0.036^{***}$	$+0.083^{***}$	$+0.060^{***}$	$+0.022^{***}$
	(0.004)	(0.008)	(0.009)	(0.009)	(0.005)
	[-0.000]	[-0.009]	[+0.014]	[+0.015]	[-0.013]
$0.5–0.75V$	$+0.041^{***}$	$+0.032^{***}$	$+0.088^{***}$	$+0.013$	$+0.028^{***}$
	(0.007)	(0.012)	(0.017)	(0.011)	(0.008)
	[-0.000]	[-0.017]	[+0.012]	[-0.002]	[+0.003]
$0.75–1.25V$	$+0.056^{***}$	$+0.074^{***}$	$+0.053^{***}$	$+0.102^{***}$	$+0.028^{***}$
	(0.006)	(0.012)	(0.011)	(0.020)	(0.007)
	[-0.000]	[-0.007]	[-0.026]	[+0.050]	[-0.010]
$1.25–2V$	$+0.019^{***}$	-0.007^{**}	-0.016	$+0.074$	$+0.024^{**}$
	(0.007)	(0.011)	(0.009)	(0.017)	(0.008)
	[-0.000]	[-0.032]	[-0.024]	[+0.048]	[+0.002]
$\geq 2V$	$+0.045^{***}$	$+0.052^{***}$	$+0.136^{***}$	-0.008^{**}	$+0.028^{***}$
	(0.008)	(0.013)	(0.014)	(0.013)	(0.008)
	[-0.000]	[-0.010]	[+0.057]	[-0.037]	[-0.003]
Panel B: different-exporter co-shipping (P^{diff})					
$<0.25V$	-0.061^{***}	-0.052^{***}	-0.087^{***}	-0.049^{***}	-0.059^{***}
	(0.006)	(0.009)	(0.009)	(0.009)	(0.006)
	[-0.000]	[-0.009]	[+0.002]	[-0.011]	[+0.008]
$0.25–0.5V$	-0.073^{***}	-0.085^{***}	-0.066^{***}	-0.075^{***}	-0.071^{***}
	(0.005)	(0.008)	(0.010)	(0.010)	(0.006)
	[-0.000]	[-0.019]	[+0.030]	[-0.029]	[+0.007]
$0.5–0.75V$	-0.076^{***}	-0.076^{***}	-0.058^{***}	-0.083^{***}	-0.080^{***}
	(0.009)	(0.017)	(0.021)	(0.016)	(0.011)
	[+0.000]	[-0.029]	[+0.049]	[-0.023]	[-0.001]
$0.75–1.25V$	-0.034^{***}	-0.075^{**}	-0.027^{***}	-0.019	-0.028^{***}
	(0.011)	(0.024)	(0.015)	(0.014)	(0.010)
	[-0.000]	[-0.025]	[+0.019]	[+0.004]	[+0.000]
$1.25–2V$	-0.051^{***}	-0.113^{***}	-0.029^{*}	-0.002^{***}	-0.065^{***}
	(0.011)	(0.028)	(0.017)	(0.017)	(0.011)
	[-0.000]	[-0.008]	[+0.005]	[+0.045]	[-0.028]
$\geq 2V$	-0.056^{***}	-0.102^{***}	-0.064^{***}	-0.067^{**}	-0.028^{***}
	(0.011)	(0.016)	(0.018)	(0.017)	(0.011)
	[+0.000]	[-0.048]	[+0.043]	[-0.029]	[+0.016]
Panel C: net co-shipping (P^{any})					
$<0.25V$	-0.009	-0.024	$+0.009$	-0.008	-0.011^{**}
	(0.008)	(0.012)	(0.013)	(0.012)	(0.009)
	[-0.000]	[-0.024]	[+0.006]	[-0.008]	[+0.011]
$0.25–0.5V$	-0.012^{**}	-0.033	$+0.023$	$+0.001$	-0.030^{***}
	(0.007)	(0.010)	(0.012)	(0.013)	(0.007)
	[-0.000]	[-0.022]	[+0.035]	[-0.013]	[-0.003]
$0.5–0.75V$	-0.012	-0.033	$+0.035$	-0.031	-0.019^{**}
	(0.010)	(0.018)	(0.022)	(0.017)	(0.012)
	[-0.000]	[-0.032]	[+0.042]	[-0.025]	[+0.005]
$0.75–1.25V$	$+0.031^{**}$	$+0.006$	$+0.029^{**}$	$+0.087^{**}$	$+0.017^{**}$
	(0.012)	(0.023)	(0.017)	(0.020)	(0.011)
	[-0.000]	[-0.022]	[-0.010]	[+0.045]	[-0.009]
$1.25–2V$	-0.031^{***}	-0.102^{***}	-0.043	$+0.064$	-0.043
	(0.012)	(0.028)	(0.019)	(0.021)	(0.013)
	[-0.000]	[-0.024]	[-0.014]	[+0.081]	[-0.029]
$\geq 2V$	-0.008	-0.035	$+0.073$	-0.067	-0.002
	(0.013)	(0.019)	(0.020)	(0.019)	(0.013)
	[+0.000]	[-0.051]	[+0.075]	[-0.047]	[+0.011]

Notes: Estimates of equation (5) at the six finer combined-size sub-bands underlying the three groups of Figure 4; U.S.-origin yellow corn, $\leq$2019, uniform swap-MCMC benchmark, weighted by the pair's combined monthly import volume. "Pooled" replaces the phase indicators with a single post-TPA indicator ($t \geq$ May 2012); phase columns are the quota-fill phases of Figure 4; the omitted baseline is the pre-TPA period through April 2012. Randomization standard errors in parentheses and null-draw means in brackets beneath (Appendix E); *, **, ***: two-sided exact randomization $p < .10, .05, .01$ against the coefficient's own null distribution, which the draws' inheritance of the realized vessel structure can center away from zero. Net co-shipping rises only where the pair's combined cargo fills roughly one vessel ($0.75–1.25V$, pooled +0.031, depletion +0.087) and turns negative just above ($1.25–2V$).

Table 4: The cost of the race: components, measurement, and magnitudes

Component	Model object	Measured as	Bottom line
The stakes (transfer, not a cost)	$R_y = r_y \overline{Q}_y$, eq. (8); importer share β, eq. (9)	Scheduled wedge × scheduled quota; median race-window FOB gap over the world price	94M USD (2014) declining to 31–35M (2020–21); mean 54M; importers retain $\beta = 0.71$–0.96
Acceleration	Carrying cost of advanced cargo-time, $c\,\eta\,x(V-x)$	In-quota cargo accepted 68–81 days earlier than the pre-TPA seasonal center (within-origin and all-origin constructions; Fig. A11); carrying costs for grain	9–34M USD per year, an upper bound on the race-attributable part
Deadfreight from unbundling	$\phi(\xi)$, eq. (3)	Excess share of underfilled (80–100% of V) solo charters in race windows vs. pre-TPA	6.1% vs. 3.8% of solo manifests; ≈0.03–0.2M USD/yr at 15–25 USD/t of unused capacity
Excess organization	F, sunk links ζ, eq. (11)	Size cells coordinating under binding FCFS but in none of the no-prize states; formation spike at quota onset (0.46 of at-risk pairs in 2014)	Real but not monetized; its persistence after the prize is Section 5.4
Forgone sourcing economies	$s\,x(V-x)$, eq. (1)	Landed-cost penalty of small solo lots vs. shared service (Figure 8)	Solo lots under 500 t pay ≈20% over benchmark; sharing removes it—the economies unbundling gives up

Notes: The ledger of Section 6. Each row names the model object, its empirical counterpart, and the measured magnitude. The transfer row sets the scale of the stakes; the remaining rows are resource costs. Acceleration and deadfreight are measured directly; organization costs are identified by participation that exists only under the binding quota but are not monetized. Timing gains are zero-sum across firms (Appendix Lemma A3), so the resource rows are spending that reallocates, without enlarging, the transfer.

A Theory appendix: the model in full

This appendix develops the model of Section 3 in full: the environment and its primitives, the results the main text states only in summary, and all proofs. Numbered results carry the prefix A; the three propositions of the main text keep their main-text numbers and are proved here.

A.1 Environment: what firms take as given

Policy and the prize path. In a quota year, tariff t_L applies to the first $\overline{Q}$ tons accepted and $t_H > t_L$ thereafter. With out-of-quota imports in every binding year (fill ratios 1.19–1.75), the value of in-quota treatment on one ton is the wedge $r = t_H - t_L$ (evaluated at the benchmark price under ad valorem rates). The quota fills over the year along a path firms take as given: the wedge is available while the quota has room, vanishes at the exhaustion date T^*, and reappears at the January reset. Of the saving on a lot, the importer retains share $\beta \in (0, 1]$ and its exporter the rest; β is measured, not modeled (Section 5.7).

Prices, volumes, and the single channel of the race. Each importer i has annual cargo $x_i > 0$, fixed. Two assumptions fix the price environment. First, the international price of corn is taken as given: Colombia's imports are small relative to the world market, so no importer's shipping choices affect the price at which cargo is procured. Second, output markets are perfectly competitive: importers resell or process the grain at prices they do not influence, so the value of in-quota treatment on a ton is the tariff saving itself, common to all firms, and no strategic interaction operates through output prices. The race therefore operates through a single channel: shipping choices change the *probability* that a lot is accepted before T^* and so receives the wedge. Nothing else in a firm's profit moves.

Cargo and vessels. The shipping technology has a single kink, on which the analysis builds. Bulk carriage becomes worthwhile only once a committed load reaches a scale V, the *dispatch threshold*: at V the per-ton cost of dispatching a vessel immediately falls to the bulk rate, so V is the smallest cargo that can sail on its own schedule instead of waiting to be consolidated onto someone else's vessel. A firm can reach V alone, if its cargo is large enough, or by pooling with others—which is what makes racing a joint-production problem: a firm short of V must either wait for the market to assemble the missing tonnage or combine with partners to make up a dispatch load, so the value of coordinating rises with both the cargo a firm has to move, x, and the scale it is missing, $V - x$—the product $x(V - x)$ that the departure cost below makes precise. Writing $x_i = n_i V + \xi_i$ with $n_i \in \{0, 1, \dots\}$ and residual $\xi_i \in (0, V]$, we first analyze firms with $n_i = 0$ and Section A.3 turns to $n_i \geq 1$, whose residual cargo drives the unbundling at the top. Every shipment is small relative to the annual market.

Departure and arrival. Normalize a full vessel's departure to time zero. A lot of size $x \in (0, V]$ using regular consolidation waits for the market to assemble its missing scale: departure $d_W(x) = \eta(V - x)$, where $\eta > 0$ is the market's assembly time per ton. An organized shared vessel that commits the dispatch load departs at zero, advancing a member's cargo by $\eta(V - x)$. Advancing cargo uses real resources—expedited procurement, inventories, rescheduling—at carrying cost c per ton per unit time: $C_A(x) = c\,\eta\,x(V - x)$; because the time advanced is $\eta(V - x)$, the carrying cost necessarily scales with η, which is why the η comparative static in Proposition A4 is signed by $\beta r\rho - c$. A shipment departing at d is accepted by customs at $d + \varepsilon$, where the shipment-specific delay ε (sailing, port congestion, filing, processing) has distribution Γ with density γ on $[0, \bar{u}]$, independent across shipments. A shipment is in-quota if accepted before T^*: $p(d) = \Gamma(T^* - d)$, with T^* *given*.

Assumption A1 (Interior race)**.** $\eta V \le T^* \le \bar{u}$: exhaustion occurs while arrival order is still uncertain. In slack years $T^* = +\infty$ and the prize term vanishes.

Proposition A1 (The probability channel under uniform delays)**.** *Let ε be uniform on $[0, \bar{u}]$ and Assumption A1 hold. Then $p(d) = (T^* - d)/\bar{u}$ and the probability gain from pooling a lot of size x onto an organized early vessel is*

$$\Delta p(x) = \rho\,\eta(V - x), \qquad \rho \equiv 1/\bar{u}, \tag{10}$$

independent of T^*. Given the prize path, each firm's mode choice is a dominant-strategy choice.

The slope ρ is the reciprocal of shipping-time dispersion: when sailing is predictable, a one-day advance is very likely to change the allocation, and the race is steep.

Corollary A2 (Scarcity and the value of timing)**.** *Under uniform delays, the private value of advancing one ton by one unit of time is $\beta r\rho$ when the quota binds within the season and zero otherwise: expected scarcity enters through $r\rho\,\mathbf{1}\{\overline{Q} < X\}$.*

Lemma A3 (Timing gains are zero-sum)**.** *Fix the fill path and let the quota bind. Aggregate in-quota tonnage equals $\overline{Q}$ under every profile of shipping choices, so any reallocation of shipping that raises one lot's probability of in-quota treatment lowers other lots' treatment by the same expected tonnage. Summed over firms, the rent gains from timing are zero; the resources spent producing them are not.*

The lemma underlies the cost accounting of Section 6 and requires no market equilibrium: it holds conditional on the realized fill path, which is how the empirical tests condition as well.

Sourcing economies and organization costs. Sharing also creates ordinary freight and procurement economies; competitive matching makes the optimized saving a function of own size, and we use $\Delta\kappa(x) = s\,x(V - x)$, $s \ge 0$, the same cargo-scale logic as acceleration; $s = 0$ is the pure-race

benchmark, and the cost moments of Section 5.6 discipline this shape directly. Organizing a group costs member i

$$F_i(N) = \underbrace{\textstyle\sum_{j\in N\setminus\{i\}} \zeta_{ij}}_{\text{relationship formation (sunk)}} + \underbrace{a_i\,\omega(Q_N)}_{\text{per-season governance}}, \tag{11}$$

with first-season cost F and continuation cost $F^m < F$; the sunk component ζ drives the dynamics of Section A.5.

A.2 The coordination game

The game is now fully specified. *Players:* importing groups, plus a competitive fringe of consolidators who offer shared-vessel contracts (departure condition, cargo shares, cost shares) with charter payments driven to cost. *Actions:* each season, a group with cargo $x \in (0, V]$ chooses regular consolidation, an organized shared vessel (if the contract attracts a dispatch load), or—jointly with others—forming a standing bloc; a group with $x \geq V$ additionally chooses how to ship its residual (Section A.3). *Payoffs:* the expected rent $\beta r\, x\, p(d)$ implied by the chosen departure, minus carrying, sourcing, and organization costs. Because prices and the fill path are given, payoffs are additive across lots and the strategic content is confined to who pools with whom.

An importer's incremental payoff from coordinating combines the sourcing saving, the acceleration cost, and the expected rent gain $\beta r\, x\, \Delta p(x)$:

$$g(x) = \big(s + \eta[\beta r\rho - c]\big)x(V - x) - F \;\equiv\; \theta\, x(V - x) - F, \qquad \theta \equiv s + \eta(\beta r\rho - c),$$

as in equation (1) of the main text. The term s collects every value of sharing that does not operate through the prize—real freight and procurement economies, but also any bargaining transfer a pooled purchase extracts from the exporter; the participation predictions do not depend on the split, and Section 5.2 bounds the transfer component directly. The term $\eta(\beta r\rho - c)$ is the *net* value of the time pooling buys: each unit of advanced cargo-time earns $\beta r\rho$ in expected rent and costs c in carrying.

The closed-form benchmark is Proposition 1(i) of the main text: with uniform delays and $\theta V^2 > 4F$, importers join iff $x \in (x_-, x_+)$, with the cutoffs of equation (2); if $\theta V^2 \leq 4F$ no one joins, and a firm with $x = V$ departs alone. Its proof, and every other proof, is collected at the end of this appendix.

The combined-size statement of Proposition 1(i) treats a coalition's pooled cargo as one lot: organized early dispatch is worth forming iff $X_B \in (x_-, x_+)$ evaluated at the coalition's joint cost F_B, with value $\theta X_B(V - X_B) - F_B$ maximal for X_B near $V/2$ and negative once X_B overshoots V; the relevant alternative is consolidating the pooled block, which departs at $\eta(V - X_B)$.

The interval shape follows from the payoff rather than from an assumption: g approaches $-F$

at both endpoints because a tiny importer moves too little cargo to recover the organization cost, while an importer near V gains too little scale or time—consolidation already serves it quickly. The structure does not depend on uniform delays: Proposition A11 below gives the interval under a general delay distribution, with a curvature condition replacing the closed form. The combined-size statement is why the co-shipping tests of Section 5 cut pairs by combined size $(x_a+x_b)/V$.

