

What an Export Licensing Regime Costs*

Philip A. Luck[†]

This version: August 2026

First draft: February 2026

Abstract

We estimate what loosening national security export controls does to trade. The 2013–2017 Export Control Reform moved less sensitive military products from the U.S. Munitions List to the lighter Commerce Control List. Decontrol raised the affected products’ exports by about 17 log points relative to not-yet-treated defense products, on the order of \$6 billion a year against a \$38 billion base. Most of that gap is a decline suffered by comparable products that remained controlled, which the decontrolled products avoided, so the estimate is a wedge between decontrol and continued control. The gain concentrated in allied and arms-customer markets, and it arrived as larger authorized transactions—the margin an order-based license taxes. Inverted through a trade elasticity measured for these products, the licensing wedge runs 2 to 3.5 percent ad valorem, for an implied administrative saving of \$0.3 to \$1.3 billion a year. In a twelve-region Armington model solved in exact hat algebra, that wedge delivers a U.S. welfare gain of \$2.9 to \$5.1 billion a year, a terms-of-trade improvement that costs allied suppliers market position and almost no real income. The agencies that issued the rules scored the filing burden at \$15 to \$50 million. Stated as a break-even, the reform paid for itself unless decontrol raised expected end-use losses above one to four cents per dollar of the affected trade. Identification rests on the rollout order having been fixed years in advance, so the estimates are read under that assumption throughout.

JEL codes: F13, F14, F52, L52.

Keywords: export controls, export licensing, ITAR, decontrol, defense trade, difference-in-differences, regulatory impact analysis.

*I am grateful to colleagues at the Center for Strategic and International Studies for comments. The views expressed are the author’s alone and do not represent those of CSIS or its funders. All errors are mine.

[†]Center for Strategic and International Studies, Washington, DC. Email: philip.a.luck@gmail.com.

1 Introduction

The 2013–2017 Export Control Reform (ECR) was arguably the largest deliberate loosening of U.S. defense export controls in decades. With export controls having been tightened and loosened in recent years—justified either as protecting national security or as promoting exports—understanding the opportunity cost of controls is central to designing effective policy. Official regulatory impact analyses price the paperwork of licensing; we estimate the cost of the licensing regime itself.

The empirical exercise is possible because the 2013–2017 ECR introduced a staggered, pre-announced loosening of controls covering a defined set of military products. Each covered Munitions List category was rewritten as an enumerated list, with less-sensitive items transferred to a new set of Commerce Control List entries—the “600 series”—while the most sensitive items remained under their existing controls. Credible estimates of the economic effects of export controls are scarce because treatment is rarely quasi-random. Moreover, the natural measure of control stringency—the rate at which exports are subject to individual review—cannot serve as an instrument because its denominator is itself affected by the controls.

The goal of this paper is to use the phased rollout of these decontrols to identify their impact on trade. The 600-series transfers avoid both problems. They are discrete, dated, and ordered years in advance, allowing us to hold the underlying item fixed while varying the regulatory regime governing it. We define treatment using pre-reform licensing records rather than realized post-reform trade, separating the effect of the policy from selection into trade. We then estimate the effect of decontrol by comparing transferred defense products with defense products that had not yet transferred.

We have four main results. First, decontrol raised the affected products’ exports by about 17 log points relative to not-yet-treated defense products (95 percent interval 8 to 26, and 15 to 18 across heterogeneity-robust estimators). The treated products also held their world-export share while comparable defense products lost roughly a quarter of theirs. Most of both gaps is the controls’ decline, so the estimate is a wedge between decontrol and continued control (Section 8). The gain landed in allied and arms-customer markets, and it registered against the market position of allied industrial suppliers rather than against Russia or China.

Second, the effect behaves as a licensing cost would: it scales with each product’s pre-reform licensing burden, and three quarters of it arrives as larger shipments—the margin an order-based authorization taxes—while the shipment-count margin is small and too imprecise to sign.

Third, we estimate an implied price wedge of controls equivalent to a 2 to 3.5 percent ad valorem tariff (8 to 13.5 percent extrapolated to a fully licensed basis), for an implied administrative saving between roughly \$0.3 and \$1.3 billion a year. The agencies’ own \$15 to \$50 million is one to two orders of magnitude below that, and the two figures are not rival estimates of one quantity: the agencies scored the filing burden on eligible transactions, while the trade response prices the whole regime that burden belonged to. Embedding the wedge in a twelve-region Armington model and solving the counterfactual equilibrium exactly puts the welfare gain at \$2.9 to \$5.1 billion a year, about 0.02 to 0.03 percent of GDP. The gain accrues as a terms-of-trade improvement, and it costs allied suppliers market position but almost no real income. That exercise also disciplines the wedge: the measured range reproduces the estimated export response, while the extrapolated range implies a response two to four times too large (Section 9).

Fourth, nothing beyond trade is detectable. Output, employment, and patenting in the exposed industries do not respond, on bounds set by the data’s resolution: the export gain is about 1.5 percent of those industries’ shipments.

The third and fourth results give the paper relevance beyond the episode. A regulatory impact analysis that scores an export license at an hour of clerical time will understate the cost of the license, by a wide margin, and the current wave of decontrol is being justified on industrial-base grounds that no public aggregate can resolve. The most recent instance, a Bureau of Industry and Security final rule relaxing controls on short-endurance commercial UAVs, cites allied market-share loss as its primary rationale—an explicit acknowledgment that the licensing regime carried a cost the reform is intended to recover.¹ Both statements are about how licensing rules get evaluated, not about 2013 alone, and both are testable claims that this design can make. Two of the paper’s devices are contributions in their own right: the scorecard of Section 14.3, which sets the estimates against the agencies’ own valuation of the relief, and the rubric of Table 6, which turns the estimates into conditions for scoring a proposed decontrol before the fact.

Section 2 places the paper in the literature and Section 3 describes the reform. In Section 4 we set out how licensing can be read as a trade cost and derive the transaction-margin prediction the data later discriminate on. Sections 5 and 7 define the treatment and the design; Section 8 presents the trade results; Section 9 solves the implied wedge in general equilibrium and shows that only the measured wedge reproduces the estimated response; Sections 10 through 10.4 assess identification. Section 11 asks where the gain landed and Section 12 how it arrived. Section 14 prices the licensing regime, scores the estimates against what the reform was promised to deliver, and states the security side as a threshold. Appendix B collects the robustness.

2 Related Literature

We build on two literatures. Methodologically, it treats a discrete product-level policy change as identifying variation, like the tariff studies of Fajgelbaum et al. (2020) and Flaaen and Pierce (2024) and the market-access study of Lileeva and Trefler (2010). The nearest ancestor is Pierce and Schott (2016), who measure a pre-announced policy change that removed an uncertainty wedge on a defined product set, with treatment intensity read off a product-level gap in the policy and not off realized trade; the licensing dose used here plays the same role. Khandelwal et al. (2013) is the closest substantive analog: what was removed there was an administrative allocation and licensing apparatus rather than a border tax, and their finding that the apparatus itself carried much of the distortion is the claim this paper’s ad-valorem equivalent puts a number on. Substantively, the paper joins the young economics of export controls, whose evidence is almost entirely about tightening. Crosignani et al. (2026) find Entity-List controls impose large costs on the U.S. firms that supply designated Chinese entities; Makioka and Zhang (2024) measure the trade collapse when Japan’s 2019 controls moved three chemical inputs to individual per-contract licensing, which is the per-order margin of Section 4 imposed as tightening; and the sanctions-trade literature prices the cost of restriction directly (Crozet and Hinz, 2020; Felbermayr et al., 2020). The closest published work on reclassification, Liu et al. (2024), exploits within-list inclusion and exclusion under the 2007 U.S. “China Rule” to measure the target country’s innovation response to tightening. We measure the controller’s export response to loosening, which is a different margin and the opposite policy. As far as we are aware, then, no estimate of the trade effect of loosening an export-control regime exists; that is the gap this paper fills.²

¹Bureau of Industry and Security, “Streamlining Export Controls for Drone Exports,” 91 Fed. Reg. (Aug. 14, 2026) (Document No. 2026-16628). The rule reclassifies ECCN 9A012.a.1 items from National Security Column 1 to Column 2, eliminating the license requirement for exports to Country Group A:1 destinations.

²Descriptive before-and-after comparisons of earlier decontrols exist. The Government Accountability Office reports that firearms exports rose about 7 percent after the 2020 transfer of USML Categories I–III to Commerce

U.S. policy has revived the question. In 2025 and 2026 controls were loosened on several margins at once, each on a stated economic rationale: a June 2025 executive order to expedite civil-drone exports and “scale up domestic production,” a September 2025 shift of military-drone review from the missile-regime presumption of denial to the standard for manned fighters, a July 2025 program to finance exports of “full-stack” American AI technology packages, and a January 2026 rule moving H200-class AI accelerators bound for China from denial to case-by-case review.³ Every one asserts that easier export will strengthen the industries that make the good, at an acceptable security cost. The evidence applies unevenly. The case most in the news, H200 accelerators to China, is a dominant-supplier sale of dual-use technology into an adversary market, to which the allied-trade results here least apply, and Section 14 says why.

Identification rests on the exogeneity of the ex-ante rollout schedule. The residual the design cannot rule out is a demand shift graded by product category and specific to U.S. suppliers; a supplier-differenced comparison rules out any shift common to U.S. and foreign suppliers. The sharpest form of that residual sits on the control side. A U.S.-specific procurement contraction falling differentially on the retained and later-transferred categories—over a window containing the 2013 sequestration and the Iraq and Afghanistan drawdown—would generate part of the pattern we measure. Three findings narrow that space. The effect survives removing the categories where the allied platform programs of these years ran; it is present in a cohort whose rule bound at the end of 2016; and domestic defense purchases by product type fell *faster* in the decontrolled categories than in the still-controlled ones, the wrong direction for the contraction story (Appendix B.4). Section 10.4 states the limits that bind, and the estimates are read under them throughout.

3 Policy background: the 2013–2017 reform

The reform originated in an official judgment that the United States controlled too much. In an April 2010 address Secretary of Defense Robert Gates observed that the government approved well over 95 percent of license cases for EU and NATO destinations. A system built for the Cold War was therefore spending most of its effort on applications it was going to grant. He proposed one that “dispenses with the 95 percent of easy cases” and in which “higher walls are placed around fewer, more critical items.”⁴ The remedy he set out was structural: four “singularities,” a single control list under a single licensing agency, with a single enforcement coordination body and a single information system. Three of the four required legislation, Congress did not pass it, and they were never delivered (Stricker and Albright, 2017). What the executive could do under existing authority was reclassify items between the two lists that already existed, and that is what the “600-series” transfers did (Fergusson and Kerr, 2020). The reform this paper measures is therefore not the reform that was announced. It is the part that survived the absence of legislation, which is also why it makes a usable experiment: reclassification required no bargaining, and it ran on an administrative timetable rather than a legislative one.

jurisdiction (GAO-25-106849, February 2025), and industry analyses credit the 2014 satellite transfer with recovering U.S. world-market share. Neither constructs a counterfactual. On economic networks and chokepoints as instruments of coercion see Farrell and Newman (2019), and for a formal treatment Clayton et al. (2023).

³Executive Order 14307, “Unleashing American Drone Dominance,” June 6, 2025 (90 FR 24727), implemented in part by “Streamlining Export Controls for Drone Exports,” 91 FR 2467 (January 21, 2026); U.S. policy update on the export of unmanned aerial systems, September 15, 2025, under Executive Order 14268 (90 FR 15631); Executive Order 14320, “Promoting the Export of the American AI Technology Stack,” July 23, 2025 (90 FR 35393); “Revision to License Review Policy for Advanced Computing Commodities,” 91 FR 1684 (January 15, 2026).

⁴Remarks by Secretary of Defense Robert M. Gates to Business Executives for National Security, April 20, 2010. The address contains no projection of exports, output or employment, a point Section 14.3 returns to.