Proposition A4 (Comparative statics)**.** *Under $\theta V^2 > 4F$: increases in s, r, β, or ρ, or a decrease in c, expand (x_-, x_+); an increase in F contracts it; an increase in η expands it iff $\beta r \rho > c$.*

Corollary A5 (Separating ordinary from FCFS-induced coordination)**.** *If $r = 0$ or the quota is slack, an importer joins only if $(s - c\eta)x(V - x) > F$. A binding FCFS prize weakly expands this ordinary region and can create one where none existed. The organizing contrast throughout the empirics is between the no-prize coefficient $\theta_0 = s - c\eta$ and the binding-FCFS coefficient $\theta_1 = s - c\eta + \beta r \rho \eta$.*

A.3 Unbundling at the top

Consider a firm with $x = nV + \xi$, $n \geq 1$. Its whole loads travel on its own charters, departing at zero with no organization cost. The margin is the residual ξ: consolidate it (cheap, late), organize it (the interval logic applies to the *lot* ξ), or *internalize* it—stretch it across the firm's own early charters at deadfreight cost $\phi(\xi)$, $\phi(0) = 0$, $\phi' > 0$.

The internalization condition is Proposition 1(ii) of the main text: internalization beats consolidation iff condition (3) holds, the internalized set is empty at $r = 0$, and it expands in r, β, ρ. Ordinary forces push co-loading in the same direction at all sizes; the internalization margin is what moves the top of the size distribution in the opposite direction.

Proposition A6 (The exporter as coordination platform)**.** *Suppose full charters require traders able to supply V per voyage—a strict subset of exporters—while consolidated cargo draws from the full exporter distribution, and suppose an organized vessel's consolidator procures the pooled cargo from a single counterparty. Then (i) members of an organized vessel mechanically share the exporter of record, so prize-induced pooling raises same-exporter co-shipping and, by diverting cargo from open consolidation, lowers different-exporter co-shipping, at* every *combined size; (ii) unbundling concentrates large firms' purchases on the large-trader subset, raising the probability that two large firms transact with the same exporter even as their co-loading falls.*

Buying through a common exporter is the cheap form of coordination: it economizes on the pairwise link costs that direct coordination requires. The recomposition of co-shipping *through* the exporter is the mechanism's dyadic signature, and the decomposition of equation (6) is built to detect it.

A.4 The prize over the quota cycle

The prize path within the year is not flat, and the coordination margin should follow it. Two forces shape the profile. First, the timing gain is concentrated on cargo *at risk*: a lot whose unadvanced arrival would fall near or after T^* gains the full wedge from advancement, while a lot that clears comfortably gains nothing. Early in the year almost no cargo is at risk; as the quota fills, the at-risk mass grows, peaking as exhaustion approaches; after T^* it is zero. Second, whole charters are committed at the season's start: a firm with $n_i \geq 1$ books its own early vessels then, so its avoidance of shared service is front-loaded. The quota reset adds the terminal margin: a unit due after exhaustion can wait for January at carrying cost c_w per unit time and enter in-quota with next season's early-access probability p_E^+.

The four-phase partition and the dead zone are Proposition 2 of the main text; the two forces above generate its profile, and the proof is below.

A.5 Institutions, blocs, and persistence

Coordination is supplied by two institutions the data can tell apart.

Proposition A7 (Participation versus partner composition)**.** *In an active-contract equilibrium, every lot on an organized early vessel has size in* (x_-, x_+) *(a small firm's whole cargo or a large firm's residual). Conditional on feasibility, the baseline imposes* no *restriction on partners' relative sizes or identities: competitive consolidation predicts partner* rotation*, so pair-level overlap statistics should show no stable matching even when shared shipping is pervasive.*

Proposition A8 (Blocs extend coordination below the firm-level cutoff)**.** *Let a bloc B pool cargo* $X_B = \sum_{i \in B} x_i \leq V$*, booked jointly at per-season governance cost* F_B^m*, with sunk pairwise formation costs* $\bar{\zeta}_B$ *per link (reduced by a common downstream use and a common port,* $F_{ij} = F_i - A_z \mathbf{1}\{z_i = z_j\} - A_h \mathbf{1}\{h_i = h_j\}$*). Then: (i) the bloc earns positive operating surplus iff* X_B *lies in the interval of Proposition 1 evaluated at* F_B^m*; (ii) a firm below* $x_-(F_B^m)$ *obtains organized early service only as a bloc member; (iii) a symmetric n-member bloc forms only if discounted surplus covers* $n(n-1)\bar{\zeta}_B/2$*, so membership is bounded and decreasing in link costs; (iv) feasible blocs are homophilous—same-use, same-port sets support more members and a wider pooled-size range.*

Corollary A9 (Two institutions, two signatures)**.** *Consolidator service creates no relationship capital: partners rotate, and stable-pair statistics are null. Bloc members co-ship with a fixed homophilous set repeatedly. Stable, recurrent pairs identify blocs only; their absence in the anonymous segment corroborates the mechanism rather than contradicting it.*

The corollary reverses the natural reading of a null result: pervasive shared shipping with *no* excess pair stability is what competitive consolidation looks like. When being late is costly, small

firms' bloc participation is *defensive*—organize or be displaced by the shipments of those who did—which is the motive reported in the field (Section 5.3).

The organization cost (11) has a sunk component $\zeta = F - F^m$. In the stationary repeated environment where the state θ_t is drawn each season, incumbents pay F^m and entrants F. The resulting band of inaction is Proposition 3 of the main text: incumbents persist in states where no entrant would form, and a permanently extinguished prize stops formation without dissolving existing groups.

Because the single coefficient θ governs the participation interval, the internalized residuals, and the formation flow, their movements across states must be consistent with one θ path; the components of θ are not separately identified. Appendix Table A1 collects the model's predictions and where each is tested.

The randomization benchmark as an exact null

Stated here, tested throughout Section 5, and constructed in Appendix D:

Proposition A10 (The randomization benchmark)**.** *Fix a cell (product, origin, arrival port, non-overlapping arrival window) and condition on the lots present with their sizes and on the realized manifest opportunity set (number of manifests and their tonnage capacities). If there is no timing prize, no organized shared service, and consolidation queueing is* exchangeable*—lots' readiness dates exchangeable across identities and sizes—then the permutation distribution generated by randomly reallocating lots to feasible manifests is the exact null distribution of any assignment statistic, including pairwise co-occurrence. Without size exchangeability, exactness holds for identity-based statistics conditional on the realized size-to-manifest allocation, and both versions should be reported. Under a positive prize, the equilibrium departs from this null in three directions: (i) excess co-occurrence of lots with sizes in (x_-, x_+), concentrated on manifests accepted before T^*; (ii) depressed co-occurrence involving the residuals of above-threshold firms where internalization is optimal; (iii) both departures increase with the expected prize and vanish in no-prize states.*

Proofs and further results

Proof of Proposition A1 With Γ uniform on $[0, \bar{u}]$ and Assumption A1, $p(d) = \Gamma(T^* - d) = (T^* - d)/\bar{u}$ on the relevant range. Pooling advances departure from $d_W(x) = \eta(V - x)$ to 0, so $\Delta p(x) = p(0) - p(d_W(x)) = \eta(V - x)/\bar{u} = \rho\eta(V - x)$, independent of T^*. With the prize path given, each firm's payoff depends only on its own mode, so the mode choice is dominant-strategy. □

Proof of Lemma A3 With the quota binding, cumulative accepted tonnage crosses $\overline{Q}$ within the year, and in-quota tonnage equals $\overline{Q}$ under every profile of shipping choices. Expected in-quota

tonnage summed across lots is therefore constant at $\overline{Q}$; any choice that raises one lot's expected in-quota tonnage lowers the total for all other lots by the same amount. Multiplying by the common wedge r gives the result. □

Proof of Corollary A2

Proof. If $\overline{Q} \geq X$ then $A(t) < \overline{Q}$ for all finite t, $T^* = +\infty$, and $p(d) = 1$ for every d, so timing has no private value. If $\overline{Q} < X$ and Assumption A1 holds, Proposition A1 gives $\partial p/\partial d = -\rho$, so advancing one ton by one unit of time raises expected retained rent by $\beta r \rho$. □

The interval under general delay technologies

Proposition A11 (Coordination interval under general technologies)**.** *Let $K(x) \equiv \Delta\kappa(x) - C_A(x)$ satisfy $K(0) = K(V) = 0$ and let, for $d = \eta(V - x)$,*

$$K''(x) + \beta r\left[-2\eta\gamma(T^* - d) - x\eta^2\gamma'(T^* - d)\right] < 0 \quad \textit{for all } x \in (0, V). \tag{12}$$

Then total coordination value $\mathcal{V}^\Gamma(x) = K(x) + \beta r x\, \Delta p(x)$ is strictly concave with zero endpoints and a unique interior peak. For common first-season cost $F > 0$: if $\mathcal{V}^\Gamma$ peaks below F no importer coordinates; otherwise there are unique cutoffs $0 < x_- < x_+ < V$ and an importer demands a shared vessel iff $x \in (x_-, x_+)$. Strict quasi-concavity of $\mathcal{V}^\Gamma$ suffices for the same conclusion.

Proof. Fix the equilibrium exhaustion time T^*. Let $d(x) = \eta(V - x)$ and $\Lambda(x) = \Gamma(T^*) - \Gamma(T^* - d(x))$. Since $d(x)$ falls in x at rate η, the argument $T^* - d(x)$ rises in x at rate η, so $\Lambda'(x) = -\eta\,\gamma(T^* - d(x))$ and $\Lambda''(x) = -\eta^2\,\gamma'(T^* - d(x))$. The second derivative of the rent-capture term $x\Lambda(x)$ is $2\Lambda'(x) + x\Lambda''(x) = -2\eta\gamma(T^* - d(x)) - x\eta^2\gamma'(T^* - d(x))$. Multiplying by βr and adding $K''(x)$ gives exactly the left side of condition (12), so $\mathcal{V}^\Gamma$ is strictly concave. Moreover, $K(0) = K(V) = 0$, the factor x makes the rent-capture term zero at $x = 0$, and $d(V) = 0$ makes it zero at $x = V$. Strict concavity and equal endpoint values imply a unique interior maximizer. A horizontal line at $F > 0$ lies above the maximum, is tangent at the maximum, or intersects the value function exactly twice; under the tie-breaking convention the first two cases give an empty participation set and the third the claimed open interval. A pointwise increase in $\mathcal{V}^\Gamma$ weakly enlarges the set on which $\mathcal{V}^\Gamma > F$. For the quasi-concavity version: the strict upper level sets of a strictly quasi-concave function on $(0, V)$ with zero endpoint values are open intervals with interior endpoints, which is all the argument uses. □

Proof of Proposition 1(i)

Proof. With common β, equation (1) gives $g(x) = \theta x(V - x) - F$. If $\theta \leq 0$, then $g(x) < 0$ for every $x \in (0, V)$. Suppose $\theta > 0$. The function is strictly concave, satisfies $g(0) = g(V) = -F$, and

reaches its unique maximum at $x = V/2$, where $g(V/2) = \theta V^2/4 - F$. It is positive for some size if and only if $\theta V^2 > 4F$. Solving $g(x) = 0$ gives equation (2); strict concavity implies $g(x) > 0$ exactly on (x_-, x_+). At $x = V$, regular consolidation already departs at zero, so an importer meeting the dispatch threshold ships alone and avoids F. □

Proof of Proposition A4

Proof. Write $\Delta = \sqrt{V^2 - 4F/\theta}$, so $x_\pm = (V \pm \Delta)/2$, with $\partial\Delta/\partial\theta = 2F/(\Delta\theta^2) > 0$ and $\partial\Delta/\partial F = -2/(\Delta\theta) < 0$. Thus higher θ lowers x_- and raises x_+; higher F does the reverse. From (1), $\partial\theta/\partial s = 1$, $\partial\theta/\partial c = -\eta$, $\partial\theta/\partial r = \eta\beta\rho$ (symmetrically for β, ρ), and $\partial\theta/\partial\eta = \beta r\rho - c$. □

Proof of Corollary A5

Proof. If $r = 0$ or $T^* = +\infty$, the rent term is zero and $g(x) = (s - c\eta)x(V - x) - F$. If $s \le c\eta$ this is negative for every interior x. If $s > c\eta$, its positive set, when nonempty, is the ordinary sourcing region. Adding a positive FCFS term raises θ, and Proposition A4 weakly expands that set. □

Proof of Proposition 1(ii)

Proof. Under consolidation, the residual earns expected rent $\beta r\,\xi\,p_W(\xi)$, receives sourcing economies $s^c\xi(V - \xi)$, and incurs no acceleration cost. Under internalization, the residual departs at zero, earning $\beta r\,\xi\,p_E$, at deadfreight cost $\phi(\xi)$ and carrying cost $c\eta\xi(V - \xi)$. The firm internalizes when $\beta r\xi[p_E - p_W(\xi)] - c\eta\xi(V - \xi) - \phi(\xi) - s^c\xi(V - \xi) > 0$; substituting $p_E - p_W(\xi) = \rho\eta(V - \xi)$ from Proposition A1 and rearranging gives (3). At $r = 0$ the left side is negative while the right side is nonnegative, so no residual is internalized; the left side is strictly increasing in r, β, ρ while the right side is constant in them, so the internalization set weakly expands. Internalized residual cargo appears on single-importer charters rather than consolidated multi-importer manifests, yielding the participation statement; the statement for (x_-, x_+) is Proposition A4. □

Proof of Proposition A6

Proof. Index exporters by per-voyage supply capability; let $\mathcal{L}$ be the subset able to deliver at least V, and let firm k's seasonal purchase distribution across exporters be π_k. Consolidated cargo inherits the consolidator's counterparty mix with support on the full exporter set; full charters require a counterparty in $\mathcal{L}$. Unbundling therefore shifts π_k toward $\mathcal{L}$ in the sense of first-order dominance over the ordered partition $\{\mathcal{L}, \mathcal{L}^c\}$. For firms k, l the season same-exporter probability is $\sum_e \pi_k(e)\pi_l(e)$; shifting both distributions toward the smaller set $\mathcal{L}$ weakly raises this inner product whenever within-$\mathcal{L}$ shares are no more dispersed than the pooled shares, strictly when $|\mathcal{L}|$ is small. The shared-vessel statement follows because an organized contract designates a single voyage counterparty for the pooled cargo. □

Proof of Proposition A7

Proof. For any lot of size y, the organized-vessel payoff relative to consolidation is $g(y) = \theta y(V - y) - F$, negative outside (x_-, x_+) by Proposition 1; a lot outside the interval never demands organized service, whatever its owner's total size, and internalization does not place the lot on a shared vessel at all. Competitive consolidators serve only subscribers, so each lot on an observed organized vessel lies in the interval. The baseline payoff depends on own lot size and on aggregate feasibility, not on a partner's size or identity conditional on feasibility; it therefore supplies no ranking of feasible partners and no assortative-matching restriction. □

Proof of Proposition A8

Proof. (i) A bloc subscribing its pooled lot X_B at per-season cost F_B^m has per-season surplus $\theta X_B(V - X_B) - F_B^m$, positive exactly between the two roots of the strictly concave quadratic evaluated at F_B^m. (ii) A firm with $x_i \le x_-(F_B^m)$ has $\theta x_i(V - x_i) \le F_B^m$, so no individual subscription is profitable even at governance cost alone; any participation must pool cargo. (iii) Per-season surplus is bounded by $\theta V^2/4 - F_B^m$, so discounted value is at most $(\theta V^2/4 - F_B^m)/(1 - \delta)$, a constant, while the sunk formation requirement $n(n-1)\bar{\zeta}_B/2$ is increasing and unbounded in n; formation therefore fails for all n above a finite $\bar{n}(\bar{\zeta}_B)$, nonincreasing in $\bar{\zeta}_B$. (iv) Common downstream use or port lowers $\bar{\zeta}_B$ through the pairwise cost reductions A_z, A_h, which weakly raises $\bar{n}$ and enlarges the set of pooled sizes for which formation is profitable. □

Proof of Proposition 3

Proof. Let $W^{in}(x), W^{out}(x)$ be continuation values of entering the season inside and outside a group (membership persists while participation continues; re-entry requires paying ζ again). Joining from outside in state θ_t yields $\theta_t x(V - x) - F + \delta \mathbb{E} W^{in}$; staying out yields $\delta \mathbb{E} W^{out}$. Remaining inside yields $\theta_t x(V - x) - F^m + \delta \mathbb{E} W^{in}$; exiting yields $\delta \mathbb{E} W^{out}$. Since $F = F^m + \zeta > F^m$ and the option-value term $\delta \mathbb{E}[W^{in} - W^{out}] \ge 0$ is common to both comparisons, every state satisfying the entry inequality satisfies the continuation inequality strictly. For the permanent-zero-prize statement, write $\tilde{X} \equiv x(V - x)$ and $\theta = s - c\eta$ constant. If incumbency is weakly profitable, $W^{in} = (\theta \tilde{X} - F^m)/(1 - \delta)$; incumbents remain iff $\theta \tilde{X} \ge F^m$. Entry pays iff $\theta \tilde{X} - F + \delta(\theta \tilde{X} - F^m)/(1 - \delta) \ge 0$, i.e., $\theta \tilde{X} \ge (1 - \delta)F + \delta F^m = F - \delta\zeta$. Since $F - \delta\zeta - F^m = (1 - \delta)\zeta > 0$, the band is nonempty with the stated width. □

Proof of Proposition A10

Proof. Under the stated null there is no organized shared service, so every multi-importer manifest is assembled by a consolidator filling capacity from the cell's queue of lots. Exchangeable queueing states directly that the induced distribution over feasible assignments is uniform; random reallocation

of lots to feasible manifests reproduces this distribution, hence the permutation distribution of any assignment statistic is its exact null distribution. When only identities (not sizes) are exchangeable, the same argument applies within the realized size-to-manifest allocation, yielding conditional exactness. Under a positive prize, Proposition 1(i) places lots with sizes in (x_-, x_+) on organized early vessels, accepted before T^* by construction, raising their mutual co-occurrence relative to exchangeable assignment; Proposition 1(ii) removes internalized residuals of above-threshold firms from the consolidation pool, lowering their co-occurrence with all other lots. Both departures scale with the mass of lots satisfying the respective participation conditions, which increases in $\beta r \rho \eta$ and vanishes when $r = 0$ or $\rho = 0$. □

Proof of Proposition 1(i), combined-size statement A coalition pooling its cargo into one lot X_B commits it toward the dispatch load and advances it by $\eta(V - X_B)$ relative to consolidating the pooled block, so its joint payoff is $\theta X_B(V - X_B) - F_B$, the single-lot objective evaluated at (X_B, F_B); the cutoffs and the interior peak at $V/2$ follow. For $X_B > V$ the pooled cargo exceeds one dispatch load and the members with whole loads face the internalization margin of part (ii). □

Proof of Proposition 2 Partition the year by the fill state. (i) At the opening, the unadvanced arrival $d_W(x) + \varepsilon$ falls before T^* with probability close to one for all but the latest lots, so $\Delta p \approx 0$ for most cargo and the sub-threshold pooling motive is near its ordinary level; firms with $n_i \geq 1$ commit charters at $t = 0$ by assumption, so their residual-internalization choice (Proposition 1(ii)) is made at the opening and their shared-service participation is lowest then. (ii)–(iii) As cumulative arrivals approach $\overline{Q}$, the mass of lots for which the unadvanced arrival straddles T^* rises; for those lots Δp approaches $\Gamma(T^* - 0) - 0$, its maximum, so the return to pooling rises with the fill state and peaks as exhaustion nears. (iv) For $t > T^*$ the wedge is zero and the participation condition reverts to $(s - c\eta)x(V - x) \geq F$, the ordinary level θ_0. A unit due at $t > T^*$ waits for the reset iff the saving $r p_E^+$ exceeds waiting cost $c_w(1 - t)$, i.e. iff $t > t_w = \max\{T^*, 1 - r p_E^+ / c_w\}$; the width $1 - t_w$ increases in r and is zero at $r = 0$. □

B Data appendix

Sources and vintages. The declaration universe comes from DIAN extracts by product. The yellow-corn base is a slim-schema extract covering 2001–2023, augmented for 2011–2012 with a full-schema vintage (the slim base carries no rows for those years; the full-schema vintage carries 1,776 and 2,365 declaration lines, respectively, with their own manifest dates and declaration types); manifest dates and anticipated-declaration flags for the slim base merge from a supplementary vintage on (form, year); the comparison products come from separate grain extracts that include 2011–2012 natively. Acceptance timestamps to the second come from a supplementary timestamp

file, 2013–2021. Quota quantities and scheduled rates are transcribed from the TPA tariff-elimination schedules and implementing decrees; the world maize price is the World Bank Pink Sheet monthly series, matched at acceptance month, with a two-month-lag variant for fixture timing; the AIS voyage sample is a purchased VesselFinder extract (209 matched U.S.-route voyages, 2012–2019). The yellow-corn source is truncated at 2023-06-30, so late-2022/2023 seasonal outcomes are coded missing, not zero.

Fields and harmonization. A declaration line records the importer's tax identifier (NIT), the cargo-manifest number, the declaration type (in particular *anticipated* declarations, filed before the vessel arrives), the exporter of record, product code, origin and port codes, net weight, FOB and customs values, the statutory tariff, and the tariff paid. NITs are harmonized by stripping the DIAN verification digit where present and mapped into ownership groups with the project concordance (36,159 lines match directly, 4,315 after digit-stripping, 2,100 retained under their own NIT); exporter names are normalized (case, punctuation, legal-form suffixes) into exporter groups.

Manifest totals and the arrival date. One manifest carries Form-500 declarations from many importers, accepted on different dates—sometimes weeks apart, occasionally across a year boundary—so manifest tonnage always sums every declaration line sharing the manifest number, across all acceptance dates, with no date, year, origin, or port splits. A manifest's arrival date is the recorded manifest date where available (89% of yellow-corn manifests); otherwise the earliest acceptance among non-anticipated declarations (anticipated acceptances precede arrival and would date it too early); otherwise the earliest acceptance. Dates outside 2000–2024 are treated as missing. Year and 14-day window assignments use the arrival date; quota accounting always uses acceptance timestamps. Manifest identity does not condition on customs office, which is innocuous: only 10 of the yellow-corn manifest IDs (0.2% of tonnage) span more than one office, and the randomization cells condition on each manifest's tonnage-dominant port. The AIS validation of both clocks and of the manifest–voyage correspondence is Table A4.

Counts and audit. The clean yellow-corn panel: 42,574 declaration lines (of 44,757 raw; lines drop for missing manifest, acceptance date, origin, port, or nonpositive tonnage) on 6,979 manifests, forming 20,087 lots; 27% of lines are anticipated declarations. A per-product audit (line counts, match rates, arrival-date sources) accompanies the data construction. The replication materials regenerate every number in the paper from the committed data extracts: all randomization is seeded, and every table and figure is written directly by the estimation programs.

Dispatch threshold and firm size. $V = 16{,}280$ t is the median tonnage of post-TPA (2013+) U.S. solo bulk yellow-corn manifests (solo = one ownership group; bulk = ≥5,000 t; records under 100 t excluded as container/LCL residuals). Alternatives bracket it: 5,607 t (median over all solo manifests) and 17,257 t (p75). Firm size is the group's pre-TPA (2004–2010) mean monthly tonnage over V; groups are never classified with post-agreement data, and post-TPA entrants receive a post-entry proxy only where noted.

B.1 Cutoff and reset validation detail

Two further discontinuities qualify how the cutoff may be read. Declaration size falls discontinuously at the cutoff (−1.6 log points): once the quota is exhausted, the large racing lots vanish and the residual queue is small consignments (Figure A10). This is selection, not a causal effect of losing access—the race itself removes the big lots before the cutoff. Consistent with composition rather than lost trade, a daily-mass outcome that retains zero-import days shows no comparable discontinuity.

The January reset closes the annual loop the model turns on: at the year boundary, in the same design, the effective tariff proportion falls by 0.997 (s.e. 0.004)—complete restoration of in-quota treatment—the tariff rate falls by 14.7 points, and declaration size jumps by 0.78 log points as bulk racing lots return. The RD estimates are robust where they should be and fragile where the design says they must be: insensitive to bandwidth, essentially unchanged when anticipated declarations are excluded (jumps of 0.631 and 9.3 points), and mechanically attenuated only in the no-donut specification that mixes regimes inside the carryover band (Table 1).

C Limitations, data decisions, and model boundaries

Three genuine limitations bound what the tests establish. First, *power*: the pairwise co-occurrence design is informative only for yellow corn. White corn and crude soybean oil, the other binding FCFS products, have no above-threshold co-shipping pairs in their race windows, so their evidence is confined to the scale and timing margins; the auctioned-product falsification of the bloc detector is in Appendix Table A18. Second, *incidence at low wedges*: the transaction-price split is measured credibly only in the high-wedge years (Section 5.7); from 2018 on the wedge is small relative to price noise, and the post-exhaustion gaps of +1 to +7 USD/t—which the paper does not model—remain unexplained measurement facts. Third, *the recomposition margin*: the pre-period falsification (Appendix Table A15) shows the same/different-exporter recomposition drifting before the TPA at roughly half the treated magnitude, so that margin is read jointly with its binding-year specificity rather than on its own; the reweighting that holds the monthly exporter composition at its pre-TPA distribution (Appendix Table A16): the margin survives at the parent-trader level and under a leave-out of the largest traders, fails at the finest entity level, and Section 5.2 states the joint reading.

Data decisions a careful reader should know: manifest totals sum all declaration lines of a manifest across acceptance dates, and manifests are dated by the recorded manifest date where available and by the earliest non-anticipated acceptance otherwise (Section 2.2)—a handful of pre-2011 manifest IDs with year-long acceptance spans look like ledger artifacts rather than voyages and are retained as recorded (excluding the 18 with spans over 180 days leaves the pre-period median at 5.2 versus 5.3 kt and the ≥10 kt share at 0.30, while mean groups per manifest falls from 3.0 to 2.8); manifest identity does not condition on customs office, which is innocuous: after harmonizing

DIAN's two office-coding vintages (letter and numeric codes, with Barranquilla and Cartagena each recoded once), only 26 of the 6,979 yellow-corn manifest IDs (0.7% of tonnage) record declarations at more than one office, 23 of them before 2013 and concentrated in the same truncated-ID family as the ledger artifacts above—in the timestamp era the count is three, all under 9,000 t—and the randomization cells still condition on each manifest's dominant port; and exporter-of-record strings carried sixty-seven residual name variants after normalization (legal-form suffixes, conjunctions, typos), merged for the exporter decomposition by the harmonization map in the replication package, with the parent-trader grouping layered on top for the reweighting check.

A few further caveats round out the list. The dispatch threshold V is estimated from post-TPA solo bulk manifests and is therefore an equilibrium object—the size bands it scales use only predetermined pre-TPA firm sizes, and the alternative thresholds reported alongside it (5,607 t all-solo median; 17,257 t p75) bracket the estimate. Re-estimating the phase profile with V moved 20% in either direction, the opening-phase avoidance of the largest pairs and the depletion-phase same-exporter margin are unchanged, while the *net*-pooling depletion peak survives raising V but dilutes to zero at $0.8V$, where the dispatch-scale group absorbs pairs that jointly cannot fill a vessel—consistent with an effect local to the threshold. The full battery under alternative V is Appendix Table A17: the sign-reversal contrast and the top-pair race deficit are robust at $V \pm 20\%$, the bloc universe and the Agricol club are unchanged there, and only the all-solo threshold (5,607 t)—which erases the sub/above-threshold distinction by construction—scrambles the profile; the yellow-corn source truncation at 2023-06-30 (late-2022/2023 outcomes coded missing); the 2014 tonnage-based crossing preceding the treatment-based timestamp by 7 days (carryover accounting; the treatment-based timestamp is used, including in the enforcement RD); entrants entering the bloc universe only through the post-entry size proxy, without which the club's youngest member would be invisible by construction; and the 2022 customs-value anomaly, excluded from incidence interpretation.

What the model leaves outside. Three margins are held fixed that a full market equilibrium would determine, and the choice is intentional. First, the *exhaustion date*: in aggregate, racing advances arrivals and so moves T^* earlier—an equilibrium feedback that can amplify reorganization, and, when early-arrival density is decreasing, make racing self-reinforcing, so that a coordinated race and a quiet season can coexist at the same primitives. The empirical tests condition on the realized fill path, so nothing in Sections 4–5 rests on this feedback; its analysis belongs to the study of FCFS's market-level effects. Second, *prices*: the split β of the wedge between importer and exporter is measured rather than derived from bargaining, and commodity prices do not respond to the race by assumption—an assumption the transaction-price evidence supports (Section 5.7). Third, *market structure downstream*: whether inherited coordination among competitors affects conduct in the domestic market is outside the model. One aggregate implication we do use is bounded and mechanical: shared service operates only if the interval's cargo can fill a vessel,

so policy changes that expand the interval can make organized service appear discretely rather than smoothly—a threshold effect consistent with the two-year completion of the reorganization in Appendix Figure A1, though we do not lean on it for attribution.

D The randomization benchmark: construction

Cells and the opportunity set. A cell is a product × origin × arrival port × non-overlapping 14-day arrival window. Within a cell the objects reallocated are lots (one ownership group's cargo on one manifest); the opportunity set held fixed is the set of manifests present in the cell together with the tonnage each actually carried. Conditioning on the cell fixes which cargo *could* physically have shared a voyage—same product, origin, port, and two-week arrival window—without conditioning on the outcomes under test (who shipped with whom, through which exporter).

Slot null. The slot null permutes manifest labels among the lots within a cell. It preserves each manifest's lot count and the set of lots present, but not tonnage feasibility: a permutation can load a manifest beyond what its vessel carried. It is the looser of the two nulls and isolates pure label exchangeability.

Capacity-feasible null (swap-MCMC). The capacity-feasible null restricts the support to assignments in which every manifest's total tonnage stays within a tolerance $\varepsilon = 5\%$ of its observed total, and samples that support *uniformly*. Because the feasible set is defined by the tonnage constraints and cannot be enumerated, we sample it with a Metropolis chain on lot-to-manifest assignments: a proposal picks two lots in a cell and swaps their manifests; the swap is accepted if and only if both affected manifests remain within tolerance. The proposal is symmetric and the acceptance rule is a flat indicator of feasibility, so the chain's stationary distribution is uniform on the connected feasible set. One persistent chain is run per sample, with 50 sweeps of burn-in from the observed allocation and a draw taken every 10–15 sweeps, so successive draws are far from the data rather than shallow restarts; a sweep is one proposal per lot per cell. Mixing is monitored by the acceptance rate ($\approx 16\%$) and by the displacement—the share of lots off their observed manifest—which rises during burn-in and then plateaus (near 0.25 for the yellow-corn panel), confirming that draws no longer track the starting allocation.

Draws and statistics. Deviation statistics use 400–600 slot draws and 100–150 capacity draws; the co-shipping difference-in-differences and its randomization inference use $D = 500$ capacity draws. On each draw the full counterfactual assignment is materialized and the co-occurrence statistic recomputed exactly as for the observed allocation.

E Randomization inference for the co-shipping DiD

The pairwise co-shipping panel is dyadic, and its two-way (importer-a, importer-b) cluster covariance is frequently not positive-definite, so analytic standard errors for the coefficients of equation (5) are undefined in many bands. Following MacKinnon and Webb (2020) we make the coefficient itself the test statistic and take its distribution from the benchmark draws.

Stacked design and recentering. We stack the observed ("real") allocation, indexed Real = 1, over the benchmark, indexed Real = 0, on the same pair-months; the benchmark outcome for a pair-month is that pair-month's co-shipping indicator averaged over the swap draws. Each coefficient β_φ on Real $\times$ Phase$^\varphi$ therefore measures the excess of real over benchmark co-shipping in phase φ relative to the pre-TPA baseline—the recentering that makes β_φ zero-mean under the null.

Frisch–Waugh and the fast null. Partial the fixed effects (importer-a, importer-b, year, calendar month) and Real out of the treatment columns by weighted least squares (weights = the pair's combined import volume), giving residualized treatments $\tilde{t}$ and fixed weight vectors; each β is then a weighted inner product of those vectors with the outcome. Because the weights depend only on the design, each swap draw d yields the full coefficient vector at the cost of one matrix–vector product on the draw's co-shipping vector $y^{(d)}$. We verified that the pooled coefficient so computed reproduces the weighted OLS coefficient to 10^{-8}. The phase interaction and the event study apply the same construction column by column on the phase and year indicators respectively.

p-values and standard errors. The randomization distribution is $\{\beta_d\}_{d=1}^{D}$. The exact upper-tail p-value is Besag–Clifford, $(1 + \#\{\beta_d \geq \beta\})/(1 + D)$, with floor $1/(D + 1)$; no variance is ever inverted. The *randomization standard error* we report—the whiskers in Figures 4 and 6—is the standard deviation of $\{\beta_d\}$, and a coefficient is significant at conventional levels when its ± 1.96 randomization-SE interval excludes zero. The external replication package (Table A13) applies the same logic at the bill-of-lading atom.