Scope of the transfers. Each covered Munitions List category was rewritten as an enumerated list, and its less sensitive items were moved to a new block of Commerce Control List entries, the 600 series, while the most sensitive items, among them fighters, bombers, and their principal subsystems, stayed on the Munitions List. Under the Munitions List nearly every export requires an individually adjudicated State Department license, and manufacturers must register with the Directorate of Defense Trade Controls whether or not they ever ship. On the Commerce list a 600-series item still requires a license to most of the world, but exports to close allies can move under License Exception STA without a case-by-case license. The reform also moved these goods from the ITAR’s zero-tolerance see-through rule to the EAR’s existing 25 percent *de minimis* threshold on re-exports, so foreign-made goods incorporating less than 25 percent 600-series content by value escape control on re-export except to arms-embargoed destinations, and it ended registration for firms dealing only in transferred items.⁵ The allied relief is narrower than it sounds. STA reaches only the roughly three dozen closest partners in Country Group A:5, subject to those conditions, so several of the largest U.S. arms customers, among them Israel, the Gulf states, and Taiwan, continued to need licenses. Decontrol changed the forum and the default terms of the licensing decision more than it removed the license itself.

The rollout sequence. The transfers rolled out category by category in an order announced years ahead, though never as a dated calendar. State announced the rewrite of every Munitions List category in a December 2010 advance notice and proposed the aircraft category in November 2011, describing it as the “second in a planned series,” and brought it into effect in October 2013. Aircraft (Category VIII) and gas-turbine engines (Category XIX) transferred first, in October 2013; ground vehicles, naval vessels, submersibles and miscellaneous articles followed in January 2014; energetic materials, protective equipment, training equipment and missile-related articles in July 2014; military electronics at the end of 2014; fire-control and optical systems, toxicological agents and directed-energy weapons at the end of 2016.⁶ One step in that sequence is a documentary anchor, not a stated intention. Removing an item from the Munitions List requires advance notification to Congress under section 38(f)(1) of the Arms Export Control Act, and State transmitted the notification of intent to move the first tranche on 18 March 2013, about a month before the final rules published and seven months before they took effect.⁷ A statutory filing required *before* removal

⁵White House, “Fact Sheet: Announcing the Revised U.S. Export Control System,” October 15, 2013. License Exception STA was established at 76 FR 35276 (June 16, 2011); the 25 percent *de minimis* threshold for 600-series content is at 15 CFR 734.4, which retains a zero threshold for arms-embargoed (Country Group D:5) destinations. The 25 percent figure was not created by the reform: BIS proposed a stricter 10 percent threshold for 600-series content, drew fourteen adverse comments, and retained the EAR’s pre-existing rule (78 FR 22660, p. 22667). STA carries consignee-statement and per-shipment notification conditions and, for 600-series items, is limited to government end-users in the eligible countries; for 600-series *end items*, the category that includes the aircraft of the largest treated family, its use additionally required a one-time prior BIS eligibility approval (15 CFR 740.20(g)).

⁶Effective dates of the Commerce transfer rules: 78 FR 22660 (October 15, 2013); 78 FR 40892 (January 6, 2014); 79 FR 264, with State companion 79 FR 34 (July 1, 2014); 79 FR 37551 (December 30, 2014); 81 FR 70320 and 81 FR 49517 (December 31, 2016). Spacecraft (Category XV) moved on a separate track to the 9x515 series rather than to the 600 series, effective June 27, 2014 (79 FR 27418), and on a legislative and not an administrative clock, which is a reason to keep it outside the treated set. Firearms and ammunition, Categories I–III, did not transfer until March 9, 2020 (85 FR 4136). Proposal-to-effective lags across the sequence run from about eighteen months to three years, most often close to two. No transfer calendar was published *ex ante*: the pre-announcement was of intent and order, not of dates, and the last tranche slipped, both to a December 31, 2016 effective date (80 FR 34572; 80 FR 34562), for administrative and political reasons unrelated to the products’ trade.

⁷Executive Communication 738, House of Representatives, March 18, 2013, transmitting “an intention to transfer jurisdictional control of certain classes of items currently on the United States Munitions List to the Commerce Control List, pursuant to Section 38(f)(1)”; initial-implementation final rules 78 FR 22660 (Commerce) and 78 FR

cannot be a response to trade realized after it. Firearms and ammunition, held back inside the reform window and not transferred until 2020, are the clean set of not-yet-treated controls the design uses. The reason for their lateness is on the record. The official who owned the category told Congress in February 2016 that over 90 percent of the items in it were not exported to governments or militaries, so it raised no interoperability question, and that although drafts had been written in 2012 “we suspended those efforts.”⁸ A customer mix that is mostly civilian is not a property of these products’ trade performance, which is what the not-yet-treated control logic requires.

The multilateral baseline. One feature of the reform decides how its competitive effects should be read. The transfer did not take these goods out of multilateral control. The 600 series is the United States’ own implementation of the Wassenaar Munitions List on the Commerce Control List: it was built by consolidating the thirteen WAML entries already carried there under the xY018 numbers, and the rule creating it states that 600-series items are controlled “because they are Wassenaar Arrangement Munitions List (WAML) and formerly USML items subject to the jurisdiction of the EAR.”⁹ What the United States gave up was the stricter *national* treatment it had layered on top of the multilateral list, not its participation in the list. One distinction keeps the reading precise. The Wassenaar Arrangement harmonizes what its members *list* and not how stringently they license it: it is a consensus regime with no common licensing standard, no verification, and explicit national discretion, so co-signatories applying the same Munitions List span a wide range of national stringency. The experiment is therefore a unilateral change in the national stringency the United States applied to goods that remained multilaterally listed, against rivals’ national implementations of a common list that did not move. That is the margin the estimates identify, and it is the one relevant to a government asking whether to keep licensing a shared list more strictly than its competitors do.

4 Conceptual framework: licensing as a trade cost

Because the license was nearly always granted, its burden was the process rather than the prohibition, which is what makes it a trade cost of the kind the empirical literature prices (Hummels and Schaur, 2013; Djankov et al., 2010). Three components can be read off the regulation. The largest falls on the *order*: a State license is issued per applicant, end user and order, valid for four years of shipments against an approved value (22 CFR 123.1(c), 123.21(a)), so each new order or customer requires a fresh adjudication and each authorization caps the value that may move under it.¹⁰

22740 (State), April 16, 2013, effective October 15, 2013.

⁸Testimony and written statement of Brian Nilsson, Deputy Assistant Secretary of State for Defense Trade Controls, House Small Business Committee, February 11, 2016. The §38(f)(1) notification for the firearms categories was not transmitted until November 13, 2019, eighteen months after the proposed rules of May 2018.

⁹78 FR 22660, at 22661 and 22668. Across 2013–2017 the Wassenaar Arrangement kept every relevant munitions category controlled—fire-control equipment (ML5), ground vehicles (ML6), vessels of war (ML9), aircraft and aero-engines (ML10), and military electronics (ML11)—with only routine technical amendments and no category added or removed, and the European Union re-adopted its Common Military List on the same categories each year. Sources: the annual lists WA-LIST (13)1, WA-LIST (14)2, WA-LIST (15)1, WA-LIST (16)1 and WA-LIST (17)1 with their Summaries of Changes; EU Common Military List, *Official Journal* C 90 (27.3.2013), C 107 (9.4.2014), C 129 (21.4.2015), C 122 (6.4.2016), and C 97 (28.3.2017).

¹⁰22 CFR 123.21(a): “A license is valid for four years. The license expires when the total value or quantity authorized has been shipped or when the date of expiration has been reached, whichever occurs first.” Registration is a separate standing cost: fees ran in three tiers under the schedule then in force, \$2,250 a year at the base and \$2,750 plus \$250 per authorization above ten at the top, paid by every firm that makes a USML item, and the Aerospace Industries Association told Congress in February 2012 that 68 percent of registered companies never export at all.

A second and smaller cost falls on the *consignment*: every shipment moves under the filing and license-decrementation apparatus of individually licensed trade. A third component, less visible because it is not a shipment license, sits above both. Defense services and technical data required a negotiated agreement with the Directorate of Defense Trade Controls under Part 124 of the ITAR, approved with provisos on timelines materially longer than a shipment license, and the Commerce regime has no analog. Sustainment, repair and return, overhaul and co-production for aircraft and gas-turbine engines run on those agreements, and those two families carry over half the dollar dose here, so the agreement layer is plausibly the largest single piece of administrative relief in the transfer. It is negotiated per program rather than per consignment, which points the same way as the per-order reading below.

What the transfer removed is narrower than “decontrol” suggests, and the oversight record enumerates it. Comparing individually licensed export against export moving under a country-based authorization, GAO identified ten compliance and enforcement activities and found that exactly three differ: review of the license application, per-transaction vetting of the parties against the watch lists, and compliance-program review and recordkeeping. End-use monitoring, selection of shipments for post-shipment verification, border inspection, seizure, investigation and penalties apply either way.¹¹ The relief therefore has two tiers. A forum shift, to Commerce adjudication with no registration and a *de minimis*, reaches every destination that previously needed a State license; STA adds a conditional allied margin on top of it. That allied tier is bounded at exactly the transactions the size margin of Section 12.2 picks up, since STA cannot be used for 600-series Major Defense Equipment on contracts above \$25 million (15 CFR 740.20(b)(3)(iv)). Large platform sales to A:5 governments thus continued to need an individual Commerce license, which is a sharper reason than destination composition for the muted A:5 premium reported below, and a reason to locate the operative mechanism in the forum shift with the license exception as its allied edge.

Predictions. In a trade model with destination-specific trade costs, and in the policy-uncertainty models where an individual license is a sunk or recurring export cost (Handley and Limão, 2015, 2017), lowering the cost of serving a set of destinations raises exports to them. Three predictions follow. First, exports and the U.S. share of the world market for the decontrolled goods rise, because the friction on U.S. supply fell while foreign competitors’ did not. Second, the gain concentrates in the destinations whose cost fell most, the allied and defense-customer markets that bore the individual-licensing burden. Third, and this is what lets the data discriminate between cost structures instead of assuming one, the per-order and per-consignment costs imply *different* transaction margins. Relief on a cost borne per consignment lowers the price of an additional shipment, so exporters de-bundle and the gain arrives as more and smaller shipments (Hornok and Koren, 2015a,b; Kropf and Sauré, 2014). Relief on a cost borne per order instead raises the size of the transaction a single adjudication supports, so the gain arrives as larger consignments against a roughly unchanged count. The per-order case has a measured tightening mirror: when Japan’s 2019 controls required an individual license for each export contract on three inputs to Korea, the affected trade collapsed, the response a per-order cost predicts run in reverse (Makioka and Zhang, 2024). Section 12.2 reports that decomposition, which the Census filing records support directly because they count consignments as well as valuing them.¹² Either version is intensive-margin in

¹¹GAO-13-119R, November 16, 2012. STA also inserts documentary substitutes: a consignee statement recording the ECCN and barring onward transfer, a Commerce compliance review for every first-time STA user, and analysis of STA shipping data. In its first twelve months STA carried 74 exporters, 604 transactions and \$33.9 million.

¹²The research history belongs here, because it bears on how the test should be read. An earlier version of this paper predicted the per-consignment case and read the export response as intensive-margin on that basis; the decomposition

the usual sense, deeper sales to buyers already served.

The per-relationship structure the third prediction turns on is written into the license exception as well as inferable from behavior. For 600-series items STA may be used only if the purchaser, intermediate consignee, ultimate consignee and end user *have previously been approved* on a license or other authorization.¹³ Case-by-case adjudication therefore survives for a counterparty the exporter has not sold to before, and is removed only for repeat business with parties already vetted. That is a regulatory statement of the mechanism, and it carries three predictions with no estimation at all: deeper sales into existing relationships, no entry into new ones, and larger authorized transactions against a roughly unchanged set of consignees. Sections 12.2 and 11.1 report all three.

Scope conditions. Decontrol here is a market-access shock. It lowers the cost of selling an existing product abroad and leaves the technology and capacity of the industries that make the good untouched, so it carries no first-order prediction for domestic innovation, output or employment, and none for whether the good reaches an unauthorized end user, which turns on enforcement and not on the licensing cost. The mechanism also presumes a competitive foreign fringe. It describes decontrol of goods that have substitutable foreign supply, and does not extend to a dominant-supplier sale into an adversary market, where the trade elasticity is small and the security question is capability transfer rather than diversion (García-Alonso and Levine, 2007), a limit Section 14 returns to. One channel runs through third-country manufacturers and not through U.S. exporters at all: because the ITAR see-through rule subjected any foreign good with U.S. munitions content to U.S. re-export control whatever the share, foreign manufacturers had a standing reason to design that content out, and a 25 percent *de minimis* removes the penalty below the threshold. That predicts a shift back toward U.S. supply against the allied industrial suppliers capable of building the substitute, which Section 11.1 finds. Finally, the framework makes no prediction at all about government-to-government procurement. Sales routed through the Defense Security Cooperation Agency go exclusively to U.S. firms and are negotiated between governments, so neither the competitive-fringe logic nor the per-order adjudication binds them. That channel turns out to carry most of the dollar increase, which is why Section 11.3 decomposes it and why the estimate we lead with is the commercial one.

5 Treatment: the 600-series transfer

Figure 1 sets out the experiment. The licensing regime that governs these products handed off at transfer, from ITAR individual licensing to a Commerce 600-series classification, and the products' exports afterward diverged from comparable defense goods that had not yet transferred. The rest of this section and Section 7 define its pieces.

A 600-series ECCN matches the pattern [0-9] [A-E] 6 [0-9] [0-9] (for example 9A610, aircraft). We flag these cells in the U.S. Census ECCN panel (29.3 million quarterly commodity-by-destination-by-license cells, 2009Q1–2026Q1) and assign each 600-series family its data-driven cohort, the

of Section 12.2 did not support it, and this framework was rewritten around the per-order structure. The framework therefore did not predict the outcome in advance, and we do not present it as having done so. What the two branches supply is a discrimination between cost structures that were both ex-ante plausible and that imply opposite signs on an observable margin, which is more than the literature on administrative trade barriers usually gets to condition on, and less than a prediction the theory passed.

¹³Note to 15 CFR 740.20(c)(1)(i). The qualifying prior approvals are a BIS license or a DDTC Manufacturing License Agreement, Technical Assistance Agreement, Warehouse Distribution Agreement, or General Correspondence approval.

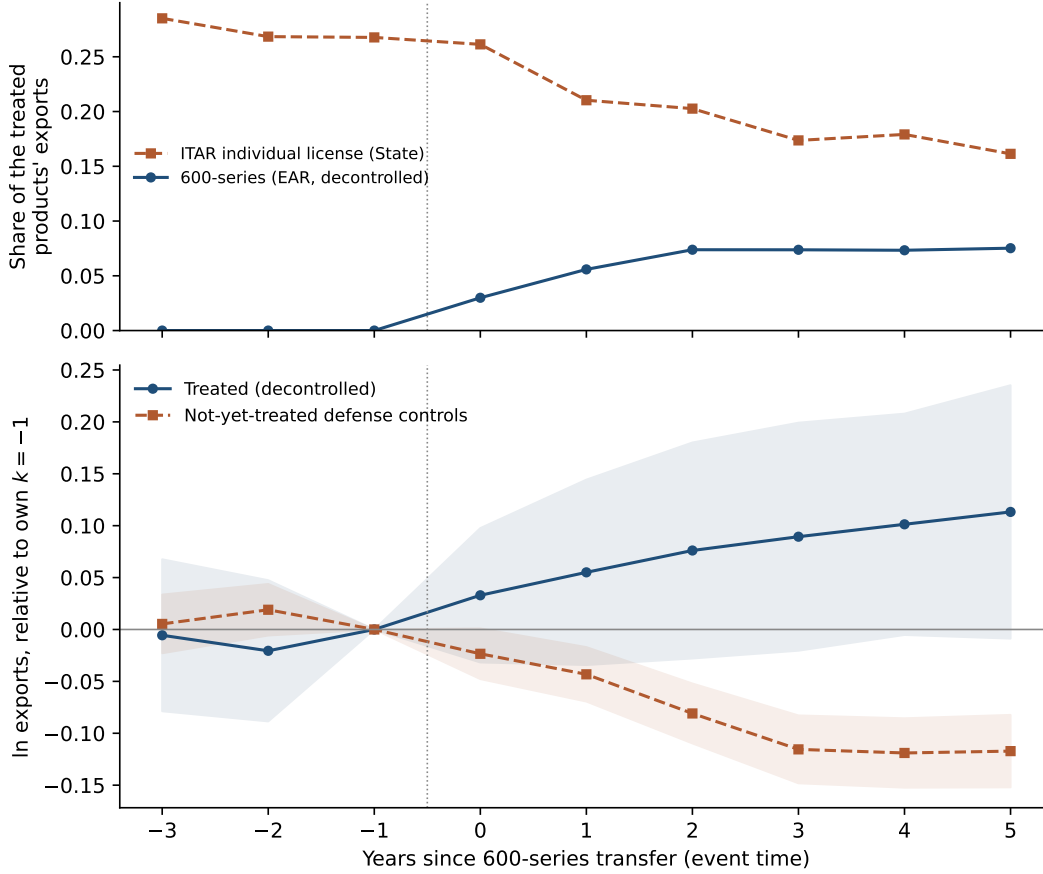


Figure 1. The decontrol experiment. Treated products are the decontrolled defense goods (the regime-gated treated set, Section 5); controls are defense products not yet transferred at the same event time. *Top:* averaged across the treated products and by event time, the share of their exports moving under ITAR individual licenses falls as the share under the decontrolled 600-series classification rises, so the regime hands off at the transfer. *Bottom:* the treated products' total exports, each measured against its own value the year before transfer and averaged across the group, separate from the control group's after event time zero. The gap that opens is the difference-in-differences estimate; bands are 95 percent intervals across products.

first calendar year in which the family carries any 600-series value in the filing records. Pre-2013 the families are absent; they appear exactly when the corresponding USML category transferred. The cohort is the first full trading year under the new classification—two of the five rules took effect within days of a year boundary, so the data-driven rule assigns those families the following calendar year—and it reproduces the documented ECR tranches exactly: 7/7 families with a known Federal Register rule date have a first-appearance year equal to the first full trading year the rule implies. Transferred value concentrates in aircraft (9x610, rule October 2013, cohort 2013), gas turbine engines (9x619, rule October 2013, cohort 2013), military electronics (3x611, rule effective 30 December 2014, cohort 2015), and ground vehicles (0x606, rule January 2014, cohort 2014). These products already exported before reform, much of it under ITAR individual licenses; at each family's rule date the 600-series classification appears and the individual-license value gives way to it while total exports keep rising (Figure 1, top panel). The transfer relabels the control regime of an existing good and creates no new trade line, so the staggered appearance of the 600-series slice, \$0 before each rule, is the candidate identifying variation.

Defining the treated set. A 600-series code marks a product as having been transferred, but two ways of turning that mark into a treated set are flawed. Restricting to products with realized 600-series trade above a value threshold (the *realized* set, 146 HS6) conditions the treated set on the outcome: products where decontrol generated little trade are misclassified as controls, biasing the effect upward. Using every product that ever carries a 600-series code (the *full-eligible* set, 1,179 HS6) avoids that but dilutes the treatment: most of those products carried only a sliver of military trade before reform, so decontrol touches a small part of the HS6 and the product-level outcome is dominated by untreated trade.

We use a third definition that conditions on the *pre-reform licensing regime* instead of on realized trade. A product is *regime-gated treated* if it ever carries a 600-series code *and* at least 5 percent of its 2009–2012 export value shipped under ITAR individual licenses (DDTC codes S05/S61/S73/S85/SAG). This identifies products that genuinely required State individual licenses before reform, the items the transfer actually decontrolled, without reference to how much trade materialized afterward.¹⁴ The gate is built on the license-code field, whose totals match published Commerce license-code aggregates to within a few percent; the post-transfer ECCN plays no role. The regime-gated set is 90 HS6 with \$33.0B in cumulative 600-series transferred trade (dominant-family Herfindahl 3,047; 9x610 aircraft is 48%), a concentration we address in Appendix B.1. This cumulative transferred value is distinct from the annual export base of these products (about \$32 billion *a year*, Section 14), which the two figures coincidentally approximate. Table 1 collects the treated-set descriptives and the estimation samples the results below are computed on.

¹⁴The gate counts individually adjudicated State authorizations and excludes the two ITAR channels that moved without one: exemption shipments (mainly the Canadian exemption, 22 CFR 126.5) and government-to-government Foreign Military Sales. Products whose pre-reform trade ran heavily through those channels are under-dosed relative to their full ITAR exposure, which makes the gate conservative for identifying State-licensed products and, if anything, attenuates the measured dose. License type codes are defined in the AES filing codes at 15 CFR part 30, Appendix C, and in U.S. Census Bureau, *Automated Export System Trade Interface Requirements* (AESTIR), Part III, Appendix F. The five gate codes are the individually adjudicated State authorizations: S05 (DSP-5, permanent export of unclassified defense articles), S61 (DSP-61, temporary import), S73 (DSP-73, temporary export), S85 (DSP-85, classified), and SAG (Part 124 agreements). C59 designates the Strategic Trade Authorization exception and S94 (DSP-94) Foreign Military Sales in the same appendix.

Table 1. The treated set, the comparison universe, and the estimation samples

<i>Panel A. Regime-gated treated products</i>	
Treated HS6, nominal / estimation sample	90 / 88
Pre-reform export base (2009–2012, annual)	\$32.1B
Post-transfer export base (2013–2024, annual)	\$37.7B
Pre-reform value under ITAR individual licenses (annual)	\$9.0B
Mean licensing dose, unweighted / value-weighted	0.27 / 0.28
Median pre-reform export base per product (2009–2012 total)	\$404M
Median pre-reform U.S. world-export share	17%
Dominant-family Herfindahl	3,047
Largest families (9x610 aircraft / 9x619 engines / 3x611 electronics)	48% / 23% / 14%
<i>Panel B. Comparison universe and stacked samples</i>	
Defense-products universe (HS6)	1,831
Candidate control HS6 (not-yet- and never-treated defense products)	1,743
Export outcome: stacked observations / treated HS6	13,155 / 88
World-share outcome (BACI): stacked observations / treated HS6	4,078 / 77

Dollar figures are from the Census ECCN export panel; the licensing dose is the 2009–2012 share of a product’s export value shipped under ITAR individual-license codes (Section 5). The Herfindahl is computed over the 600-series families’ shares of cumulative transferred value. The defense-products universe comprises HS6 that ever received a 600-series transfer or carried at least 5 percent ITAR individual-license value pre-reform (Section 7); each sub-experiment’s controls are the universe members not treated within its window. Eleven treated codes lack a BACI world-share panel, leaving 77 for that outcome. Scripts 63, 103.