F Additional results and robustness

This appendix collects the robustness detail referenced from Section 5.

Common counterparties (T4). Appendix Proposition A6 predicts that co-loading and same-exporter purchasing move together as two faces of one reorganization. The level of the common-exporter rate is uninformative on its own—it rises just as fast in no-quota soybean meal, where there is no race anywhere, because the world's grain traders consolidated over this period—so only the within-year regime contrast can speak. It carries the predicted sign: among co-active pairs in post-period yellow corn, common-exporter purchasing is about eight percentage points higher in race windows than after exhaustion (0.771 against 0.695; Table A12). This margin is secondary to

the T7 decomposition in the main text.

Size profile (T2). Re-estimating equation (4) with the dispatch threshold moved 20% up leaves the $[0.35, 0.70)$ multi-share coefficient at -0.093 ($p < .001$); moving it 20% down attenuates it to -0.063 ($p = .34$) as the bin boundaries reshuffle groups—robust to where vessel scale is placed, not to erasing the distinction. A Benjamini–Hochberg adjustment across all twelve coefficients of Table 2 leaves the $[0.35, 0.70)$ multi-share effect as the sole survivor at $q = .05$. The full battery under alternative thresholds is Table A17.

Benchmark deviations (T3): inference, dose, fleet. Pooling the randomization draws across years, the race-window top deficit has $z = -10.3$ (post-exhaustion -7.1; no-race -0.5), while the mid excess pools to $z = 1.5$. The dose design regresses each race year's top-pair excess on the scheduled out-of-quota wedge: -1.75 pairs per percentage point (HC1 s.e. 0.38; exact permutation $p = .005$ over all 8! reassignments of the wedge vector, $n = 8$), against no dose response for the mid excess (+0.08, exact $p = .92$) and a right-signed but imprecise dose on the unbundling margin's solo near-V manifest share (+0.003 per pp, exact $p = .25$). The placebo dose on no-quota soybean meal is -0.91 per pp (exact $p = .033$)—the schedule declines monotonically with time, so the wedge partly proxies a common bulk-market trend, and the conservative yellow-corn reading is the differential, -0.84 pairs per pp (Table A21). The common-support trim restricts both regimes to the p10–p90 overlap of manifest tonnage within each year (the full fleets differ: 1,166 race-window manifests at a median 15.5 kt against 849 post-exhaustion at 12.2 kt); the top deficit remains large and significant in both regimes (-85.0 race, -78.8 post-exhaustion, both $p < .005$) while the mid excess keeps its race-only sign (+7.7 vs. -6.7; Table A11).

Within-year profile (T7): falsification detail. Under scheduled-phase assignment the dispatch-scale net profile is flat and insignificant (+0.022, +0.011, +0.007 across opening, stockpiling, depletion, net of null-draw means) and the largest pairs' front-loaded avoidance attenuates (-0.011, n.s.); the same-exporter margin is instead elevated in *every* phase of binding years and in neither slack year—a binding-year-level phenomenon rather than a within-year one. In the 2004–2011 pseudo-treatment design (May 2008), all coefficients are judged against null-draw means because pre-period co-shipping trended upward mechanically: dispatch-scale net co-shipping shows no positive pre-period effect (-0.018 pooled, -0.004 in pseudo-depletion), while same-exporter co-shipping was already drifting above the benchmark (+0.029 pooled, +0.056 in late-year months) and different-exporter below it (-0.05 to -0.10), at roughly half the treated magnitudes. The drift is not an artifact of a shrinking exporter set (active exporters per month rose from 6.3 pre-TPA to 9.1 in binding years; the exporter volume HHI fell from 0.13 to 0.11), and a joint randomization test confirms the pre-period profile is not flat ($p = .002$)—which is why the identifying contrast rests on

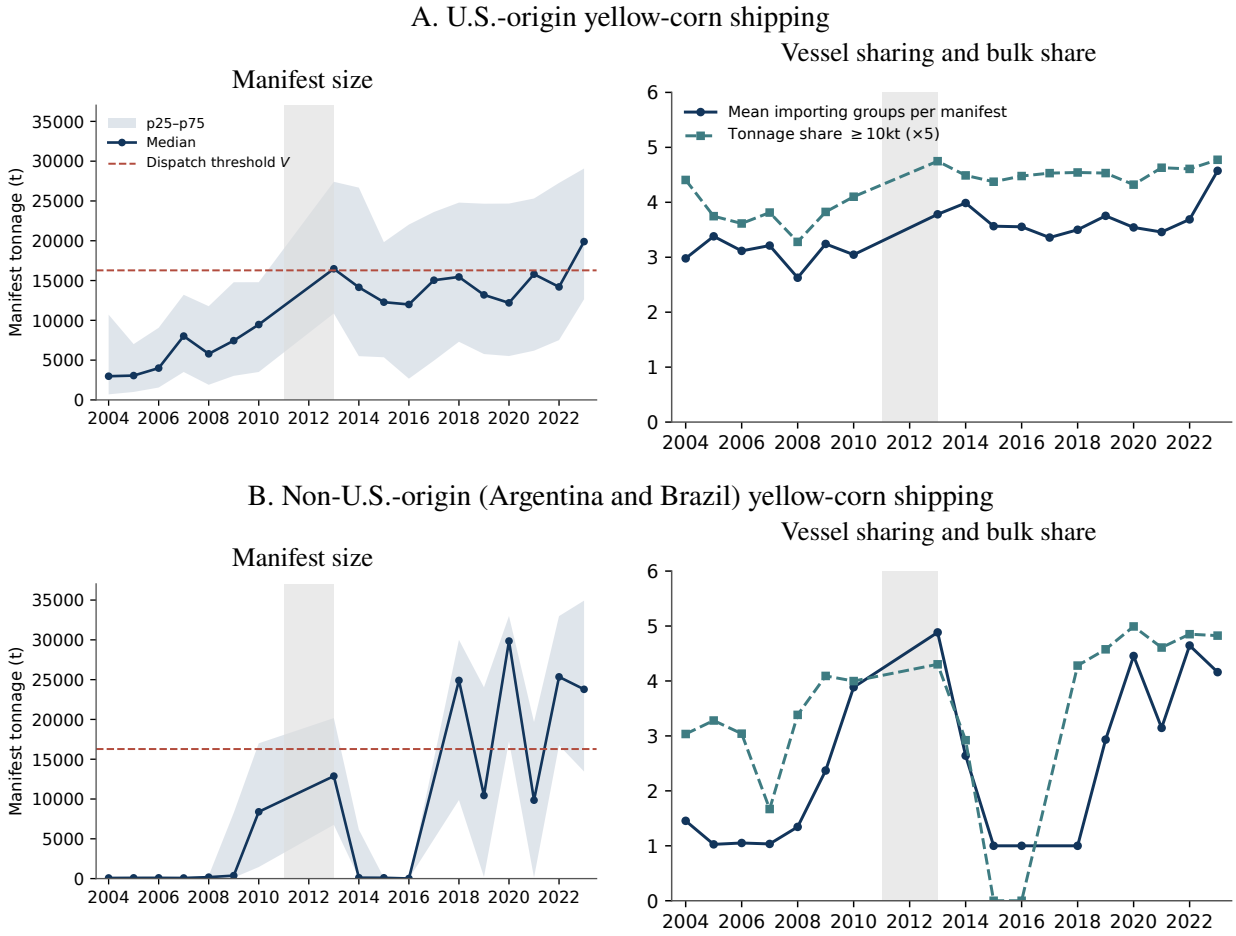
the discrete binding-year step above every pre-period value and its absence in slack years, never on pre-period flatness.

Within-year profile (T7): reweighting detail. Exporter strings carry sixty-seven residual name variants after normalization, merged under a conservative fuzzy harmonization that ships with the replication package; the decomposition is estimated at both the harmonized-entity and the parent-trader level because the platform operates at the level of the trading house. Weights are the ratio of the exporter's pre-TPA volume share to its share in the event's year, capped at five, zero for exporters with no pre-TPA sales. One caution: the composition-fixed margin is also elevated in the slack cell—a year split shows this is entirely 2013 (2012 is too thin to estimate) and largely a weight artifact: under a cap of 2 the slack coefficient falls to +0.023 ($p = .08$) while every binding-phase coefficient stays at the randomization floor.

Within-year profile (T7): further checks. Aggregating the panel to pair-quarters leaves the recomposition at the randomization floor in every band (dispatch-scale same-exporter +0.064) with the net dispatch-scale coefficient right-signed (+0.009), so selection into monthly co-activity is not the margin. These checks come from a robustness battery that re-implements the design end-to-end; its volume-weighted pooled dispatch-scale net coefficient, +0.015, sits within half a randomization standard error of the main implementation's +0.012—cross-implementation agreement on the same ≤2019 sample—and its unweighted variant, +0.014, is qualitatively identical to both. The external replication at the bill-of-lading atom is Table A14.

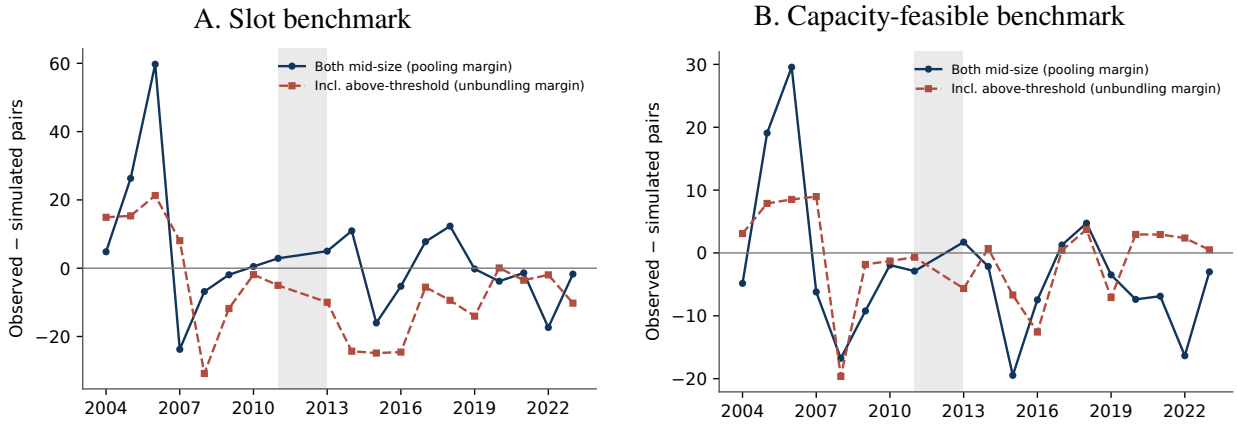
G Appendix tables and figures

Figure A1: The reorganization of yellow-corn shipping: U.S. vs. non-U.S. origin



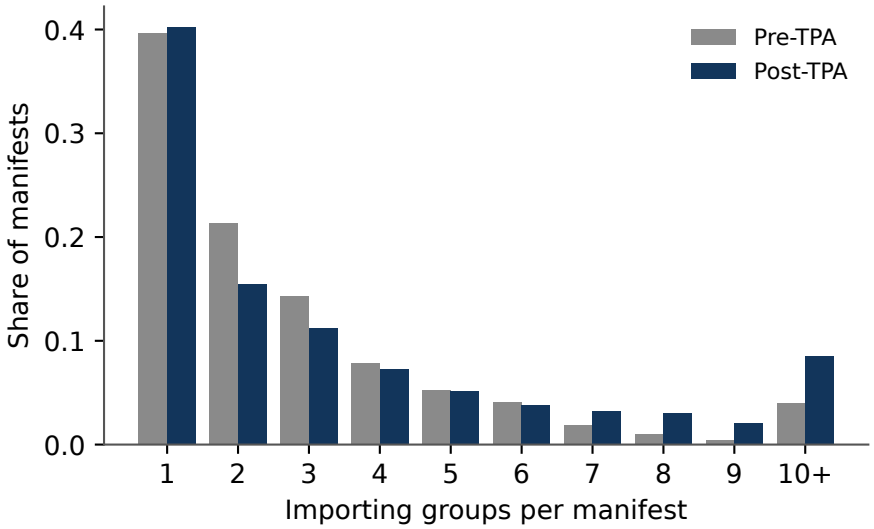
Notes: One point per year; manifests aggregate all declaration lines with a common manifest number across acceptance dates, dated by arrival. Left panels: median manifest tonnage with p25–p75 band; dashed line marks the dispatch threshold V = 16,280 t. Right panels: mean importing ownership groups per manifest and the tonnage share on ≥10 kt manifests, rescaled (×5) to share the axis. Axes are shared across the two origins. The shaded band marks the 2011–2012 transition (TPA in force May 2012); the transition years are plotted, but U.S.-origin volumes there are thin (71 manifests in 2011, 13 in 2012) because pre-TPA corn was overwhelmingly Mercosur-sourced—the same years carry 422 and 471 non-U.S. manifests.

Figure A2: Benchmark deviations by year and pair category



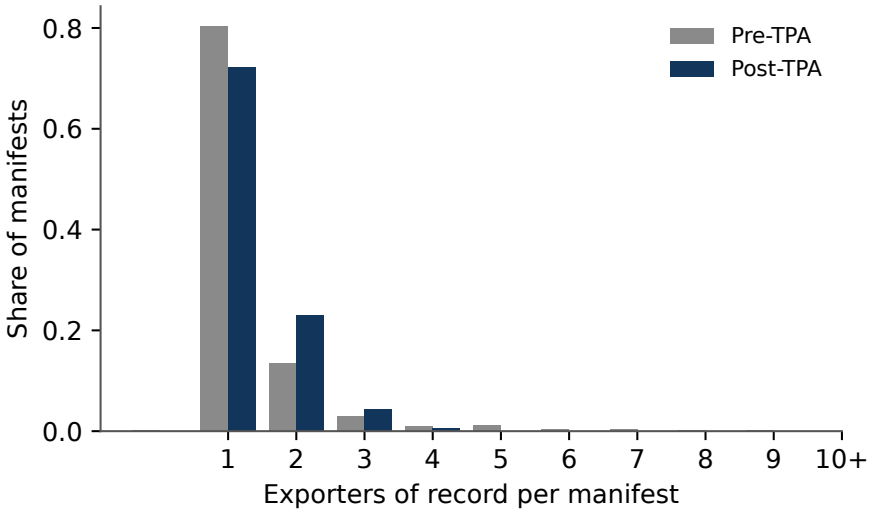
Notes: Observed minus simulated distinct co-occurring group pairs per year, U.S. yellow corn, randomized reassignment of lots to manifests within year × port × 14-day cells (slot: label permutation, 600 draws; capacity: uniform swap-MCMC over tonnage-feasible allocations within 5% of manifest totals, 150 draws). The two theory-relevant pair categories are shown; the all-pairs series is in Table A9. Descriptive complement to the pooled tests of Section 5.2.

Figure A3: Importing groups per manifest, pre- vs. post-TPA



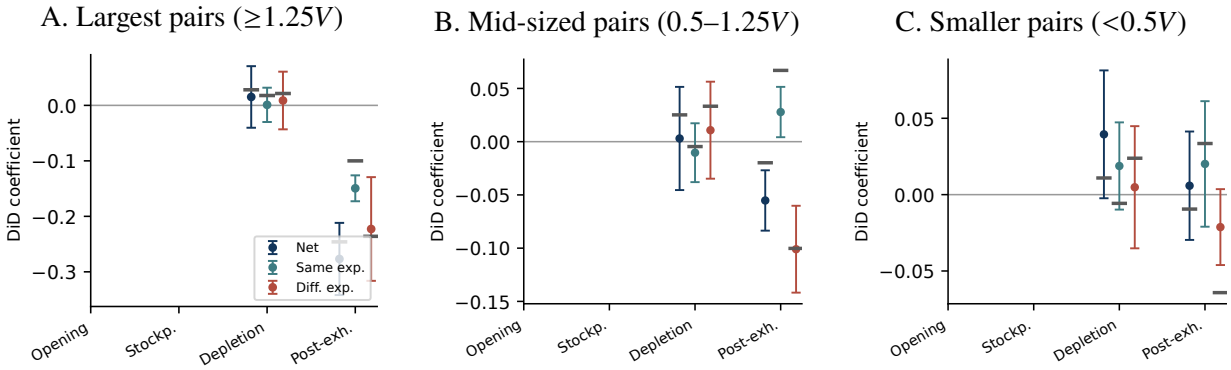
Notes: Distribution of the number of distinct importing ownership groups per manifest, U.S.-origin yellow corn; shares of manifests, counts capped at 10+. Pre-TPA: 2004–April 2012 (987 manifests); post-TPA: May 2012–2023 (2,393 manifests). The post-TPA distribution shifts right: shared vessels carry more importers each.