6 Data and outcomes

We study two outcomes. The first is the log of total U.S. exports of the HS6, from the Census panel (complete calendar years 2009–2025 of a filing-record panel that runs through 2026Q1). This is not mechanical: the pre-transfer military variant already entered Census totals as ITAR S-code value, so a post-transfer increase reflects product export growth, not a code relabel. Appendix B.12 confirms that no treated code is a newly created HS6 line, including at the 2017 nomenclature revision, and that removing special-provision and revision-vintage codes leaves the estimate unchanged. The second is the U.S. share of *world* exports of the HS6, computed from BACI (CEPII’s reconciliation of UN Comtrade), HS12 for 2012–2024 and HS92 for 2009–2011, over the defense-relevant chapters (HS 84, 85, 87, 88, 89, 90, 93). We model the share in inverse hyperbolic sine, so its coefficients read as *proportional* changes in the share and not as changes in percentage points; because the treated products’ base world shares run well above one percent, a coefficient of +0.15 is a rise of about 15 percent in the share, which translates to a few percentage points on those base shares. Every treated product has a pre-reform world share above one percent (median 17 percent), so the transformation is effectively a log across the treated set, and re-estimating with the log of the share on the positive-share subsample returns +0.169, confirming the proportional reading (Appendix B.3). The share outcome differences out global demand: a rise in it cannot be explained by a U.S.-wide defense-export boom. The principal datasets are summarized in Table 7.

7 Empirical design

The comparison is restricted to a *defense-products universe*: HS6 that ever received a 600-series transfer or carried at least 5 percent ITAR S-code value in the pre-period. Controls are therefore still-ITAR or decontrolled-later defense products; no commercial good enters the comparison. To handle staggered timing we use a stacked estimator (Cengiz et al., 2019): one clean sub-experiment per cohort g , comprising the cohort- g treated units and only controls not treated within g ’s window $[g - 3, g + 5]$, each sub-experiment with its own unit and time fixed effects. No unit treated within a sub-experiment’s window enters that sub-experiment’s controls, so the design avoids the negative weighting of two-way fixed effects under heterogeneous timing (Goodman-Bacon, 2021; de Chaisemartin and D’Haultfœuille, 2020).¹⁵ Late-transferring categories, including firearms (USML Categories I–III, transferred in 2020), serve as clean not-yet-treated controls for the early cohorts.¹⁶ Transferred products below the regime gate are neither treated nor used as controls: their own 2013–2017 cohort falls inside every sub-experiment’s window, so dropping them from the universe leaves the estimate unchanged (Appendix B.3). The controls are never- and later-transferred defense products, and restricting them to the never-600 set alone leaves the export effect if anything larger, at reduced precision. Because treated and control defense products are made by overlapping firms and bought by overlapping customers, spillovers from treated to control products—substitution that would make the control path partly treatment-caused—are a design assumption and not a given. Appendix B.4 tests the assumption directly: the sharpest substitution prediction, that the closest substitutes decline most, is rejected (same-heading controls declined less, and excluding them raises the estimate), while the control group’s decline turns out to be specific to U.S. suppliers of those products, which rules out a worldwide demand shock and is consistent with continuing ITAR status itself costing the controls ground. Standard errors are clustered on HS6. The estimating equation, run on the stacked data, is

$$y_{i,t,g} = \alpha_{i,g} + \gamma_{t,g} + \sum_{k \neq -1} \beta_k \mathbf{1}\{t - g = k\} \cdot \text{Treat}_i + \varepsilon_{i,t,g}, \quad (1)$$

where i indexes HS6, t year, g cohort, and k is event time relative to transfer. In words, β_k is the average gap between the decontrolled products and the not-yet-transferred controls k years from transfer, each measured against its own value the year before; the pooled effect reported below replaces the event-time indicators with a single post-transfer dummy. Only the definition of Treat_i changes across the three treated sets; the universe, controls, windows, and fixed effects are held fixed, so the estimates are directly comparable.

Reading conventions. Because we report many specifications, the rule for reading them is stated once here and applied throughout. Every pooled estimate is accompanied by the largest absolute pre-period event-study coefficient from its own event study, written $\max|\text{pre}|$, and where a Rambachan and Roth (2023) sensitivity analysis is available by its breakdown value. We treat an estimate as supported when $\max|\text{pre}|$ is below roughly a third of the average post effect *and* the breakdown value is at least 0.20; estimates failing either condition are reported as directional; we do not read them as magnitudes, and we say which condition they fail. The rule binds against us twice.

¹⁵See Roth et al. (2023) for a synthesis of the recent difference-in-differences literature on designs with heterogeneous treatment timing.

¹⁶The 2020 rule moved Categories I–III to a 500-series Commerce classification, so the data-driven 600-series cohort rule leaves firearms products without stray 600-series codes coded as never-treated; six firearms HS6 sit in the control pools through the panel windows. Appendix B.4 shows the estimates are unchanged when these controls are censored at the 2018 proposed rule or the 2020 effective date, or dropped.

Both cases belong here, ahead of the estimates they qualify. The world-share headline sits at the boundary of the first condition ($\max|\text{pre}|$ 0.049 against an average post effect of 0.147), which is one reason we treat export volume as the primary outcome; and the out-of-sample firearms exercise of Appendix B.4 fails it outright, which is why we do not count that exercise as validation of the design. The rule is a reading convention and not a pre-test: a pre-period deviation is informative only against the size of the effect it might explain, which is the quantity the breakdown value measures directly.

8 Main results

In Table 2 we report the pooled ATT under all three treated definitions, for both outcomes. The realized set reproduces a large headline (volume +0.181, share +0.106); the full-eligible set roughly halves it (volume +0.061, share +0.047). The regime-gated set recovers the effect at the headline magnitude (volume +0.170, share +0.147) while *not* conditioning on realized trade. The pattern says the full-eligible attenuation is mostly treatment dilution: the 1,089 products the full set adds beyond the regime-gated set carried little pre-reform military trade and pull the average toward zero. The magnitudes are economically sensible: exports rise about 17 log points (roughly 18 percent) and the world-market share of the decontrolled products by a similar order over five years. Figure 3 plots the regime-gated event studies: both outcomes rise after transfer, and the export pre-trend, the weakness of the realized set, is flat (largest pre-period coefficient 0.028).

One restriction on the headline belongs here and not in a robustness table. Most of the *dollar* increase in these products' exports ran through government-to-government Foreign Military Sales, a channel an individual export license never governed (Section 11.3). We therefore report the effect on commercial, individually licensable trade as the paper's headline: estimated on all non-FMS value the same design returns +0.165 (SE 0.048, wild-cluster $p = 0.0005$) against +0.170 on the total, on a cleaner pre-period. The two figures are close because the FMS share of these products' value is small even though the FMS share of the dollar *change* is large, and every result below that could be affected by the channel is reported both ways.

The two outcomes are clean on complementary margins, and we treat the export-volume result as primary and the world share as corroborating. The export effect is the more robustly inferred: it survives every clustering level and inference method in Appendix B.8 and the removal of the aircraft family, but it does not difference out global demand. The world-share effect does difference out demand, so it cannot be explained by a U.S.-wide defense-export boom, but it is the more fragile of the two under the dose concentration: it weakens at chapter-level clustering and on the thinnest control set (Appendices B.8, B.3). As the data section notes, the arcsinh share coefficient reads as a proportional change; the tables report that specification.

Figure 2 shows the same comparison in group means, not regression coefficients: within each product, $\ln$ exports and the world-share transform relative to the product's own value the year before transfer, averaged over the treated and over the control products. Before transfer the treated and control paths sit together; after it they separate. For exports the separation is roughly symmetric, the treated products rising about eleven log points (twelve percent) over five years while the controls fall about twelve, and the gap of roughly twenty-three log points is the pooled estimate at that horizon. For world share it is not symmetric. The treated products hold their share close to flat while comparable defense products lose about twenty-five percent of theirs, so the estimated share gain is mostly these products avoiding a decline that reached the rest of the defense sector. The pooled coefficient of +0.147 is smaller than that twenty-five-percent endpoint because it averages

Table 2. Pooled ATT by treated-set definition (same universe, controls, and fixed effects)

Treated-set definition	n HS6	ln U.S. exports	U.S. world share	max pre
Realized ($\geq$ \$50M 600-series trade)	144	+0.181 (.033)	+0.106 (.039)	.082 / .021
Full-eligible (ever 600, any value)	1,177	+0.061 (.023)	+0.047 (.031)	.019 / .013
Regime-gated (ever 600 $\cap$ ITAR _{ind} $\geq$ 5%)	88	+0.170 (.045)	+0.147 (.048)	.028 / .049

Stacked DiD, reference $k = -1$, not-yet-treated defense controls, SEs clustered on HS6. Columns vary only the treated-set rule; the defense universe, controls, windows, and fixed effects are identical. “ITAR_{ind}” is the 2009–2012 share of HS6 export value under ITAR individual-license codes **S05/S61/S73/S85/SAG**. “max |pre|” lists the largest absolute pre-period event-study coefficient for the volume / share outcome. “ n HS6” is the estimation-sample count for the export outcome (the share outcome loses codes without a BACI panel: 77 for the regime-gated row); the nominal regime-gated definition contains 90 codes, per Table 1. The code-count accounting is in Appendix B.12. All estimates are significant at the 1 percent level except the full-eligible world-share estimate, which is not significant at conventional levels.

over the post window and is an inverse-hyperbolic-sine change, the same endpoint-versus-average gap the export series shows above.

The control decline is not a world-demand story. Within the same control products, non-U.S. suppliers’ exports did not fall at all, while U.S. exports of them fell seventeen log points relative to non-U.S. exports (Appendix B.4). The control path thus embodies the trajectory of U.S. defense products that remained under ITAR, which is the policy-relevant counterfactual if continuing control itself cost them ground to foreign suppliers and a contaminated one if decontrol diverted U.S. sales from them. Stated as an estimand, what the design identifies is the policy wedge between decontrol and continued ITAR treatment: what moving a product off the individual-licensing regime was worth relative to leaving it on. We use the term wedge throughout to mark that the counterfactual is continued control, and the two readings coincide only to the extent that the controls’ decline is itself the cost of continuing control. Appendix B.4 tests that attribution where it can: the substitution reading fails where its prediction is sharpest, on the closest substitutes, which declined less rather than more.

Two readings survive that test. Continuing ITAR treatment cost the still-controlled products ground to unconstrained foreign suppliers, in which case the control path is the policy-relevant counterfactual and the wedge is the object we want; or a U.S.-specific procurement contraction, the post-2012 drawdown and sequestration working through categories served largely by U.S. incumbents, drove the decline, in which case part of the measured wedge is not a licensing effect. The trade data do not adjudicate between them, and the supplier-differenced bound of Section 10.3 does not either, because a contraction operating through incumbency is U.S.-supplier-specific by construction. A measured bound bears directly on the choice: domestic federal defense purchases by product type contracted *more* in the decontrolled categories than in the still-controlled ones, a mean log change of -0.34 against -0.14 from 2009–2012 to 2014–2018 (BEA NIPA Table 3.11.5), a differential of 13 to 20 log points with the wrong sign for the drawdown reading (Appendix B.4). Section 14.2 prices the less favorable reading as a scenario.

Robustness summary. Table 3 collects the specifications that matter most in one place, several of which are developed later in the paper and appear here so the result can be read against its main challenges at once. Across two heterogeneity-robust estimators, the measured demand controls of Section 10.3, the supplier-differenced triple difference, the removal of the eleven miscoded arms codes, and a window ending before the 2018 tariffs, the export effect stays between $+0.137$ and $+0.229$ against the baseline $+0.170$. Removing the aircraft and gas-turbine engine families outright,

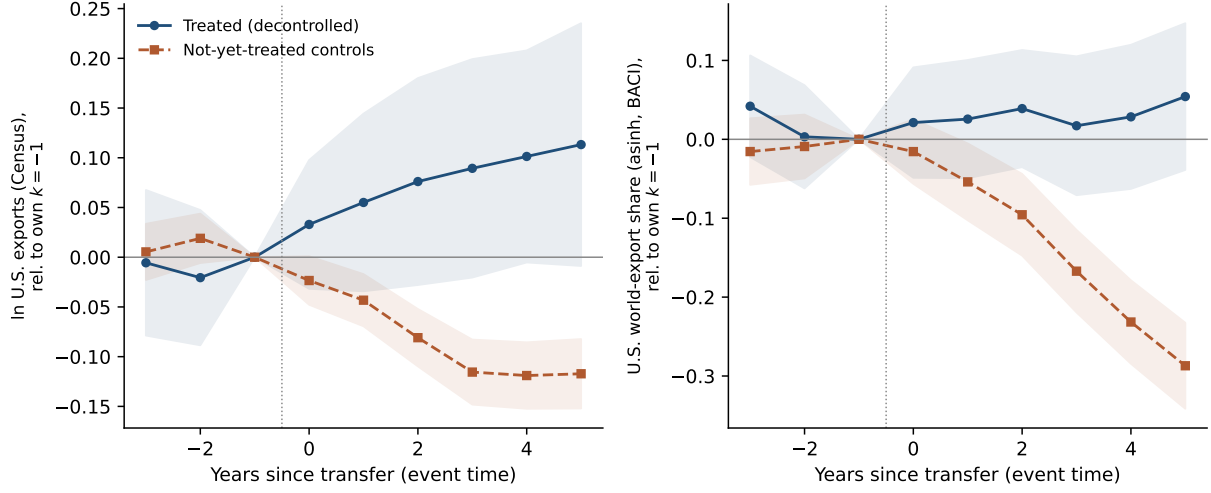


Figure 2. The difference-in-differences in group means. Each series is a within-product change relative to the product’s own value the year before transfer ($k = -1$), averaged across the treated (decontrolled) products and across the not-yet-treated defense controls; bands are 95 percent intervals across products. *Left*: ln U.S. exports. *Right*: the U.S. world-export share (inverse hyperbolic sine, BACI). The vertical gap after $k = 0$ is the estimated effect. For the world share the gap opens mainly through the controls’ decline, not an absolute rise in the treated products’ share.

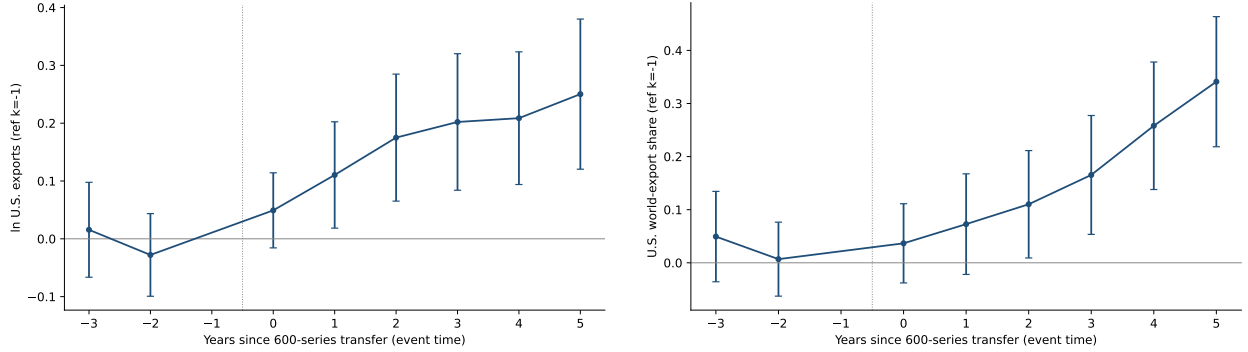


Figure 3. Stacked event studies, regime-gated treatment. *Left*: ln U.S. exports (primary). *Right*: U.S. world-export share (demand-differenced). 95% confidence intervals, clustered on HS6. The largest pre-period coefficient is 0.028 on exports and 0.049 on the share, against average post effects of 0.170 and 0.147: flat on exports, and at the boundary of the reading rule of Section 7 on the share.

the two categories where the allied fighter programs of these years ran, leaves $+0.137$ on exports and *raises* the world-share effect to $+0.193$, on cleaner pre-periods than the full sample. The effect is also positive in every transferred family, chapter, and cohort, and stable as the regime gate tightens from 2 percent to 10 percent.

Because the dollar dose is concentrated in the aircraft family, we also stress the inference. The export ATT keeps a restricted wild-cluster bootstrap $p \leq 0.025$ and a randomization-inference $p < 0.001$ even when clusters are coarsened from HS6 to the HS2 chapter, which collapses the aircraft mass into a single cluster. The world-share ATT behaves the same way except at the most demanding cut, chapter-level clustering of the seven-chapter share panel, where the wild-cluster p rises to 0.10 while randomization inference holds at 0.032 (Appendix B.8). Appendix B collects the

Table 3. The regime-gated effect across the specifications that challenge it

Specification	ln U.S. exports	U.S. world share
Baseline stacked estimator (Table 2)	+0.170 (.045)	+0.147 (.048)
<i>Estimator</i>		
Callaway–Sant’Anna	+0.184 (.048)	+0.163 (.037)
Sun–Abraham	+0.150 (.042)	+0.135 (.036)
<i>Demand</i>		
+ measured category demand, SIPRI (preferred)	+0.142 (.048)	+0.134 (.048)
+ product-level world-excl.-U.S. exports	+0.229 (.055)	+0.149 (.043)
Supplier-differenced triple difference ^a	+0.188 (.055)	—
<i>Treated set</i>		
Dropping the aircraft and engine families ^b	+0.137 (.047)	+0.193 (.057)
Dropping the eleven miscoded arms codes	+0.161 (.048)	+0.160 (.050)
<i>Window</i>		
Through 2017 only, before the 232/301 tariffs ^c	+0.138 (.044)	+0.063 (.046)

Pooled ATT on the regime-gated treated set; standard errors clustered on HS6 in parentheses. All rows use the 88-code estimation sample of Table 1, which is 13,155 stacked observations for the export outcome and 4,078 for the world share (77 treated codes); rows that restrict the treated set or the window use the corresponding subset. Appendix B.7 reports the estimator comparison. Sections 10.3, B.11, B.1 and B.6 develop each row; this table gathers them.

^a Treated versus control in the U.S.-versus-non-U.S. export gap, before versus after transfer, so demand common to a product’s suppliers cancels. It states the world-share comparison as an explicit triple difference and has no separate share counterpart.

^b **9x610** and **9x619** removed jointly, leaving 48 treated products with no exposure to the fighter and gas-turbine programs of 2014–2018. Pre-period deviations are 0.016 and 0.020, cleaner than the full sample’s.

^c The world-share effect matures over the fourth and fifth post-transfer years, so it is not identified on a window ending in 2017; the pre-2018 dose-response, which has more power, carries the tariff check for that outcome (Appendix B.11).

details.

The licensing channel. The trade gain and the removal of the individual-license burden are visible together in the dynamics. The individual-license share is flat before the transfer and then declines gradually, reaching about ten points by the fifth year, while the export response is flat before and builds over the same horizon. The two move in step (Appendix Figure 5): trade rises as the individual-license burden falls. Neither margin is a sharp jump at the effective date; the licensing transition is itself a multi-year adjustment, through grandfathered licenses and contract cycles. It is the co-movement, not a discontinuity, that ties the trade gain to the licensing relief.

9 The estimate in general equilibrium

The reduced form of the previous section is silent on two questions. Is a 17-log-point export response the right order of magnitude for a licensing wedge of the size these data imply? And what is the response worth once wages and prices adjust? Both can be addressed by embedding the wedge in a

quantitative trade model and solving for the counterfactual equilibrium exactly, which also relaxes the partial-equilibrium restriction that the administrative-saving calculation of Section 14.2 carries.

Model. We use a multi-region, two-sector Armington model in exact hat algebra (Costinot and Rodríguez-Clare, 2014). There are twelve regions: the United States, the eight allied suppliers and customers of Section 11.1, China, Russia, and a residual rest of the world. There are two sectors, T for the 88 regime-gated treated products and O for all other goods. Demand is CES across origins within a sector, with elasticity σ_s , and Cobb–Douglas across sectors with country-specific shares $\alpha_{j,s}$. Labor is the single factor and trade deficits are held at their baseline levels. Write $\hat{x}$ for the ratio of a variable’s counterfactual value to its baseline value, $\lambda_{ij,s}$ for the baseline share of region j ’s spending on sector s sourced from i , Y_i for income, and D_j for the deficit. The counterfactual equilibrium solves

$$\pi'_{ij,s} = \frac{\lambda_{ij,s}(\hat{w}_i \hat{\tau}_{ij,s})^{1-\sigma_s}}{\sum_k \lambda_{kj,s}(\hat{w}_k \hat{\tau}_{kj,s})^{1-\sigma_s}}, \quad \hat{P}_{j,s} = \left[\sum_i \lambda_{ij,s}(\hat{w}_i \hat{\tau}_{ij,s})^{1-\sigma_s} \right]^{\frac{1}{1-\sigma_s}}, \quad (2)$$

together with sector budgets $E'_{j,s} = \alpha_{j,s}(\hat{w}_j Y_j + D_j)$ and market clearing

$$\hat{w}_i Y_i = \sum_s \sum_j \pi'_{ij,s} E'_{j,s}, \quad (3)$$

after which real income follows as

$$\hat{W}_j = \hat{w}_j / \prod_s \hat{P}_{j,s}^{\alpha_{j,s}}. \quad (4)$$

In words, the first equation says that spending shifts toward the origins whose delivered costs fall, at a rate governed by σ_s , and the price index aggregates those delivered costs; market clearing requires each region’s wage bill to equal what the world spends on its goods; and real income in (4) is the wage deflated by the price indexes, so the terms-of-trade improvement reported below is the gap between the two. Decontrol enters as a fall in the iceberg cost on U.S. exports of sector T , so that $\hat{\tau}_{US,j,T} = 1/(1 + \bar{\tau})$ for the ad-valorem wedge $\bar{\tau}$, applied either to every destination or to allied destinations alone. Baseline bilateral trade is BACI for 2012, the last full pre-reform year, in which U.S. exports of sector T are \$76.7 billion of \$1.46 trillion of total U.S. goods exports. At $\hat{\tau} = 1$ the solver returns $\hat{w} = \hat{W} = 1$ to machine precision, and the largest market-clearing residual across every run reported below is 1×10^{-13} .¹⁷

Elasticities and calibration. Two elasticities and one calibrated share do the work. For the treated sector we set $\sigma_T = 6.5$, the median Armington elasticity measured for these products in Appendix B.10, and carry their trade-weighted median of 10 as a sensitivity. For all other goods $\sigma_O = 4$, within the range of median estimates in Broda and Weinstein (2006). The U.S. home share in sector T is set to 0.85. That share is not observed at the product level, and it is calibrated from these products’ \$32 billion of annual exports against the \$400 to \$425 billion of shipments in the industries that make them (Section 13); because those industries are broader than the products, the figure is an assumption and results are reported from 0.75 to 0.95. Other regions’ home share in T is 0.5 and the goods share of expenditure is 0.5. The wedge is the object of Section 14.2: 2 to 3.5 percent as measured, and 8 to 13.5 percent when extrapolated to a fully licensed basis by the treated set’s mean dose.

¹⁷Script 122. Solved by multiplicative dampened iteration on the wage vector, with world tradable income as numeraire; an additive update diverges at $\sigma_T = 10$. Domestic absorption is not observed at the product level and is calibrated, as described below.

The export response the model implies. Solved forward at each candidate wedge, the model predicts the quantity the paper estimates, and the comparison disciplines which wedge can be carried downstream. At a wedge of 3.5 percent the model raises U.S. exports of the treated sector by 17.0 percent, against an estimated 17 log points, or 18.5 percent. At 2 percent it gives 9.5 percent, so the measured range brackets the estimate. The extrapolated range does not survive the same test. At 8 percent the model implies a 41 percent export response and at 13.5 percent a 74 percent response, two to four times what the data show. This is a consistency check and not an independent test, since the wedge is itself an inversion of the estimated effects through the same elasticity, and a wedge recovered that way should roughly reproduce the effect it came from. What the exercise adds is the verdict on the rescaling: multiplying the wedge by the inverse of the mean dose and then applying it to the whole sector counts the dose twice, because the estimated effect is already an average over a set that was 27 percent licensed, and the model makes the resulting overshoot explicit. The measured 2 to 3.5 percent is therefore the figure the general-equilibrium arithmetic supports, which is how Section 14.2 carries it. The elasticity is the sensitive input and it points the same way: at $\sigma_T = 10$ a 3.5 percent wedge implies a 28 percent export response, above what the data show, so $\sigma_T = 6.5$ paired with a wedge near 3.5 percent is the internally consistent combination.