Figure A4: Exporters per manifest, pre- vs. post-TPA



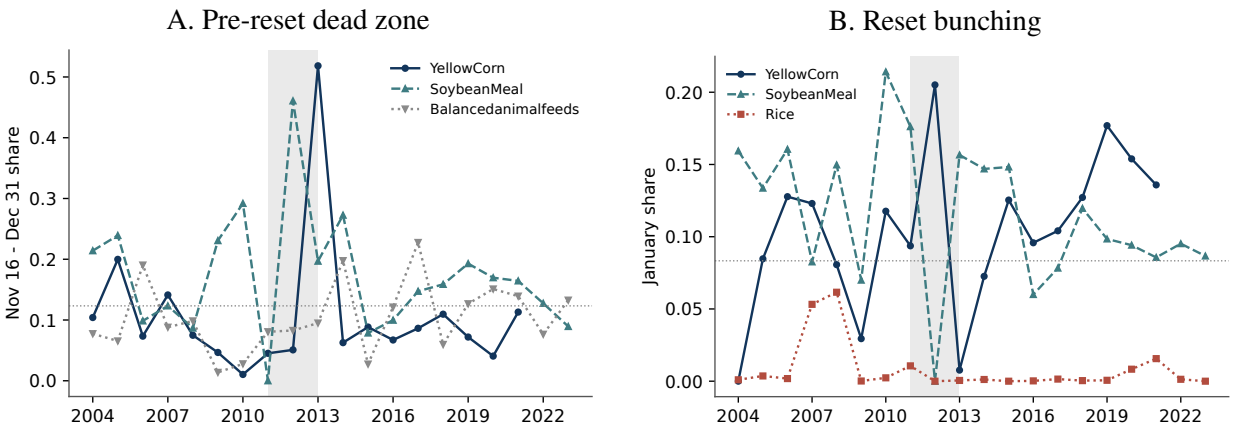
Notes: Distribution of the number of distinct exporters of record per manifest, same sample and construction as Figure A3. Single-exporter voyages carry 61% of post-TPA U.S. volume (47% of shared, multi-importer cargo): even as vessels came to carry more importers, the cargo increasingly travels under one exporter—the platform structure behind the same-exporter margin of Figure 5.

Figure A5: The co-shipping DiD on non-U.S. routes: spillover, not race



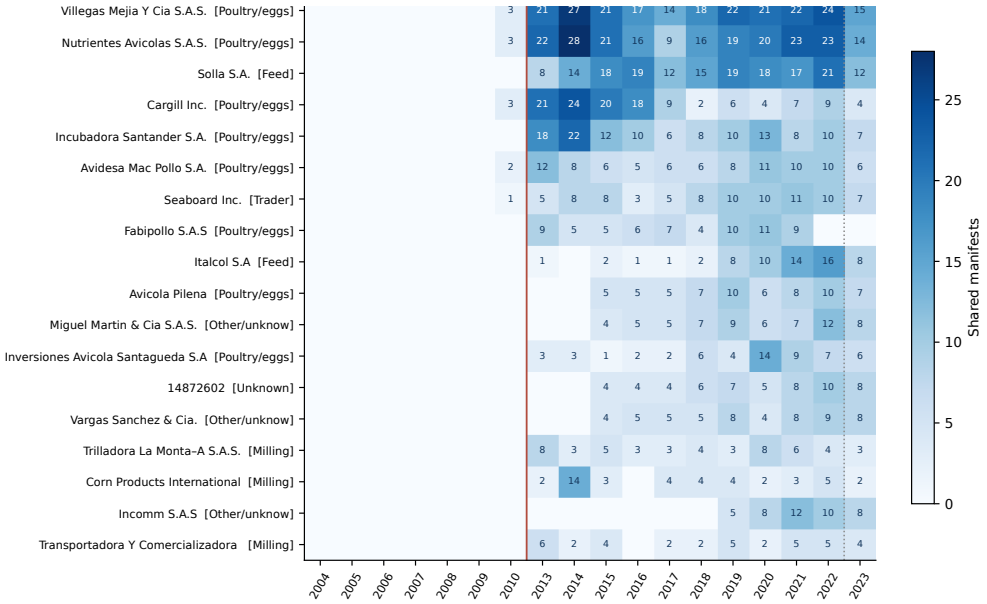
Notes: Equation (5) estimated on non-U.S.-origin (Argentina and Brazil) yellow corn, same size groups, benchmark, weights, and inference as Figure 4; phases follow the U.S. quota's fill calendar, and are thin early in binding years because non-U.S. arrivals concentrate after the U.S. quota is spent. No first-come right attaches to these routes, so this is a spillover test: same-exporter co-shipping is elevated among smaller pairs—the organization built to race the U.S. quota redeployed on Mercosur purchases—but every race-specific margin is absent: no different-exporter withdrawal, no dispatch-scale recomposition, no depletion-phase peak.

Figure A6: The dead zone and the reset



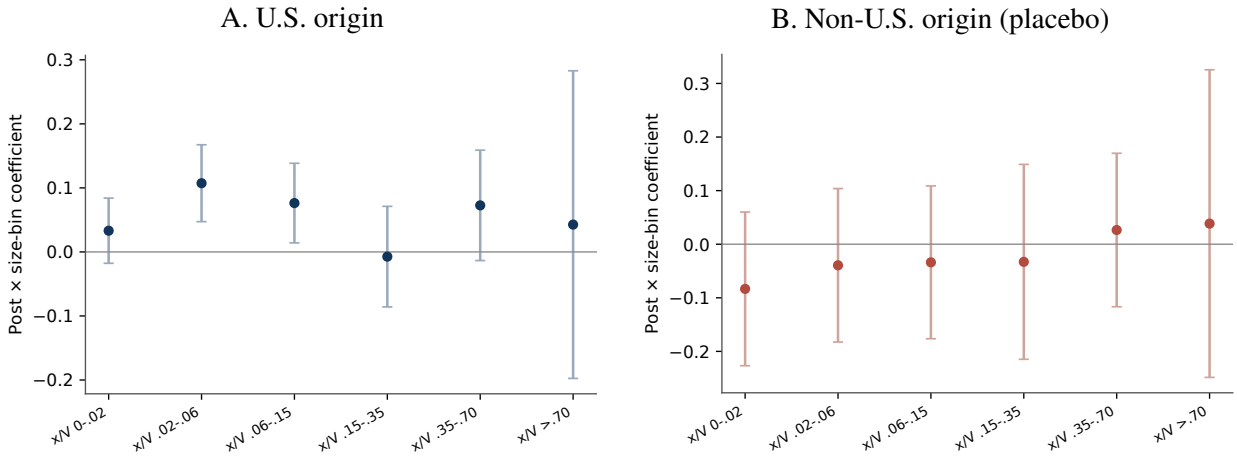
Notes: Panel A: Nov 16–Dec 31 share of annual U.S. tonnage by product-year (dotted line: uniform-arrival share). Panel B: January share (dotted line: 1/12). Acceptance-dated; yellow-corn 2022–23 late shares coded missing (source truncation). Shaded band: 2011–12 (Table A22).

Figure A7: The Agricol club in the manifests



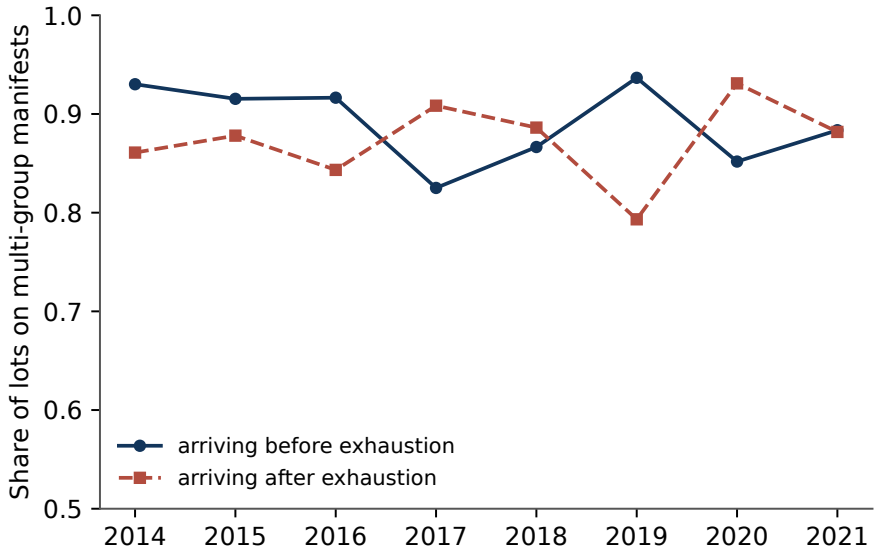
Notes: Manifests shared with Agricol per year, top 18 partners, with downstream sectors in brackets. Solid vertical line: TPA post-period begins; dotted line: 2023 unlimited quota. The interview-verified club members recur among the top partners every race year; large traders and feed firms also appear, because organized vessels carry many groups—which is why detection requires the benchmark-adjusted edge statistics of Table A24.

Figure A8: Pre/post participation change by size, U.S. and non-U.S. origin



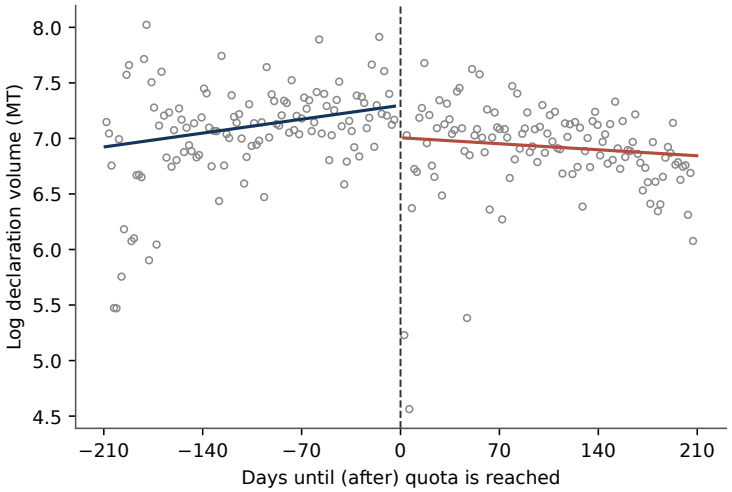
Notes: Post × size-bin coefficients on the multi-group share by pre-TPA size bin, group-year level; OLS with group and year fixed effects; bars are 95% CIs, SEs clustered by ownership group; 2011–2012 are on the pre side of the Post indicator. This is the *across-period* participation margin: in U.S. cargo (Panel A) it is positive at nearly all sizes—the equilibrium switch raised sharing broadly—so it does not display the sign-reversing profile, which is a *within-year* prize margin (race vs. post-exhaustion; Figure 3). In non-U.S. cargo (Panel B), to which no prize attaches, the across-period profile is flat at every size.

Figure A9: Lot-level co-loading before vs. after realized exhaustion



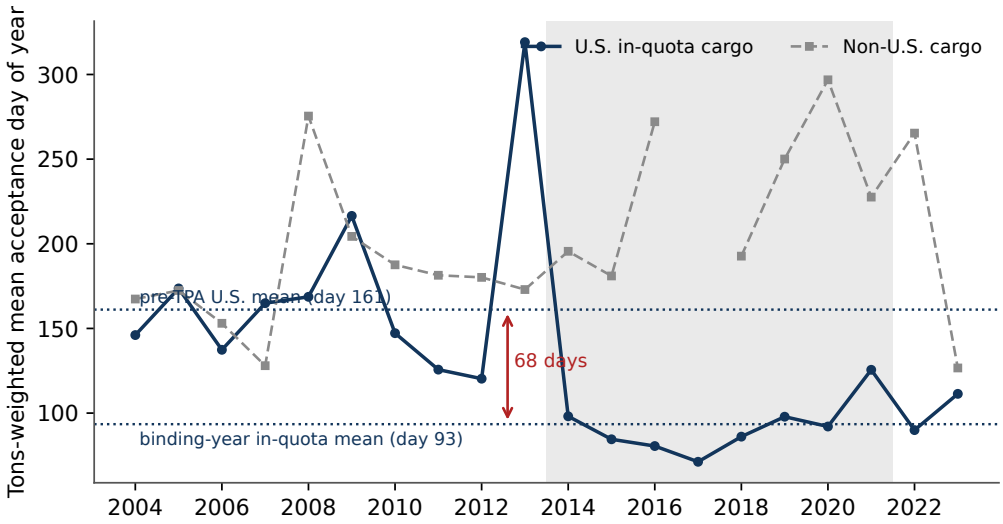
Notes: Share of U.S. yellow-corn lots on multi-group manifests, arriving before vs. after the reconstructed quota crossing, by year; years with a crossing only.

Figure A10: Declaration size at the exhaustion cutoff



Notes: Binned means of log declaration volume against days from the exhaustion timestamp (donut spec of Table 1); RD estimate −1.59 (s.e. 0.28).

Figure A11: Construction of the advancement estimate: seasonal center of gravity of yellow-corn cargo



Notes: Tons-weighted mean acceptance day of year, by year: U.S.-origin in-quota cargo (accepted on or before the year's exhaustion date in binding years; all U.S. cargo otherwise) and non-U.S. cargo. Shading marks the binding years (2014–2021). Dotted lines mark the pre-TPA U.S. mean (day 161) and the binding-year in-quota mean (day 93); the pre-TPA all-origin mean is day 174. The within-origin advancement (68 days) holds the harvest calendar fixed by construction; the all-origin gap (81 days) includes the origin-switch composition. Non-U.S. cargo shifts *later* in binding years—it serves the post-exhaustion market—so the untreated origin cannot bound the harvest-calendar shift, and the within-origin comparison is the bound. These are the day counts behind the acceleration entry of Table 4.

All tables referenced in the text follow. Standard errors in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$. No regression pools products: every estimated specification is single-product and labeled with its allocation regime.

Table A1: From theory to tests

The labels T1–T5 and T7 index the paper's tests; each appears in its results paragraph.

Prediction (result)	Distinguishing feature	Test (section)
Prize exists, assigned by arrival order	—	Cutoff ledger (§2.3)
Sign reversal at vessel scale (Prop. 1)	Opposite signs across V, in combined pair size	T2, T3, T7 (§5.2)
Coordination follows the fill cycle (Prop. 2)	Depletion peak; front-loaded avoidance; post-exhaustion reversion	T7 phase profile (Figs. 4–5)
Recomposition through the exporter (Prop. A6)	Same-exporter up, different-exporter down, at all scales	T7 (§5.2); T4 (App. F)
Partner rotation off blocs; homophilous standing blocs (Prop. A8, Cor. A9)	Null pair stability is a prediction	Bloc detector, Agricol validation (§5.3)
Formation responds to prize; dissolution does not (Prop. 3)	Asymmetry itself testable; 2023 reversion of flows only	T1 (§5.4)
Dead zone widens with wedge; January bunching (Prop. 2)	Dose in predetermined schedule	T5 (§5.5)
Cost-shape discipline: single-peaked K, $\theta_0 > 0$ small, deadfreight ϕ	Measured in no-prize states	κ moments (§5.6)
Timing gains zero-sum given the fill path (Lem. A3)	Zero-sum reallocation without market equilibrium	Cost ledger (§6)

Table A2: Products under TPA tariff-rate quotas, by allocation rule and trade value

Product group	Allocation	2012 imports (t)	FOB 2013–2022 (M USD)
Yellow corn	FCFS	3,362,493	9,867
Sorghum	FCFS	628,735	145
Crude soybean oil	FCFS	244,540	2,545
Rice	Auction	119,022	935
White corn	FCFS	69,062	706
Animal feeds	FCFS	68,300	991
Chicken leg quarters	Auction	39,596	–
Dried beans	FCFS	25,591	–
Milk powder	FCFS	22,847	–
Pet foods	FCFS	8,495	–
Glucose	FCFS	4,519	–
Processed dairy products	FCFS	4,312	–
Bovine meat offal	FCFS	2,392	–
Cheese	FCFS	1,460	–
Standard quality beef	FCFS	415	–
Ice cream	FCFS	265	–
Yogurt	FCFS	81	–
Spent fowl	FCFS	72	–
Butter	FCFS	69	–

Notes: The nineteen product groups placed under tariff-rate quotas by the U.S.–Colombia TPA: seventeen allocated first come, first served, two by auction through export trading companies. 2012 import volumes (metric tons) from the project's TRQ schedule compilation. FOB values (million USD, all origins, 2013–2022) computed from the DIAN declaration extracts for the six products digitized for this paper; "–" = product not yet digitized. Yellow corn's trade value is four times the next TRQ product among those digitized—the basis, with its bulk technology and repeatedly binding quota, for its role as the paper's leading case.