Welfare. At the measured wedge the reform raises U.S. real income by 0.018 to 0.032 percent, which is \$2.9 to \$5.1 billion a year on 2012 GDP (Table 4, Panel A). The gain arrives as a terms-of-trade improvement: the U.S. nominal wage rises about a tenth of a percent at the 3.5 percent wedge while the U.S. consumption price index rises about seven hundredths, and equation (4) is the difference between them. Three features of the number matter for reading it. It scales with the model’s sector- T base, which is BACI’s reconciled \$76.7 billion rather than the \$32 billion these codes carry in the U.S. filing records, a gap of about 2.4 times that Appendix B.12 attributes to the redistribution of civil-aviation trade reported under a catch-all commodity code. The welfare figure therefore prices the same HS6-total object as the \$6 billion gross flow of Section 14.1 and inherits the same upper-bound reading: scaled to the filing-record base for these codes it would run about \$1.2 to \$2.1 billion a year. It is larger than the \$0.3 billion administrative saving of Section 14.2 because it is a wider object, counting the consumption and reallocation gains from lower delivery costs on all of the sector’s trade and not only the resource cost of paperwork on the \$9.0 billion that crossed individual State licenses. And it remains a gain from a trade-cost reduction, so it values at zero whatever the screening was buying, exactly as the break-even of Section 14.2 does.

Incidence on allied suppliers. Restricting the shock to allied destinations, the channel the license exception actually opened, cuts the U.S. gain by more than half, to \$1.3 to \$2.3 billion at the measured wedge. The model also prices what Section 11.1 finds in the data, that the U.S. gain registered against allied industrial suppliers. Those suppliers lose market share, and almost no welfare. Korea, the largest estimated loser of position, gives up 0.009 percent of real income at the measured wedge and 0.038 percent under the extrapolated one; the Netherlands loses 0.002 and 0.007 percent; Japan, Germany, and Italy are marginally positive throughout, gaining as consumers of cheaper U.S. goods roughly what they lose as competing producers. Two readings follow. The transfer between allied producers that Section 14.2 declines to count as an efficiency gain is real in market position and second-order in welfare, and an allied ledger of this reform is close to the U.S. ledger alone, which is the accounting the break-even assumes.

Table 4. General-equilibrium effects of the decontrol shock (exact hat algebra)

Wedge $\bar{\tau}$	United States			Real income, allied suppliers (%)				
	$\hat{W}_{US}$ (%)	\$bn/yr	$\Delta X_{US,T}$ (%)	KOR	JPN	DEU	ITA	NLD
<i>Panel A. All destinations</i>								
2.0% (measured)	+0.018	2.9	+9.5	−0.005	+0.002	+0.001	+0.001	−0.001
3.5% (measured)	+0.032	5.1	+17.0	−0.009	+0.003	+0.002	+0.002	−0.002
8.0% (extrapolated)	+0.076	12.4	+41.1	−0.021	+0.007	+0.006	+0.005	−0.004
13.5% (extrapolated)	+0.137	22.3	+73.7	−0.038	+0.013	+0.012	+0.010	−0.007
<i>Panel B. Allied destinations only</i>								
2.0% (measured)	+0.008	1.3	+4.3	+0.003	+0.002	+0.002	+0.001	−0.000
3.5% (measured)	+0.014	2.3	+7.6	+0.006	+0.003	+0.003	+0.002	−0.000
8.0% (extrapolated)	+0.034	5.5	+18.3	+0.014	+0.008	+0.008	+0.005	−0.000
13.5% (extrapolated)	+0.061	9.9	+32.8	+0.027	+0.016	+0.016	+0.010	+0.001

Twelve regions, two sectors, one factor; equations (2)–(4). Benchmark calibration: $\sigma_T = 6.5$, $\sigma_O = 4$, U.S. sector- T home share 0.85, other-region sector- T home share 0.5, goods expenditure share 0.5. Baseline trade is BACI 2012; GDP is World Bank WDI for 2012, so dollar figures are 2012 dollars. $\Delta X_{US,T}$ is the model’s U.S. export response in the treated sector, to be read against the estimated +0.170 log points, or +18.5 percent, of Table 2: the measured wedges bracket it and the extrapolated ones exceed it by two to four times. “Measured” and “extrapolated” distinguish the wedge as estimated from the wedge rescaled to a fully licensed basis by the mean dose of 0.27 (Section 14.2). Sensitivities, all at the 3.5 percent wedge and all destinations: $\sigma_T = 10$ gives +0.051 percent (\$8.4 billion) with a +28.2 percent export response; the U.S. home share from 0.75 to 0.95 gives \$5.2 to \$4.6 billion; the other-region home share from 0.25 to 0.75 and the goods share from 0.4 to 0.6 give \$4.2 to \$6.6 billion. Full grid in `122_ge_hat_algebra.csv`.

What the calculation does not settle. The model is a demand system with one factor, so it prices a delivery-cost reduction and cannot speak to entry, capacity, or the industrial-base arguments the reform was sold on, which is consistent with the absence of any measurable domestic response in Section 13. Its calibrated shares are assumptions and not measurements, though the U.S. home share matters little: across 0.75 to 0.95 the gain at the measured wedge moves only from \$5.2 to \$4.6 billion, and the widest excursion over the other-region home share and the goods expenditure share is \$4.2 to \$6.6 billion. The elasticity is where the real sensitivity sits, and it is bounded by the consistency check above. Nothing here changes the identification of the export response, which the next section takes up.

10 Identification

10.1 Timing exogeneity

The identifying variation is transfer timing, used through the not-yet-treated controls. Three checks bear on its exogeneity, each computed on the regime-gated treated set. First, the selection-on-decline concern, that items were decontrolled because their military relevance was fading, is not supported. Regime-gated treated products were not declining before transfer (mean 2009–2012 log-export slope +0.027), and their pre-trend is statistically indistinguishable from that of never-treated defense controls (−0.048; Welch $t = 0.87$, $p = 0.39$). The differential’s confidence interval is nonetheless wide: extrapolated over a five-year post window the point differential implies a cumulative bias of +0.376 with interval [−0.507, +1.259], larger than the ATT. We therefore print that arithmetic in Appendix B.4 and do not rest on the failure to reject. Second, among treated products the

transfer cohort is not predicted by the product’s pre-transfer trend (correlation -0.07 ; $p = 0.52$) or its pre-period S-code share ($p = 0.95$); it is predicted by size, with larger products transferring earlier ($p < 0.01$), the level correlation we return to below and that the unit fixed effects absorb. Third, a joint Wald test that the regime-gated pre-period event-study coefficients are zero is not rejected for either outcome (trade $\chi^2_2 = 1.52$, $p = 0.47$; share $\chi^2_2 = 1.74$, $p = 0.42$).

Beyond the data, the institutional record bears on how the timing was set, and what the contemporaneous documents say differs from what was said about them afterwards. The order was fixed years before the trade it might have tracked (Section 3), and the two determinants that appear in the record at the time concern neither trade nor markets. The December 2010 advance notice gives an administrative one, that the lists were “too big and complex to be reviewed in their entirety all at once” and were taken category by category “in order to make the project more manageable,” with military vehicles first because that category was the drafting template the rest would follow. Officials testifying in April 2013 described the remaining categories as either undergoing or awaiting interagency review. The strongest anchor is the statutory one: the section 38(f)(1) notification for the first tranche was filed in March 2013, before the rules published and seven months before they bound.

The interoperability criterion the reform is usually credited with belongs to a later and more qualified place in the record. It was stated plainly for the first time in February 2016, by the Deputy Assistant Secretary who owned the categories, under adversarial questioning about why firearms had not yet moved, and attributed back to Gates’s 2010 address, which does not contain it.¹⁸ We therefore do not claim the criterion was stated at the time, and we do not treat it as reassuring. Read at face value it cuts the other way: revising first the categories most central to operating alongside allied forces selects, by construction, on allied-market importance. That is a threat this design bounds, not a fact that helps it. The design nonetheless keeps it from driving the estimate. Unit fixed effects absorb the levels, and the controls are other defense products of similar strategic standing, so identification needs only that timing within the design is not set by product-level trade *trends*, which the balance tests above support and which Section 10.3 confronts directly with measured allied demand.

One category illustrates the point because its timing was set by something unrelated to the review queue entirely. Firearms and ammunition (Categories I–III) were held back inside the reform window for the customer-mix reason set out in Section 3, a property of who buys them and not of how they were selling. Their lateness has nothing to do with the early cohorts’ trade, which is why they serve as a clean not-yet-treated control.

Because the rollout was announced ahead of time, one might worry the response is anticipation of the announcement and not a reaction to the transfer. It is not. Re-centering each treated product on its proposed-rule year, the export path does not turn up at the announcement; it begins to build only after the rule takes effect, about two years later. The gradual response is adjustment after the rule binds, which is when the regime actually changes. The check has a control-side counterpart, since every control category eventually received its own proposed rule inside some cohort’s window. Controls do not break downward at their own announcements either—the one marginal coefficient is a positive share shift that would bias the estimate toward zero—and censoring control observations at their own announcement dates leaves the estimate intact (Appendix B.4).

¹⁸House Committee on Small Business, hearing of February 11, 2016, printed pp. 17–18 and 40, is the criterion’s first appearance in the record; it is absent from the seven Congressional Research Service vintages published between 2011 and 2014, entering that series only in the July 2017 edition, and Fergusson and Kerr (2020) pick the sentence up from the 2016 hearing.

Figure 6, in Appendix B.4, shows the staggered design one cohort at a time: within each transfer cohort the treated products’ exports and those of the contemporaneous not-yet-treated controls track each other before the cohort’s own rule date and part after it, and the divergence recurs at each cohort’s own event zero rather than at a common calendar year, the pattern a staggered treatment predicts and a common defense-sector shock does not.

10.2 Sensitivity to pre-trend violations

The pre-period coefficients are small and jointly insignificant, but a formal sensitivity analysis is more informative than a pre-test. Under the relative-magnitude restriction of Rambachan and Roth (2023), the regime-gated effect stays significant for post-period trend violations up to about $\bar{M} = 0.3$ of the largest pre-period deviation on each of exports and world share; beyond that the robust confidence set includes zero, and a smoothness restriction is no more favorable, breaking down sooner (Appendix B.9). This is a modest breakdown, and it quantifies the identifying assumption without removing it: with two clean pre-periods and a six-year building post-window, pre-trend flatness alone cannot carry the design.

The same exercise applied to the specification that does most of the work against the demand objection, the supplier-differenced triple difference of Section 10.3, comes out better and not worse. Its pre-period deviations are somewhat larger (max $|\text{pre}|$ 0.060) but its average post effect is larger still (+0.209, a ratio of 3.5), and the robust confidence set excludes zero out to $\bar{M} = 0.4$, above the breakdown of either headline estimate: the specification that answers the demand objection tolerates more pre-trend violation than the results it defends.¹⁹ A pre-trend is not fragility on its own; what matters is its size against the effect it might explain.

Identification therefore rests on the ex-ante announced rollout schedule set out above; the cross-sectional dose-response of Section 12.1 corroborates it without using the event-time path at all.

10.3 Category-level demand

The confound the timing evidence cannot rule out is a demand shock concentrated in the categories that transferred first; the leading candidate is an allied-rearmament wave in the interoperability-central items after 2014. Three tests bound how much of the reduced form such a confound could carry: one absorbs it, one measures it, and one differences it out across suppliers.

The first absorbs it. Adding chapter-by-year fixed effects to the stacked design removes any demand shock common to a product category, so the effect is identified only from within-chapter treated-versus-control variation. Under this demanding specification the export ATT is +0.14 and the world-share ATT roughly halves to +0.08, neither significant at conventional levels ($t = 1.6$ and 1.8). The loss of precision is partly mechanical: the aircraft family, half the dollar dose, is almost entirely treated and so contributes little within-chapter identification.

The second measures the confound instead of absorbing it. From SIPRI’s arms-transfer registers we build category-by-year delivery volumes, globally and to the A:5 allies, mapped to the 600-series families through the regulation-text crosswalk, covering 93.8 percent of the dollar dose. Two facts emerge (Appendix Table 11). First, the premise of the demand story is only half right: allied deliveries of the early-transferred categories were roughly flat after 2014, about +6 percent, and

¹⁹Identical at every fixed-effect rung, including product-by-year absorption, which is the numerical signature of the orthogonality discussed in Section 10.3. Script 100.

the early-versus-late gradient is driven by a collapse of allied procurement in the late and control categories. It is not monotone in transfer order, armoured vehicles being a 2014 cohort that fell as hard as the controls, and within the early categories allied purchases tilted toward non-U.S. suppliers, the opposite of the buy-American tilt the confound requires. Second, controlling for the measured demand path leaves the effect standing: the preferred specification, world plus allied non-U.S.-supplied deliveries of the unit’s own category, gives +0.142 on exports and +0.134 on the share, each surviving wild-cluster inference at the one percent level with flat pre-coefficients. Those rows use the full sample, imputing the category covariate for the units SIPRI does not map; on the SIPRI-mapped subsample alone the same-sample export baseline is an uninformative -0.098 (SE 0.234, $N = 964$, 64 treated codes), so the SIPRI control bounds the confound on the sample it covers and does not re-establish the headline there. The demand rebuttal that does not depend on SIPRI’s coverage is the product-level one: controlling for world-excluding-U.S. exports of each HS6, a demand measure available for every treated product and exogenous to U.S. decontrol, leaves the export ATT at +0.229 and the share at +0.149, both significant at one percent. Even a deliberate over-control that includes U.S. deliveries to allies, part of the outcome itself, leaves +0.099 ($t = 1.7$) and +0.117 ($t = 2.4$). Appendix B.5 reports the ladder and the two features of the SIPRI control that qualify how far it reaches.

The third test differences the confound out along the supplier dimension. A demand shock to a product category in a year raises imports of that category from every supplier of it, so it cancels in a within-product comparison of U.S. against non-U.S. exports of the same HS6. Stacking that comparison on the baseline windows and forming the triple difference, treated versus control in the U.S.-versus-non-U.S. export gap and before versus after transfer, gives +0.188 (SE 0.055, $t = 3.4$). The estimate is numerically unchanged whether the specification absorbs demand at the cohort-by-year, the chapter-by-year, or the product-by-year level, which is the signature of the orthogonality it exploits, and restricting it to non-aircraft treated codes leaves it at +0.188 ($t = 3.3$), ruling out the concern that Airbus and Embraer civilian-aviation comparators drive the result. It keeps the within-category power that chapter-by-year absorption gives up and does not depend on SIPRI’s product coverage, and its own pre-period tolerates violations up to $\bar{M} = 0.4$, above the breakdown of either headline estimate. Its content is the same avoided-decline structure—a U.S.-specific decline in the control products, not an absolute rise in the treated ones. That is, it is not an independent identification leg but the world-share result read against demand.

What none of the three differences out is a demand shift that is both graded by category and specific to U.S. suppliers in the early-transferred categories. That residual is narrower than the general confound and it runs against the SIPRI evidence above, in which allied purchases in those categories tilted toward non-U.S. suppliers. Its concrete form is the Foreign Military Sales channel, supplied only by U.S. firms and so invisible to the supplier difference, which is why Section 11.3 takes it up and why the commercial-trade estimate is the headline. A separate concern, treated-set contamination and the absence of a clean same-tier control, is taken up in Appendix B.6: the headline is robust to dropping the eleven miscoded arms codes, but no public-data control isolates the sensitive items retained *within* the transferred categories, which is the reason the within-category intensity evidence of Section 12 carries weight.

10.4 Remaining limits of the design

One concern is now addressed, one remains. The regime gate disciplines treatment with the trustworthy license-code field, while candidate-set membership had used the 600-series ECCN, a field that over-populates relative to the codes filers actually use. Rebuilding membership from a

regulation-text crosswalk, the HS headings the ECR 600-series final rules cover and no ECCN data field, drops a large stray-code contamination at the candidate stage (538 of the 1,179 ECCN-eligible HS6 lay outside any defense chapter) and leaves the regime-gated estimate unchanged; a version that takes both membership and transfer timing from the regulation still returns +0.20 on exports ($t = 2.6$) and +0.11 on world share ($t = 2.0$), the export estimate carrying a larger pre-period lead than the baseline (0.12 against 0.02). The identification is carried by the pre-2013 license gate; the candidate trigger only scopes the set.

The remaining limit concerns the pre-reform measurement of the treated good. Before 2013 the 600-series ECCNs do not exist, so the treated goods appear in the Census panel as ITAR individual-license exports at the HS6 level—the same field the regime gate uses. These S-code exports are the available pre-reform proxy for military trade in the treated HS6, and they are what the gate is built on: at the 5 percent threshold the pre-reform baseline is at least 5 percent individually licensed, at 10 percent at least 10 percent, and so on. The limit is that the HS6 outcome used in the DiD is the product total, not the S-code slice alone, because no consistent “military-content” series spans the reform: the natural candidate, summing S-code (pre) and 600-series (post) value, is mechanically dominated by the code switch—every shipment that previously carried an S-code license relabels to a 600-series ECCN at transfer, so the post-reform sum includes the pre-reform S-code flow by relabeling rather than by trade growth, and the two cannot be separated without a counterfactual. The product-level outcome therefore measures the treated good plus any co-moving civilian trade in the same HS6.

Several strands of evidence bear on how much that matters. Section 11.1 indicates the gain tracks defense-trade destinations and does not spread as civilian trade would; the supplier- differenced bound of Section 10.3 rules out demand shocks common to U.S. and non-U.S. suppliers of the same product. The destination triple difference of Section 11.2 points the same way, significant on both bilateral sources and larger on the U.S. filing records that sit closer to the military variant, though its pre-period depends on a sample rule that narrows the comparison toward continuously served markets. The gate sweep of Appendix B.2 bounds the concern more directly: at gates where the pre-reform baseline is majority military by construction the effect does not weaken, and a civilian co-movement large enough to account for it would have to exceed 40 log points above controls, implausible given the supplier difference and the defense-customer destination concentration. The residual concern is a differential shock specific to U.S. civilian exports of the transferred categories, narrower than the general confound.

11 Composition of the gain: markets, suppliers, and channels

The estimates so far average over products and destinations. Three cuts of the same variation say where the gain actually arrived: out of whose market share, into which markets, and through which contracting channel. The first two bear on the question Section 10.4 leaves open, whether the good that moved is the decontrolled military variant or civilian trade in the same code. We take the supplier comparison first, because it is the better identified of the two: it removes demand for the product in the market by construction, and the sample rule it needs costs it little. The destination test follows, resting on the same pre-window sample rule at a higher price. The third cut bears on whether licensing relief is the operative mechanism, and it is where the dollar magnitude is least well attributed.

11.1 Market position against rival suppliers

The Census panel behind Section 8 tracks the treated good but sees only one side of the market. Reconciled bilateral trade sees all of it, at the cost of resolving only the HS6. Used for what it can do, it answers a question the Census data cannot: whether the United States gained position against the other suppliers selling the same product into the same market.

We build an exporter-by-importer-by-HS6-by-year panel from BACI covering 2009–2019, 31 exporters and 228 importers across 1,682 products, about 91 percent of world trade in them. Collapsing each cell to the United States against the sum of all other suppliers makes the importer-by-product-by-year fixed effect analytic: with two sources per cell, absorbing it is algebraically the same as taking the within-cell difference, so the outcome is $\ln(\text{U.S.}) - \ln(\text{rivals})$ and demand for that product in that market and year is removed by construction. The stacked design, treated set, and cohorts are the paper’s own. One sample rule matters. Cells are required to be observed in all three pre-treatment years of their own sub-experiment, a restriction on pre-treatment presence that selects on nothing realized after transfer. Without it the deep-lead coefficient is estimated on a thinner set of cells than the reference period, which is what generates an apparent pre-trend in this panel.²⁰

In Table 5 we collect the estimates of this section. Pooled, the United States gains +0.116 against rival suppliers of the same product in the same market (SE 0.023, $t = 4.96$, wild-cluster $p < 0.001$), with a maximum pre-period deviation of 0.024 and a Rambachan–Roth breakdown of 0.30, matching the world-share headline. The point estimate is stable restricting controls to the same HS4 (+0.106, breakdown 0.30), dropping the Chapter-93 contamination (+0.105, breakdown 0.30), and dropping aircraft and engines (+0.116). The last of those fails the reading rule of Section 7 on both conditions ($\max|\text{pre}|$ 0.074, breakdown 0.10), so it bears on the direction and not on the magnitude, which is the same standing we give the non-A:5 destination cut below, whose pre-period stands in nearly the same ratio to its effect. Because treatment is assigned at the product-category level, we subject it to the same cluster coarsening the headline faces in Appendix B.8: the wild-cluster p runs 0.000, 0.000 and 0.006 as clusters go from HS6 to HS4 to the 81 HS2 chapters, and the allied cut runs 0.001, 0.002 and 0.012.²¹ By destination the gain is twice as large in the STA-eligible allied markets, +0.141 ($t = 3.66$), and that is the only destination cut whose pre-period supports it: its breakdown of 0.40 exceeds that of either headline estimate, while the non-A:5 arms buyers (+0.131) carry a pre-period deviation more than half the size of their effect and other markets (+0.071) tolerate no violation at all. Section 11.2 reaches the same allied concentration on a different contrast, comparing destinations and not suppliers, and on a treated set the same sample rule thins considerably further, so the two corroborate each other without being nested.

Holding the cell set fixed across rivals, so that rows are comparable and only the supplier in the denominator changes, the U.S. gain is against Korea (+0.285), Japan (+0.165), the Netherlands (+0.129), Italy (+0.113), and Germany (+0.102), each surviving a restricted wild cluster bootstrap at the one percent level with a breakdown of at least 0.20. It is not against the closest defense-industrial partners: the United Kingdom does not survive the fixed-cell restriction, and France and Canada

²⁰Weighting cells by pre-reform size also flattens the pre-period, but it does so by discarding power: across a ladder of weighting schemes the maximum pre-period deviation falls in lockstep with the Kish effective sample size, which reaches 0.96 percent of the nominal count at the most aggressive rung. We do not use it. Scripts 97–99.

²¹Script 105. The sample rule also has a natural alternative. Requiring units to be present in every event year of their sub-experiment, rather than only in the three pre-treatment years, gives +0.089 (SE 0.025, wild-cluster $p = 0.001$, and $p = 0.022$ under chapter clustering) on a third of the observations. We prefer pre-window completeness because it conditions on nothing realized after transfer, but the balanced estimate is the smaller of the two and is reported here, not left in the code.

fail outright. Nor is it against Russia or China, whose estimates tolerate no pre-period violation at all. Such displacement as the data show runs against second-tier allied industrial suppliers, not against strategic rivals.

The word “gain” in these rows is relative, and the supplier-level decomposition of Appendix B.4 says which side of the contrast moved: the U.S.-versus-foreign gap on the treated products is roughly flat after transfer, while the gap on the still-controlled products falls. The coefficients therefore do not record the United States taking sales from these suppliers in the decontrolled goods. They record the United States continuing to cede position in the goods it kept licensing above its rivals’ national standards, and ceasing to cede it where control was lifted. The competitive cost is borne on the products still over-controlled, which is the over-compliance reading below stated in levels.