Table A3: Exact cutoff ledger

Year	Quota (mt)	Treatment-based	Tonnage-based	Δ (hours)	Fill ratio	TS match
2013	2,205,000	–	–	–	0.23	1.000
2014	2,315,250	2014-06-27 08:28	2014-06-19 23:59	-176.5	1.57	0.972
2015	2,431,013	2015-06-02 16:41	2015-06-02 16:41	0.0	1.75	1.000
2016	2,552,563	2016-05-17 00:04	2016-05-17 00:04	0.0	1.66	1.000
2017	2,680,191	2017-06-12 14:17	2017-06-12 14:17	0.0	1.68	0.999
2018	2,814,201	2018-06-18 14:25	2018-06-18 14:08	-0.3	1.72	1.000
2019	2,954,911	2019-08-18 10:26	2019-08-18 10:26	0.0	1.19	1.000
2020	3,102,656	–	2020-07-11 13:14	–	1.53	1.000
2021	3,257,789	–	2021-11-06 09:38	–	1.19	1.000

Notes: Treatment-based vs. tonnage-based quota-exhaustion timestamps, U.S. yellow corn. Reconstruction: declarations ordered by acceptance timestamp, tons accumulated to the scheduled quota. Δ = tonnage-based minus treatment-based, in hours. The treatment-based timestamp is the acceptance time of the last declaration receiving in-quota tariff treatment; no exhaustion time is published by DIAN. Fill ratio = annual U.S. arrivals over the scheduled quota. TS match = share of declarations matched to a timestamp. 2013 never crosses (fill 0.23). For 2020–2021 no treatment-based timestamp is available, and the tonnage-based reconstruction dates those crossings. The 2014 gap (−176 h) reflects in-quota carryover accounting; the treatment-based timestamp is used wherever available.

Table A4: Ship-tracking (AIS) validation of the shipment objects

Panel A: match and clock validation	
Matched voyages (U.S. route, 2012–2019)	209
Distinct vessels	181
Declaration lines covered	1,596
Physical voyages matching exactly one manifest	100%
|Manifest date − AIS arrival| ≤ 1 day	72%
|Manifest date − AIS arrival| ≤ 3 days	90%
Median gap (days)	0
Lines accepted before physical arrival (anticipated filings)	21%
Median acceptance lead of *Anticipada* declarations (days)	−4
Panel B: acceptance lag after physical arrival, by arrival month	
Jan.–Sep. arrivals (n = 159): median last-acceptance lag (days)	3
Oct. arrivals (n = 9)	13
Nov. arrivals (n = 12)	10
Dec. arrivals (n = 29)	25
Dec. arrivals: share with *first* acceptance in Jan. or later	28%

Notes: Purchased VesselFinder AIS voyage sample matched to DIAN yellow-corn declarations by port, date, and cargo. Panel A validates the paper's data conventions: the manifest is the physical vessel voyage, the recorded manifest date is the AIS arrival date, and anticipated declarations are accepted before the ship arrives (hence their exclusion from arrival dating). Panel B is the physical trace of strategic acceptance timing: December-arriving cargo waits in port while its filings are deferred to the January reset (Section 5.5).

Table A5: Market reorganization by allocation regime

Regime	Product	Period	Manif.	Med. t	Grp/man.	Multi	Man.≥10kt	Tons≥10kt
FCFS binding	Crudesoybeanoil	2004-2010	37	1	1.62	0.30	0.00	0.00
	Crudesoybeanoil	2013-2022	202	2,000	1.40	0.28	0.03	0.15
	Crudesoybeanoil	2023	1	4,996	2.00	1.00	0.00	0.00
	WhiteCorn	2004-2010	195	2,200	2.26	0.46	0.07	0.34
	WhiteCorn	2013-2022	415	6,254	1.79	0.39	0.25	0.50
	WhiteCorn	2023	44	6,875	1.50	0.39	0.34	0.66
	YellowCorn	2004-2010	990	5,345	3.07	0.61	0.31	0.76
	YellowCorn	2013-2022	2,320	13,462	3.60	0.60	0.61	0.90
	YellowCorn	2023	75	19,893	4.57	0.69	0.83	0.95
FCFS slack	Balancedanimalfeeds	2004-2010	1,279	38	1.57	0.30	0.00	0.00
	Balancedanimalfeeds	2013-2022	3,052	19	1.20	0.16	0.00	0.02
	Balancedanimalfeeds	2023	229	19	1.27	0.20	0.00	0.00
FCFS mixed	Sorghum	2004-2010	3	3,139	1.33	0.33	0.33	0.81
	Sorghum	2013-2022	8	8,483	1.00	0.00	0.38	0.77
Auction	Rice	2004-2010	176	4	1.26	0.07	0.02	0.64
	Rice	2013-2022	634	20	1.28	0.11	0.07	0.70
	Rice	2023	60	35	1.33	0.10	0.07	0.96
No quota	Othersoybeans	2013-2022	736	4,400	2.90	0.69	0.11	0.29
	Othersoybeans	2023	77	4,747	2.96	0.69	0.13	0.29
	Otherwheats	2013-2022	603	6,506	2.71	0.60	0.33	0.68
	Otherwheats	2023	37	5,400	3.05	0.68	0.27	0.65
	SoybeanMeal	2004-2010	472	2,631	2.89	0.61	0.03	0.17
	SoybeanMeal	2013-2022	1,353	4,359	3.52	0.64	0.16	0.43
	SoybeanMeal	2023	141	6,600	4.33	0.77	0.35	0.71

Notes: U.S.-origin manifest scale and sharing. One row per product × period. Manifest tonnage sums all declaration lines sharing the manifest number across acceptance dates; manifests are dated and assigned to periods by arrival date (manifest date where available, else earliest non-anticipated acceptance). Median t = median manifest tonnage; Groups/man. = mean importing ownership groups per manifest; Multi share = share of manifests with ≥2 groups; last two columns = share of manifests (count and tonnage-weighted) at ≥10,000 t.

Table A6: The twenty largest yellow-corn importers, 2013–2023

Rank	Firm	Ownership group	Downstream activity	Mt	Share (%)
1	Solla S.A.	Solla	Feed manufacturing	5.93	11.5
2	Contegral	Grupo Bios	Feed manufacturing	5.73	11.1
3	Italcol S.A.	Italcol	Feed manufacturing	5.68	11.0
4	Italcol de Occidente	Italcol	Feed manufacturing	4.13	8.0
5	Alimentos Finca	Grupo Bios	Feed manufacturing	3.60	7.0
6	Ingredion Colombia	Corn Products Intl.	Corn wet-milling	2.92	5.7
7	AVSA (Mac Pollo)	Avidesa Mac Pollo	Poultry	2.30	4.5
8	Pollos El Bucanero	Cargill (2017–)	Poultry	2.11	4.1
9	C.I. ADM Colombia	ADM	Grain trader	1.91	3.7
10	Seaboard Overseas Colombia	Seaboard	Grain trader	1.80	3.5
11	Albateq	Albateq	Feed manufacturing	1.63	3.2
12	Avidesa de Occidente	Avidesa Mac Pollo	Poultry	1.50	2.9
13	Incubadora Santander (Huevos Kikes)	Incubadora Santander	Eggs	1.11	2.2
14	Cooperativa Colanta	Colanta	Dairy (member feed)	0.80	1.6
15	CIPA	Seaboard	Feed manufacturing (plus pork)	0.62	1.2
16	Alimentos Concentrados del Caribe	Olímpica	Poultry, eggs, feed	0.59	1.1
17	Avícola El Madroño	Italcol	Eggs	0.56	1.1
18	Santa Anita Nápoles	Villegas Mejía y Cía	Eggs	0.53	1.0
19	Campollo	Cargill (2018–)	Poultry	0.53	1.0
20	Porcicultores APA	Italcol	Pork and feed	0.53	1.0

Notes: Universe of DIAN yellow-corn import declarations, 2013–2023, all origins (total 51.56 million t); harmonized importer NITs, ownership groups from the project concordance. Shares are of total volume. Eighteen of the twenty are direct consumers of the grain; the two exceptions (ranks 9–10) are the local importing arms of the traders ADM and Seaboard. The top ten firms account for 70.0% of volume; the top ten ownership groups for 84.9%. Agricol (Agrícola Colombiana) ranks 29th at 0.49%. Firm classifications from company websites and filings: `solla.com`, `contegral.co`, `grupobios.co`, `italcol.com`, `ingredion.com`, `macpollo.com`, `albateq.co`, `huevoskikes.com`, `colanta.com`, `cipa.com.co`, `santaanitanapoles.com`, `avicolaelmadrono.com`, `porcicultoresapa.com`; Cargill acquisitions of Bucanero and Campollo from Cargill/PR Newswire releases (2017, 2018). In 2004–2010 the composition was similar (14 of the top 20 were direct consumers), with a larger merchant-importer fringe.

Table A7: The twenty largest exporter entities on U.S.-origin shipments, 2013–2023

Rank	Exporter entity (as recorded)	Type	Mt	Share (%)
1	Seaboard Overseas Ltd.	Grain trader	9.03	23.5
2	Bunge Latin America	Grain trader	7.66	19.9
3	CAI Trading (Cargill)	Grain trader	6.28	16.3
4	ADM Americas	Grain trader	4.14	10.8
5	Zen-Noh Grain	Producer-owned cooperative	2.75	7.2
6	Luzar Trading	Grain trader	2.32	6.0
7	LDC Trading (Louis Dreyfus)	Grain trader	1.16	3.0
8	Interamerican Commodities	Grain trader	1.12	2.9
9	CHS Inc.	Producer-owned cooperative	0.98	2.5
10	Gavilon Grain	Grain trader	0.48	1.2
11	COFCO Resources	Grain trader	0.36	0.9
12	World Grain	Grain trader	0.32	0.8
13	Luzar LLC	Grain trader	0.30	0.8
14	Graneles de Chile	Grain trader	0.21	0.6
15	Interstate Commodities	Grain trader	0.20	0.5
16	Bunge North America	Grain trader	0.16	0.4
17	COFCO Brazil Overseas	Grain trader	0.14	0.4
18	Cargill Inc.	Grain trader	0.11	0.3
19	Cargill America	Grain trader	0.10	0.3
20	FCStone Merchant Services	Grain trader	0.10	0.3

Notes: U.S.-origin yellow-corn shipments 2013–2023 (total 38.48 million t); the exporter entity is the counterparty name recorded by DIAN, with obvious spelling variants normalized. Shares are of U.S.-origin volume. Aggregating entities to parents: Seaboard 9.03, Bunge 7.85 (incl. Agritrade), Cargill 6.50, ADM 4.14, Zen-Noh 2.75, Luzar 2.62, Louis Dreyfus 1.16, Interamerican 1.12, CHS 0.98 million t. Seventeen of the twenty entities are merchant grain traders; the producer-owned cooperatives Zen-Noh and CHS are the closest to origin, and no U.S. farming operation appears. In 2004–2010 the same names dominated both U.S.- and South American-origin shipments (ADM 26.4%, Bunge entities 25.2%, Cargill/CAI 10.0% of U.S.-origin volume; ConAgra Trade Group, sold in 2008 to become Gavilon, 5.8%). Corporate links documented in SEC Exhibit-21 subsidiary filings (`sec.gov`), `panjiva.com` (CAI Trading–Cargill), `zgcusa.com`, `chsinc.com`, `luzartrading.com`, `seaboardcorp.com`, and PR Newswire (Viterra–Gavilon, 2022).

Table A8: Link formation and re-linking: annual rates

	Yellow corn (FCFS binding, race windows)				Soybean meal (no quota)	
Year	New links	At-risk pairs	Formation	Re-link	Formation	Re-link
2004	270	378	0.71	–	0.31	–
2005	112	291	0.38	0.90	0.29	0.55
2006	140	359	0.39	0.80	0.19	0.59
2007	100	413	0.24	0.55	0.16	0.47
2008	31	206	0.15	0.45	0.19	0.63
2009	54	319	0.17	0.34	0.18	0.47
2010	31	154	0.20	0.31	0.00	0.22
2011	2	48	0.04	0.17	0.19	0.41
2012	0	3	0.00	0.00	0.27	0.40
2013	59	174	0.34	0.45	0.17	0.55
2014	146	320	0.46	0.60	0.16	0.51
2015	34	119	0.29	0.51	0.24	0.66
2016	18	81	0.22	0.57	0.25	0.78
2017	6	67	0.09	0.46	0.22	0.78
2018	14	76	0.18	0.49	0.24	0.81
2019	7	65	0.11	0.61	0.07	0.72
2020	13	53	0.25	0.63	0.31	0.72
2021	8	46	0.17	0.58	0.23	0.74
2022	7	39	0.18	0.52	0.28	0.64
2023	4	21	0.19	0.56	0.11	0.80

Notes: Yellow-corn race windows vs.
no-quota soybean meal. Top-50 pre-TPA groups per product, U.S. origin. A link = co-shipment on the same manifest; for yellow corn, counted in race windows of crossing years (all windows otherwise). Formation = new links (none in prior 3 years) over at-risk pairs; Re-link = share of ever-linked active pairs that co-ship again. Soybean meal has no race window; its columns use all windows (products are never pooled in any estimation).

Table A9: Randomization-benchmark co-occurrence

Regime	Category	Observed	Simulated	Excess
Slot permutation				
Race	Mid pairs	893	868.9	+24.1
Race	Top pairs	852	989.3	-137.3
Race	All pairs	2,493	2,578.8	-85.8
Post-exh.	Mid pairs	619	623.7	-4.7
Post-exh.	Top pairs	681	765.4	-84.4
Post-exh.	All pairs	1,854	1,934.4	-80.4
No race	Mid pairs	2,299	2,251.1	+47.9
No race	Top pairs	2,046	2,058.4	-12.4
No race	All pairs	6,385	6,097.8	+287.2
Capacity-feasible swaps (uniform swap-MCMC)				
Race	Mid pairs	893	921.6	-28.6
Race	Top pairs	852	873.5	-21.5
Race	All pairs	2,493	2,566.7	-73.7
Post-exh.	Mid pairs	619	653.2	-34.2
Post-exh.	Top pairs	681	688.3	-7.3
Post-exh.	All pairs	1,854	1,924.7	-70.7
No race	Mid pairs	2,299	2,311.0	-12.0
No race	Top pairs	2,046	2,043.0	+3.0
No race	All pairs	6,385	6,173.2	+211.8

Notes: By regime and pair category, U.S. yellow corn, 2014–2021. Cells: year × port × 14-day arrival window × regime (donut and dead zone excluded). Observed and simulated = distinct co-occurring group pairs summed over years; excess = observed − simulated. Slot: 400 draws permuting manifest labels within cells; capacity: 100 draws from a swap-MCMC sampling tonnage-feasible allocations (every manifest total within 5% of observed) uniformly—symmetric pairwise-swap proposals with burn-in and thinning, so draws are independent of the observed assignment. Mid = both groups $x/V \in [.02, .35)$; Top = at least one group $x/V \geq .35$. "No race" pools the product's own no-race windows (2004–2012, 2013, 2022–23; 2011–2012 restored). Single-product sample; comparison samples (non-U.S. corn, soybean meal) estimated separately.

Table A10: Coordination test across products, by allocation rule

Product	Rule	Pairs	Top excess	Mid excess	All excess
Yellow corn	FCFS, binding	2493	-137.3^{***} (−14%)	$+24.4^{*}$ (+3%)	-86.0^{***} (−3%)
White corn	FCFS, binding	125	—	−0.2 (−0%)	+1.0 (+1%)
Crude soybean oil	FCFS, binding	20	—	+0.5 (+33%)	$+4.1^{*}$ (+26%)
Balanced animal feeds	FCFS, slack	116	—	+2.5 (+12%)	−7.6 (−6%)
Rice	Auction	100	—	+0.8 (+36%)	-11.8^{**} (−11%)
Soybean meal	No quota	2366	-41.4^{***} (−9%)	+0.1 (+0%)	-94.0^{***} (−4%)
Other soybeans	No quota	45	—	—	+0.1 (+0%)

Notes: The main coordination test—observed minus simulated distinct co-occurring ownership-group pairs against the slot null (400 draws)—run on every digitized product, U.S.-origin, pooled over years. "Pairs" is the observed co-occurring-pair count in the window the excess is measured on: the race window for binding FCFS products, all windows pooled otherwise. "Top" = at least one above-threshold group ($x/V \geq .35$); "Mid" = both mid-size; "All" = every pair. "—" means no pair of that category exists (e.g. no above-threshold group co-occurs). Stars from the one-sided randomization p in the direction of the estimate: * $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$. The sign-reversing racing signature (top deficit with mid excess) is present only under binding FCFS yellow corn; no-quota soybean meal reproduces the top deficit alone (scale-driven unbundling), and the auctioned and slack products show neither margin. White corn and crude soybean oil have too few above-threshold co-shipping pairs for the test to have power. Sorghum (negligible U.S. imports) and other wheats (no co-occurring pairs) are omitted.

Table A11: Fixed-fleet (common-support) benchmark

Regime	Category	Observed	Simulated	Excess	p (lower)
Race	Mid pairs	711	703.3	+7.7	0.713
Race	Top pairs	634	719.0	-85.0	0.000
Race	All pairs	1,886	1,940.9	-54.9	0.000
Post-exh.	Mid pairs	565	571.7	-6.7	0.320
Post-exh.	Top pairs	593	671.8	-78.8	0.000
Post-exh.	All pairs	1,651	1,724.6	-73.6	0.000

Notes: U.S. yellow corn, crossing years, both regimes restricted within each year to manifests inside the common support of manifest tonnage (p10–p90 overlap of the race and post-exhaustion fleets; 4,946 lots retained). Slot null: 400 draws permuting manifest labels within year × port × 14-day × regime cells on the restricted sample. p (lower) = share of draws with statistic ≤ observed. Because the trim conditions on the chosen fleet—removing the large solo charters the race created—the matched-fleet race-vs.-post contrast is a lower bound on the race effect.

Table A12: Common-exporter rates

Product	Regime	Period	Common-exporter rate	Pair-years
Soybean meal (no quota)	norace	post	0.813	2,595
Soybean meal (no quota)	norace	pre	0.561	3,692
Yellow corn (FCFS binding)	postexh	post	0.695	1,717
Yellow corn (FCFS binding)	race	post	0.771	1,927

Notes: Common exporter on different manifests, by product and regime. Share of co-active group pairs (top-50 pre-TPA groups, U.S. origin) sourcing from a common exporter on *different* manifests in the same year. Products tabulated separately; the soybean-meal pre→post rise (0.55→0.81) with no race window shows the level tracks trader consolidation, so only the within-year race vs. post-exhaustion contrast (0.771 vs. 0.695) is informative about coordination.

Table A13: Co-shipping DiD under randomization inference

	Top 10	11–35	36–55	56–70	71+
Panel A: any co-shipping					
Real × Post-TPA	-0.0242	0.0049***	0.0013	0.0050***	0.0036***
	[1.000]	[0.002]	[0.132]	[0.002]	[0.002]
Panel B: co-shipping through the same exporter					
Real × Post-TPA	0.0039**	0.0089***	0.0011	0.0022**	0.0037***
	[0.024]	[0.002]	[0.146]	[0.030]	[0.002]
Panel C: co-shipping through different exporters					
Real × Post-TPA	-0.0280	-0.0040	0.0002	0.0028***	-0.0001
	[1.000]	[1.000]	[0.377]	[0.004]	[0.994]
Panel D: same-exporter tie (different manifests)					
Real × Post-TPA	-0.0021	-0.0013	-0.0043	0.0024*	0.0036***
	[0.760]	[0.808]	[1.000]	[0.058]	[0.002]

Notes: Real×Post-TPA DiD coefficient on the pairwise co-shipping probability by importer rank tier (columns), U.S. yellow corn, ≤2019, estimated separately by tier with rank-a, rank-b, year and quarter fixed effects, relative to the simulated physical benchmark (arrival-clocked cells, bill-of-lading parcels, uniform swap-MCMC null). Randomization-inference p-values in brackets: the coefficient itself is recomputed on each of $D = 500$ swap-null draws via Frisch–Waugh, p = Besag–Clifford exact upper tail $(1 + \#\{\beta_d \geq \beta\})/(1 + D)$, floor $1/501 \approx .002$ (MacKinnon and Webb, 2020); no cluster variance is estimated. Negative coefficients are never starred (upper-tail test). * $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$. Estimates from S. Fernández-Chacón's replication package.

Table A14: Co-shipping DiD: internal and independent implementations

		This paper		Fernández-Chacón		
Tier	Signal	β	RI p	β	RI p	Pair-months
Top 10	Any	+0.012	0.092	−0.024	1.000	2,900
Top 10	Same exporter	+0.049	0.002	+0.004	0.024	2,900
Top 10	Different exporters	−0.041	1.000	−0.028	1.000	2,900
11–35	Any	+0.013	0.038	+0.005	0.002	5,199
11–35	Same exporter	+0.040	0.002	+0.009	0.002	5,199
11–35	Different exporters	−0.041	1.000	−0.004	1.000	5,199
36–55	Same exporter	+0.025	0.102	+0.001	0.146	472
56–70	Same exporter	−0.026	1.000	+0.002	0.030	116
71+	Any	−0.038	1.000	+0.004	0.002	4,638
71+	Same exporter	+0.042	0.002	+0.004	0.002	4,638
71+	Different exporters	−0.094	1.000	−0.000	0.994	4,638

Notes: Real×Post-TPA coefficient on pairwise co-shipping, ≤2019, swap-MCMC null, $D = 500$, exact Besag–Clifford upper-tail p (floor $1/501$). "This paper": manifest atoms, arrival clock with 2011–2012 restored, our pre-TPA (2004–2010) tonnage ranks, panel of within-tier rank pairs by arrival month conditional on joint activity, rank-a, rank-b, year and quarter FE, β by Frisch–Waugh (validated against OLS to 10^{-8}). "Fernández-Chacón": an independent implementation by a coauthor of companion work (bill-of-lading parcels, his clock and ranks; Table A13). Magnitudes are not comparable across columns (different atoms and conditioning); signs and exact inference are. The three well-populated tiers agree on every qualitative verdict—same-exporter positive at the p-floor, different-exporter at or below chance, Top-10 recomposition; the thin tiers (472 and 116 pair-months in our panel) are where the implementations diverge.

Table A15: Placebo and falsification variants of the phase design

Combined size	Phase	Net (any)	Same-exporter	Different-exporter
Panel A. Phases assigned by the scheduled fill path (baseline sample)				
$< 0.5V$	Pooled	-0.011^{**} (0.006)	$+0.044^{***}$ (0.004)	-0.069^{***} (0.005)
	Opening	-0.012 (0.008)	$+0.045^{***}$ (0.006)	-0.070^{***} (0.006)
	Stockpiling	$+0.007$ (0.008)	$+0.050^{***}$ (0.006)	-0.054^{***} (0.006)
	Depletion	-0.025^{***} (0.006)	$+0.038^{***}$ (0.005)	-0.082^{***} (0.006)
	Slack	-0.039^{**} (0.015)	$+0.010$ (0.008)	-0.051^{***} (0.012)
0.5–$1.25V$	Pooled	$+0.012$ (0.008)	$+0.047^{***}$ (0.005)	-0.047^{***} (0.007)
	Opening	$+0.022^{*}$ (0.013)	$+0.068^{***}$ (0.008)	-0.048^{***} (0.013)
	Stockpiling	$+0.011$ (0.012)	$+0.041^{***}$ (0.011)	-0.042^{***} (0.009)
	Depletion	$+0.007$ (0.009)	$+0.035^{***}$ (0.006)	-0.052^{***} (0.008)
	Slack	-0.022 (0.016)	$+0.005$ (0.008)	-0.041^{**} (0.017)
$\geq 1.25V$	Pooled	-0.013 (0.010)	$+0.040^{***}$ (0.006)	-0.056^{***} (0.009)
	Opening	-0.011 (0.013)	$+0.052^{***}$ (0.009)	-0.075^{***} (0.012)
	Stockpiling	-0.005 (0.013)	$+0.041^{***}$ (0.008)	-0.044^{***} (0.011)
	Depletion	-0.020^{*} (0.010)	$+0.031^{***}$ (0.007)	-0.053^{***} (0.009)
	Slack	-0.044^{***} (0.013)	-0.024^{***} (0.005)	-0.012 (0.013)
Panel B. Pre-period falsification: 2004–2011 only, pseudo-treatment May 2008				
$< 0.5V$	Pooled	-0.030^{***} (0.007)	$+0.015^{**}$ (0.005)	-0.048^{***} (0.006)
	Opening	-0.032^{**} (0.013)	$+0.020$ (0.012)	-0.051^{***} (0.014)
	Stockpiling	-0.043^{***} (0.009)	-0.012^{*} (0.006)	-0.035^{***} (0.009)
	Depletion	-0.019^{**} (0.010)	$+0.032^{***}$ (0.008)	-0.056^{***} (0.008)
0.5–$1.25V$	Pooled	-0.018^{*} (0.010)	$+0.029^{***}$ (0.009)	-0.053^{***} (0.009)
	Opening	-0.025 (0.019)	$+0.035$ (0.020)	-0.061^{**} (0.025)
	Stockpiling	-0.033^{**} (0.014)	-0.003 (0.013)	-0.035^{***} (0.011)
	Depletion	-0.004 (0.015)	$+0.056^{***}$ (0.013)	-0.067^{***} (0.012)
$\geq 1.25V$	Pooled	-0.038^{***} (0.014)	$+0.033^{***}$ (0.009)	-0.088^{***} (0.014)
	Opening	-0.049^{***} (0.013)	-0.009 (0.009)	-0.047^{***} (0.011)
	Stockpiling	-0.021 (0.015)	$+0.019$ (0.017)	-0.069^{***} (0.020)
	Depletion	-0.045^{***} (0.019)	$+0.048^{***}$ (0.010)	-0.105^{***} (0.018)

Notes: Variants of equation (5), U.S.-origin yellow corn, uniform swap-MCMC benchmark, volume weights, randomization inference with 500 draws. Panel A assigns quota-fill phases by the *scheduled* fill path (quota/12 per month: opening January–March, stockpiling April–July, depletion August–December) instead of the realized path; sample and baseline as in Table 3. Panel B estimates the identical design on 2004–2011 data only, with a pseudo-treatment date of May 2008, pseudo-phases by scheduled thirds, and predetermined sizes from pre-2008 volumes. Because the estimator's exact null distribution need not be centered at zero (the benchmark-mean outcome co-moves with the residualized treatment; in Panel B the shift is large), every entry reports the coefficient *net of its null-draw mean*, with the randomization standard error in parentheses; *, **, ***: two-sided exact randomization $p < .10, .05, .01$. Panel A: the dispatch-scale net profile is flat under scheduled phases and the largest pairs' front-loaded avoidance attenuates, while same-exporter co-shipping is elevated in every phase of binding years and in neither slack year. Panel B: the net margin shows no positive pre-period effect; the same/different recomposition shows pre-existing drift at roughly half the treated magnitude.

Table A16: Same-exporter co-shipping at the pre-TPA exporter composition

	Parent-trader level			Entity level
Phase	Baseline	Composition-fixed	Leave-out top 3	Composition-fixed
Pooled	+0.053*** (0.005)	+0.034*** (0.004)	+0.026*** (0.005)	−0.002 (0.003)
Opening	+0.083*** (0.011)	+0.045*** (0.008)	+0.047*** (0.009)	+0.005 (0.005)
Stockpiling	+0.078*** (0.011)	+0.075*** (0.013)	+0.059*** (0.010)	−0.001 (0.005)
Depletion	+0.047*** (0.016)	+0.018 (0.011)	+0.020 (0.014)	−0.004 (0.004)
Post-exhaustion	+0.033*** (0.006)	+0.013*** (0.004)	+0.005 (0.004)	−0.007** (0.003)
Slack	−0.001 (0.008)	+0.080** (0.031)	−0.010*** (0.004)	+0.037* (0.019)

Notes: Same-exporter co-shipping DiD of equation (5), dispatch-scale pairs (0.5–1.25V), U.S.-origin yellow corn, ≤2019, uniform swap-MCMC benchmark, volume weights, 500 draws; every entry is the coefficient net of its null-draw mean with the randomization standard error in parentheses; *, **, ***: two-sided exact randomization $p < .10, .05, .01$. Exporter identities merge 67 residual name variants (harmonization map in the replication package); the parent-trader grouping further consolidates each trading house's vehicles (e.g. Seaboard, Bunge, Cargill/CAI, ADM, LDC). *Composition-fixed*: each same-exporter co-shipping event is re-weighted by $\rho_{ey} = s_e^{\text{pre}}/s_e^{y}$, the exporter's pre-TPA volume share over its share in the event's year (clipped to $[0, 5]$; zero for exporters with no pre-TPA sales), applied identically to the real allocation and to every null draw, so the test is exact and the outcome is co-shipping at the pre-TPA exporter composition. *Leave-out top 3*: binary same-exporter flag ignoring the three largest binding-year traders. The margin survives composition-fixing at the parent level and the leave-out; it disappears at the finest entity level, where binding-year events run through post-TPA entities (incumbent parents' dedicated vehicles and genuine entrants)—concentration the platform mechanism predicts but a strict reweighting cannot separate from composition change. The slack-year elevation of the composition-fixed series is the caution discussed in the text; split by year it is entirely 2013's (2012 is too thin to estimate), and it falls to +0.023 ($p = .08$) under a weight cap of 2 while the binding-phase coefficients are unchanged. Because the baseline column uses the harmonized identities, its entries differ slightly from Figure 5 and Table 3, which use the raw exporter strings.

Table A17: The battery under alternative dispatch thresholds

	T2 race × $[0.35, 0.70)V$		T3 race-window excess		Bloc detector	
Threshold	Multi share	Solo full	Top pairs	Mid pairs	Recurrent/flagged	Agricol
V (16,280 t)	−0.089*** (0.029)	+0.006* (0.004)	−137.7 [.000]	+24.7 [.068]	379 / 164	7/7
$0.8V$ (13,024 t)	−0.063 (0.066)	+0.005 (0.004)	−162.0 [.000]	+39.4 [.002]	339 / 151	7/7
$1.2V$ (19,536 t)	−0.093*** (0.017)	+0.056 (0.036)	−150.5 [.000]	+32.8 [.022]	380 / 165	7/7
All-solo median (5,607 t)	+0.034 (0.036)	−0.001 (0.011)	−167.9 [.000]	+74.5 [.000]	227 / 111	5/7

Notes: Every component of the battery re-estimated with the dispatch threshold moved to the indicated value: size bins in x/V', the solo-full cut at $0.8V'$, and the bloc universe at $x/V' \leq 0.35$ with the entrant proxy rescaled. *T2*: race-window coefficient for the $[0.35, 0.70)$ bin from equation (4) (multi-group share and solo-full-vessel share; SEs clustered by group; *, **, ***: $p < .10, .05, .01$). *T3*: per-year distinct co-occurring pairs summed over years (the convention of Table A9), observed minus slot-null mean, race windows, top and mid categories; the allocation draws are common across thresholds, so columns differ only in classification; randomization p (lower-tail for top, upper-tail for mid) in brackets. *Bloc detector*: recurrent sub-threshold pairs / pairs flagged at $p < .05$ against the fine null; Agricol column is interview-named members recovered. The sign-reversal contrast, the top-pair deficit, and the club are unchanged at $V \pm 20\%$—the mid excess, if anything, strengthens under the alternative thresholds—while the all-solo threshold erases the sub/above-threshold distinction by construction and scrambles the T2 profile even as the top-pair deficit, driven by the same very large firms under any cut, persists.

Table A18: The bloc detector across allocation rules

Product (rule)	Lots	Recurrent pairs	Flagged $p < .05$	Flag rate	Chance	FDR-surviving
Yellow corn (FCFS, race windows)	4,239	379	164	0.43	19	131
Rice (auction)	649	9	3	0.33	0.5	3
Soybean meal (no quota)	4,071	366	104	0.28	18	80

Notes: The bloc-detection procedure of Section 5.3 applied unchanged (sub-threshold universe against the corn dispatch threshold, entrant proxy, joint-manifest count statistic, fine null permuting manifest labels within year × port × 14-day window, 300 draws, recurrence ≥4 distinct years in 2014–2021) to U.S.-origin lots of each product. Yellow corn uses race windows; the auctioned and no-quota products have no race windows by construction, so all windows in those years enter—the conservative choice. "Chance" is the expected false-positive count at $\alpha = .05$. Joint booking above chance marks organized consolidation generally (soybean meal); what distinguishes the auctioned product is the near-absence of any recurrent sub-threshold pair landscape: per thousand lots, yellow corn and soybean meal have ninety recurrent pairs each, rice fourteen—a six-fold thinner pair structure even relative to its smaller lot universe.

Table A19: The distribution of the retained rent within importers

	Groups	Share of in-quota tonnage			Retained rent 2014–21
		2014–21	2014	2021	(M USD, $\beta = 0.71$–0.96)
Above threshold ($x/V \geq .35$)	6	0.66	0.69	0.69	195–264
Mid-sized (.02–.35)	33	0.28	0.26	0.23	82–110
Small (<.02)	16	0.02	0.03	0.01	6–8
Post-TPA entrants	53	0.05	0.02	0.07	11–15
Memorandum (overlapping) groups:					
Five largest groups	5	0.66	0.69	0.69	195–264
Agricol club (7 firms)	7	0.07	0.07	0.04	21–28
All detected bloc members	49	0.32	0.25	0.30	90–122

Notes: In-quota tonnage and the implied retained rent by predetermined size class, U.S.-origin yellow corn, binding years 2014–2021. A lot's quota year is the year of its first acceptance; the lot is in-quota if that acceptance falls on or before the year's treatment-based crossing date (Table A3)—a day-level, lot-level proxy for the declaration ledger that reproduces the scheduled annual quotas within 1–4%. Rent per ton is the scheduled out-of-quota rate applied to the World Bank annual mean maize price, giving a proxy pool of 414M USD over 2014–2021 against the 434M of equation (8) (the difference is price-timing in the benchmark); shares are unaffected by that timing. Retained rent applies the high-wedge-year incidence range $\beta = 0.71$–0.96 (Section 5.7) uniformly; the buyer-type premia of Table A29 are too close to support class-specific β's. Memorandum rows overlap the size classes (the five largest groups are five of the six above-threshold groups; the sixth, Agro Grain, ships 0.05% of in-quota tonnage, so the five-largest row coincides with the six-group class at the displayed precision; the Agricol club members are small and mid-sized; detected bloc members span classes). Size classes use predetermined pre-TPA sizes; entrants are groups with no pre-TPA activity.

Table A20: The residual-value test: largest pairs' co-shipping by $\xi(V-\xi)$

	Same-exporter		Net co-shipping	
Phase	High $\xi(V-\xi)$	Low $\xi(V-\xi)$	High $\xi(V-\xi)$	Low $\xi(V-\xi)$
Pooled	$+0.030^{***}$ (0.007)	$+0.053^{***}$ (0.010)	-0.021^{*} (0.013)	-0.004 (0.015)
Opening	$+0.041^{***}$ (0.012)	$+0.101^{***}$ (0.017)	-0.016 (0.021)	$+0.014$ (0.023)
Stockpiling	$+0.047^{***}$ (0.012)	$+0.084^{***}$ (0.016)	-0.035^{*} (0.019)	$+0.016$ (0.023)
Depletion	$+0.018$ (0.014)	$+0.041^{**}$ (0.017)	-0.018 (0.019)	-0.024 (0.023)
Post-exhaustion	$+0.025^{***}$ (0.008)	$+0.026^{**}$ (0.010)	-0.015 (0.013)	-0.010 (0.015)
Slack	-0.018^{***} (0.005)	-0.036^{***} (0.008)	-0.052^{***} (0.018)	-0.014 (0.017)

Notes: Equation (5) on pairs with combined size $\geq 1.25V$, split at the pair-universe median of the residual value $\xi(V-\xi)$ of the pair's largest member, where $\xi = x - \lfloor x/V \rfloor V$ from predetermined size; U.S.-origin yellow corn, $\leq$2019, swap-MCMC benchmark, volume weights, 500 draws. Every entry is the coefficient net of its null-draw mean with the randomization standard error in parentheses; *, **, ***: two-sided exact $p < .10, .05, .01$. If the mid-cycle co-shipping of large firms worked through the residual margin (Proposition 1(i) applied to the residual lot), it would concentrate in the high-$\xi(V-\xi)$ half; instead the net stockpiling uptick vanishes against its null in both halves and the same-exporter elevation is present in both—larger in the *low* half—the signature of the platform channel of Proposition A6(ii), which carries no ξ-gradient. The table's underlying file also reports the benchmarked off-manifest common-exporter margin (below null in every binding phase, above it only in slack years), the pair-quarter aggregation, and the unweighted estimates discussed in the text. ξ from monthly mean size is a coarse proxy for the shipment-cycle residual; proxy noise would attenuate a true gradient toward equality but not reverse it.

Table A21: Dose designs under exact permutation inference

Margin (per pp of scheduled wedge)	n	Slope	Exact perm. p
Top-pair excess (race windows, slot null)	8	-1.75	0.005
Mid-pair excess (race windows, slot null)	8	$+0.08$	0.922
Solo $\geq 0.6V$ manifest share (race windows)	8	$+0.003$	0.252
Placebo: soybean-meal top-pair excess (no quota)	8	-0.91	0.033

Notes: Each race year's (2014–2021) statistic regressed on the scheduled out-of-quota tariff for that year, fixed by the agreement. Exact permutation p: the OLS slope is recomputed under all 8! = 40,320 reassignments of the wedge vector across years; p = share of reassignments with $|\hat{\beta}|$ at least the observed. The placebo runs the identical regression on no-quota U.S. soybean meal's top-pair excess (its own windows) against the yellow-corn wedge: the schedule declines monotonically in time, so part of any dose response is a common trend; the yellow-corn−placebo differential is -0.84 pairs per pp.

Table A22: Dead-zone dose-response

	YellowCorn (FCFS binding)	WhiteCorn (FCFS binding)	Crudesoybeanoil (FCFS binding)
Panel A: late-year share (Nov 16 – Dec 31)			
Scheduled out-of-quota tariff (pp)	-0.0009	-0.0050	-0.0033
	(0.0020)	(0.0033)	(0.0053)
Panel B: January share			
Scheduled out-of-quota tariff (pp)	-0.0049***	0.0006	-0.0085
	(0.0018)	(0.0029)	(0.0430)
Pre-TPA mean late share	0.092	0.127	0.333
Binding-years mean late share	0.080	0.067	0.020
Crossing years (obs.)	8	9	4

Notes: Scheduled-wedge dose, estimated separately by product. One OLS regression per product (never pooled): product-year late-arrival share (tonnage share accepted Nov 16–Dec 31) or January share on the scheduled out-of-quota tariff (percentage points), crossing years only (the dead zone is defined conditional on realized exhaustion). Heteroskedasticity-robust (HC1) SEs in parentheses. Yellow-corn 2022–23 late shares coded missing (source-file truncation). * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A23: Detected coordination communities

Bloc	Members	Formed	X_B/V	Poultry share	Modal sector	Active 2023	Joint man. 22–23	Club overlap
1	20	2018	0.44	0.20	Other/unknown	8	37	3
2	17	2016	1.44	0.35	Poultry/eggs	13	67	4
3	12	2018	0.23	0.33	Feed	9	11	0

Notes: Sub-threshold groups, U.S.
yellow-corn race windows, 2014–2021. Communities (greedy modularity, size ≥ 3) of the significant-edge graph: edges = pairs co-shipping in ≥ 4 race years whose joint-manifest count exceeds the fine permutation null ($p < .05$; Table A24). X_B/V = members' pooled pre-TPA (or post-entry) monthly tonnage over the dispatch threshold. Formed = first race year by which all significant edges had first co-occurred. Club overlap = members in the interview-verified Agricol seven-firm club. Communities are coarser than the club: traders bridging producer groups pull communities together.

Table A24: Validation: the Agricol club

Pair	Years	Joint manifests	Null mean	Ratio	p (fine)	p (coarse)
Agrícola Col. – Nutrientes Av.	8	77	36.6	2.10	0.000***	0.000***
Agrícola Col. – Villegas Mejía	8	75	36.8	2.04	0.000***	0.000***
Nutrientes Av. – Villegas Mejía	8	72	35.9	2.01	0.000***	0.000***
Incub. Santander – Nutrientes Av.	8	54	28.7	1.88	0.000***	0.000***
Agrícola Col. – Incub. Santander	8	56	30.7	1.83	0.000***	0.000***
Incub. Santander – Villegas Mejía	8	61	36.0	1.70	0.000***	0.000***
Fabipollo – Inv. Santagueda	7	25	16.9	1.48	0.003***	0.000***
Avícola Pilena – Inv. Santagueda	7	18	13.2	1.36	0.010**	0.000***
Avícola Pilena – Fabipollo	7	23	18.6	1.24	0.073*	0.000***
Agrícola Col. – Inv. Santagueda	6	22	17.8	1.24	0.097*	0.000***
Inv. Santagueda – Nutrientes Av.	6	19	16.1	1.18	0.160	0.023**
Inv. Santagueda – Villegas Mejía	6	21	17.9	1.17	0.163	0.010**
Incub. Santander – Inv. Santagueda	6	19	16.6	1.14	0.230	0.017**
Fabipollo – Villegas Mejía	8	26	25.9	1.00	0.557	0.073*
Agrícola Col. – Avícola Pilena	6	17	18.6	0.92	0.790	0.177
Avícola Pilena – Villegas Mejía	6	17	19.1	0.89	0.853	0.210
Avícola Pilena – Incub. Santander	6	14	16.3	0.86	0.893	0.390
Fabipollo – Nutrientes Av.	6	19	22.1	0.86	0.900	0.577
Avícola Pilena – Nutrientes Av.	6	14	16.8	0.83	0.903	0.420
Fabipollo – Incub. Santander	7	20	24.3	0.82	0.933	0.680
Agrícola Col. – Fabipollo	6	20	24.4	0.82	0.950	0.523

Notes: Joint-booking intensity within the interview-verified seven-firm club. All within-club pairs with ≥ 4 race-year recurrence (Avícola Pilena, an entrant, enters the universe through the post-entry size proxy). Joint manifests = distinct shared manifests in race windows 2014–2021; null mean = mean count across 300 draws permuting manifest labels within year × port × 14-day window (fine; conditions on co-presence, isolating joint booking); coarse null permutes within year × port (booking + synchronization). All seven members carry significant edges (8 internal edges at fine-null $p < .05$); at that threshold the club resolves into the four-firm core and the Santagueda-centered periphery, with the bridging edges significant only under the coarse null. * $p < .10$, ** $p < .05$, *** $p < .01$. Figure A7 shows the underlying raw co-shipping pattern.

Table A25: Cost shape: solo vs. shared landed cost

Lot size	Solo ratio	Shared ratio	n solo	n shared
10-500t	1.205	1.025	2,450	7,673
500-2k	0.990	0.991	1,463	11,500
2k-8k	0.990	0.992	919	5,988
>8k	0.994	0.996	90	455

Notes: Landed-cost ratio by importer lot size, no-prize states, yellow corn. FOB unit value over the World Bank maize benchmark, no-prize states only (pre-TPA U.S., non-U.S. origins, 2023). Solo = single ownership group on the manifest. The 10–500 t solo penalty (1.20 vs. 1.02) signs $\theta_0 = s - c\eta > 0$ at small scale; parity above 500 t.

Table A26: Cost shape: deadfreight moments

State	Solo manifests	Share 60–100% of V	Share underfilled 80–100%	Median tons
Race windows (binding years)	461	0.111	0.061	6,256
Pre-TPA (2004–10)	393	0.094	0.038	1,662
Other post-TPA states	502	0.116	0.032	4,212

Notes: Solo-manifest tonnage near the dispatch threshold V. U.S.-origin yellow-corn solo manifests. Underfilled = manifest at 80–100% of V = 16,280 t (an early charter sailing below capacity). The race-window excess over pre-TPA (6.1% vs. 3.8%) measures the deadfreight cost ϕ accepted to unbundle.

Table A27: Incidence: price gaps by regime

	(1)	(2)	(3)	(4)
Panel A: U.S.-origin declarations				
Race window	5.85	4.34	1.14	0.65
	(3.82)	(5.62)	(2.91)	(4.36)
Post-exhaustion	2.99	1.25	6.88**	6.61
	(4.19)	(4.99)	(3.04)	(4.22)
Observations	26,963	19,968	26,963	19,968
Panel B: non-U.S.-origin declarations (falsification)				
Race window	-5.26	0.71	-4.77	4.83
	(6.53)	(3.98)	(13.25)	(7.08)
Post-exhaustion	1.97	10.24	6.51	25.03*
	(6.97)	(6.38)	(14.02)	(12.90)
Observations	12,630	7,497	12,630	7,497
Benchmark	Contemp.	Contemp.	2-mo lag	2-mo lag
Month FE	Yes	Yes	Yes	Yes
Importer–exporter pair FE	No	Yes	No	Yes

Notes: Transaction-price gap over the world price, yellow corn only. Single-product declaration-level regressions: gap (USD/t) = FOB unit value − World Bank maize benchmark, on race-window and post-exhaustion indicators (omitted: non-binding periods). Regime boundaries from the frozen cutoff ledger. SEs clustered by year in parentheses; pair-FE columns restrict to pairs with ≥ 20 declarations. Stars from asymptotic t: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A28: Incidence: annual gaps and implied exporter share

Year	Decl.	Median gap	Gap (2-mo lag)	Wedge (USD/t)	Exporter share	Binding
2004	1,130	1.58	0.01	–	–	No
2005	1,324	0.67	1.64	–	–	No
2006	1,538	-3.87	3.69	–	–	No
2007	1,811	-0.19	3.89	–	–	No
2008	1,693	0.98	3.41	–	–	No
2009	706	6.03	1.96	–	–	No
2010	295	2.51	4.18	–	–	No
2013	246	8.28	-1.82	42.3	0.196	No
2014	1,117	1.83	6.04	40.4	0.045	Yes
2015	1,185	5.84	1.85	28.7	0.203	Yes
2016	1,178	5.47	4.48	23.6	0.231	Yes
2017	1,056	5.79	7.58	19.9	0.290	Yes
2018	1,092	-2.66	8.38	17.5	-0.152	Yes
2019	1,281	0.98	3.34	14.2	0.069	Yes
2020	1,291	6.79	4.56	10.0	0.681	Yes
2021	1,290	-1.95	17.75	10.8	-0.181	Yes
2022	1,150	-16.64	15.99	6.8	-2.462	No
2023	1,542	9.08	-2.82	0.0	–	No

Notes: Median gap, statutory wedge, and implied exporter share, U.S. yellow corn. Race-window declarations in binding years, all declarations otherwise. Wedge = scheduled out-of-quota rate × benchmark price. Exporter share = median gap over wedge $(1 - \beta)$. 2018 and 2021 negatives and the 2020 outlier reflect benchmark-timing noise in spike years; 2022 excluded from interpretation (customs-value anomaly).

Table A29: Incidence: gaps by buyer type

Buyer type	Decl.	All	Race windows	Post-exhaustion
Above-threshold	8,227	4.41	3.18	6.65
Agricol club	2,117	5.32	3.99	7.76
Other sub-threshold	8,122	4.94	3.78	7.62

Notes: Median gap, post-TPA U.S. yellow corn, USD/t. Buyer types from the ownership-group concordance: Agricol club = the seven interview-verified members; above-threshold = pre-TPA size $x/V \geq 0.35$. Race-window medians: above-threshold 3.2, other sub-threshold 3.8, Agricol club 4.0 USD/t—differences within noise, and notably the organized club pays *no less* than unorganized sub-threshold buyers: pooling buys a vessel's departure, not a better price (Section 5.2).