Read against Section 3, what these rows price is the competitive cost of unilateral over-compliance: what a supplier gives up, in market position against rivals licensing the same multilaterally listed goods under their own national rules, by licensing above them. In supplier-competition models the relative stringency of national regimes is what decides who captures a contested sale (Levine and Smith, 2000; García-Alonso and Levine, 2007); this puts a number on that, and the cost was paid to Korea, Japan, the Netherlands, Italy and Germany rather than to Russia or China. No product-level estimate of it has appeared elsewhere; Fan et al. (2026) establish an aggregate counterpart, documenting that joint membership in multilateral export control regimes raises bilateral trade substantially. The comparison assumes rival suppliers’ own regimes and export postures were stationary, which the design does not test. The failures of that assumption do not share a direction: European small-arms and Middle East licensing tightened in these years, inflating the affected rival coefficients, while Japan’s April 2014 Three Principles replaced a near-total export ban mid-window and Korea’s state-backed export expansion ran through the same years, working the opposite way on exactly the two largest coefficients. We therefore read the cross-rival ranking at its sign only.

These coefficients also record a private competitive return: whether the stricter treatment the United States gave up was buying security the multilateral baseline does not is the question Section 14 states as a threshold instead of answering.

The extensive margin is a precise null, and it does two jobs. Conditioning on the United States selling into a cell drops 946,089 cell-years, 32 percent of all cells with a market. Adding them back with an inverse-hyperbolic-sine share defined at zero leaves the pattern intact (pooled +0.077; A:5 allies +0.090), and a linear probability model for whether the United States sells at all returns -0.003 with a standard error of 0.003. There was no entry response. The gain is deeper sales into relationships that already existed, the margin the per-order reading of Section 4 implies, and the absence of entry is also why conditioning on positive flows does not bias the estimates above.

The evidence has two limits. The dollar-weighted version of the result is not established: Poisson estimation on levels, which retains the zero cells, returns +0.060 with a standard error of 0.053. What the bilateral panel supports is a statement about relative position within a market and product, and we state it only that way, never as an aggregate dollar claim. The more serious limit is that this evidence locates the gain without identifying the good. The effect does not scale with the military content of the code, and the highest-content bin, though large, rests on thirteen products under any sample rule and carries a pre-period deviation larger than its own effect. Reconciled data can say that the gain landed in allied markets and registered against allied suppliers’ position; they cannot say that the good which moved was the decontrolled military variant. That claim continues to rest on the Census evidence and the licensing gate.

The obvious way to remove it is not available. A supplier comparison built on SIPRI’s arms-transfer register would be immune to the civilian co-movement that defeats the dose test, because

the register records military transfers only; but it records complete major weapons, and the reform retained the most sensitive complete platforms while moving less sensitive items and components, so 91 percent of the transferred-category value it carries is whole aircraft, armoured vehicles and ships. A military-only source composed mostly of the platforms the reform kept cannot check a result about the components it moved, and the exercise we ran on the four rivals with enough deals bounds nothing in either direction.

11.2 Destination evidence on the treated good

A second cut asks the same question across destinations, not across suppliers. In a triple difference with product-by-year and destination-by-year fixed effects, the differential gain to allied and major-arms-customer destinations is +0.39 (SE 0.11, $t = 3.6$) on reconciled third-party trade—an inverse-hyperbolic-sine coefficient, read as log points—and the estimate barely moves whether the pre-window-completeness rule of Section 11.1 is imposed or not. The U.S. filing records give +0.73 (SE 0.14, $t = 5.2$) on the same specification, and they are the source closer to the decontrolled good—within the treated codes, aircraft and engines are 30 percent of their treated value against 75 percent in the reconciled data—but their estimate rises from +0.27 when the pre-window rule is imposed, so we read it as corroboration on the more military-concentrated source, an estimate for continuously served trade relationships and not the headline. Under the rule both estimates hold a largest pre-treatment coefficient below 0.03; both clear a wild cluster bootstrap at better than the one-percent level; and both rise rather than fall when the fighter-program destinations are excluded, alone or together with the Gulf states. The gain concentrated in the markets that buy U.S. military goods, which is hard to reconcile with civilian co-movement in the same code. Appendix B.13 reports both sources, the sample rule, and what the rule selects on.

11.3 Contracting channels

The Census license-type field records how each shipment was authorized. Measured as the 2014–2018 average against the 2012 baseline, the treated products’ \$4.3 billion annual increase decomposes by contracting channel. Direct commercial sales, the net change in pre-reform ITAR-individual value plus post-reform 600-series value, account for \$0.2 billion, 5 percent. The Strategic Trade Authorization exception (license type C59), created by the reform and essentially unused beforehand because the exception did not exist for ITAR-regulated items, accounts for \$0.4 billion, 9 percent; its rise from near zero is the clearest reform-specific signal in the license record, since the exception applied to 600-series items under the EAR but not to ITAR-licensed goods. Together these reform-attributable channels are \$0.6 billion, 14 percent of the increase. Foreign Military Sales (license type S94), government-to-government procurement routed through the Defense Security Cooperation Agency exclusively to U.S. firms, account for \$2.7 billion, 64 percent. The remaining \$0.9 billion, about 22 percent, is other commercial EAR-authorized value, which was not individually licensed and enters the dose analysis below.

That concentration bears directly on the licensing interpretation, and it has to be read for what it is. The 64 percent is a share of the *dollar* change, and the dollar change is dominated by a small number of large, lumpy government contracts. The estimate the paper reports is a log effect on product-level exports, and removing the FMS channel entirely barely moves it: the same stacked design run on all non-FMS value returns +0.165 (SE 0.048, max |pre| 0.032, wild-cluster $p = 0.0005$) against +0.170 on the total. The two statistics answer different questions, and only one of them is the paper’s estimate. The headline export effect does not rest on Foreign Military Sales; the dollar

Table 5. U.S. market position against rival suppliers of the same product

	Estimate	(SE)	max pre	Breakdown
<i>Panel A. Pooled and by destination</i>				
All treated products	+0.116	(0.023)	0.024	0.30
A:5 allied markets	+0.141	(0.039)	0.022	0.40
Non-A:5 arms buyers	+0.131	(0.035)	0.082	0.05
Other markets	+0.071	(0.036)	0.013	0.00
<i>Panel B. Robustness of the pooled estimate</i>				
Same-HS4 controls only	+0.106	(0.029)	0.031	0.30
Dropping chapter 93	+0.105	(0.023)	0.023	0.30
Dropping aircraft and engines	+0.116	(0.025)	0.074	0.10
Balanced-sample alternative	+0.089	(0.025)	—	—
<i>Panel C. Fixed-cell rivals^a</i>				
vs. Korea	+0.285	(0.075)	0.105	0.30
vs. Japan	+0.165	(0.052)	0.043	0.50
vs. Netherlands	+0.129	(0.044)	0.079	0.30
vs. Italy	+0.113	(0.038)	0.068	0.30
vs. Germany	+0.102	(0.033)	0.025	0.20
<i>Not surviving the bootstrap or breakdown criterion:</i>				
vs. United Kingdom	+0.067	(0.044)	0.051	0.05
vs. France	+0.083	(0.051)	0.158	0.20
vs. Canada	−0.097	(0.042)	0.096	0.00
vs. China	+0.116	(0.046)	0.089	0.00
vs. Russia	+0.114	(0.081)	0.132	0.00
<i>Panel D. Extensive margin^b</i>				
Share at zero, pooled	+0.077	(0.016)	0.030	0.10
Share at zero, A:5 allies	+0.090	(0.026)	0.008	0.40
U.S. presence (LPM)	−0.003	(0.003)	0.007	0.00
Poisson on levels	+0.060	(0.053)	—	—

Outcome is $\ln(\text{U.S.}) - \ln(\text{all other suppliers})$ within an importer $\times$ HS6 $\times$ year cell, so importer-by-product-by-year demand is differenced out by construction. Stacked cohort design with the pre-window-completeness sample rule; standard errors clustered on HS6. “Breakdown” is the Rambachan–Roth relative-magnitudes value at which the robust confidence set first admits zero; the paper’s own headline breakdowns are 0.30 on both outcomes. Every estimate significant at conventional levels here also survives a restricted wild cluster bootstrap at $p \leq 0.032$, and the pooled and A:5 cuts survive coarsening the cluster to the 81 HS2 chapters ($p = 0.006$ and $p = 0.012$). Scripts 97–99, 105.

^a The (importer, HS6) cell set is held constant across rivals, so only the supplier in the denominator changes: 131,920 of 331,402 cells. The first five rows survive the wild cluster bootstrap at the one percent level with a breakdown of at least 0.20; the rows below the divider fail on one criterion or both.

^b Conditioning on a positive U.S. flow drops 946,089 cell-years, 32 percent of cells with a market. The share-at-zero rows use an inverse hyperbolic sine of the U.S. share defined at zero; the presence row is a linear probability model for whether the United States sells in the cell at all. Poisson estimation on levels retains the zero cells and is reported to show that the dollar-weighted version of the result is not established.

increase largely does.²²

²²The channel-level difference-in-differences on FMS (+0.76) and on direct commercial sales (+0.48) carry pre-period deviations of 0.30 and 0.15, and are reported as descriptive magnitudes only. The remaining channel, commercial and other EAR-authorized value, returns +0.067 (SE 0.096, wild-cluster $p = 0.53$) on the cleanest pre-period of the

Whether the FMS rise is itself a consequence of the reform is a question the evidence here does not close. The data establish only that the treated products' FMS value grew faster than both its own envelope and its controls: from 2012 to 2016 it rose 85 percent, against 18 percent for worldwide FMS deliveries and a 17 percent decline for the control products, a ranking a general expansion of allied procurement does not predict.²³ The test that would settle it lacks the power to. Restricting to bilateral FMS flows and absorbing destination-by-year fixed effects compares treated against control defense products bought by the same government in the same year, which no country-level rearmament path survives; on 7,683 product-destination-year cells that comparison returns +0.02 with a standard error of 0.18 and a pre-period deviation of 0.61. It cannot distinguish zero from the aggregate FMS rise, and we report it as uninformative and not as a null. Because the channel cannot be cleared, the estimates we lead with are the commercial ones, and the dollar figures of Section 14.1 are read net of it.

12 Mechanism and intensity

If licensing relief is what moved trade, the effect should scale with how much licensing each product bore, and it should arrive on the transaction margin the license actually taxed. Both are testable on this variation and neither uses the event-time path that the sensitivity analysis of Section 10.2 constrains. Together they say whether the mechanism the framework names is the one visible in the data.

12.1 A cross-sectional dose-response

A confound that adjusts as smoothly as the effect does is hard to rule out on the event-time path, which is what makes the sensitivity analysis conservative. Treatment intensity supplies the alternative. Among the products that transferred to a 600-series code, the amount of decontrol varies: some shipped a large share of their pre-reform value under individual ITAR licenses, others almost none. If licensing relief drives the trade gain, the post-versus-pre change in exports should rise with that pre-reform dose, within the transferred set and within sector.

It does. Regressing the long difference in log exports, the mean over the second through fifth years after transfer minus the mean over the four years before, on the pre-reform individual-license share gives a slope of +0.48 ($t = 2.6$) with chapter fixed effects and heteroskedasticity-robust standard errors ($N = 1,157$ transferred products), and +0.51 dropping aircraft; for world share the slope is +0.46 ($t = 3.7$; $N = 608$). Both slopes survive inference that does not lean on the heteroskedasticity-robust standard error: clustering on chapter tightens them instead of widening them ($t = 3.0$ for exports across 65 chapters, $t = 5.5$ for world share across 7), and a within-chapter permutation test that reshuffles the dose among products of the same chapter returns $p = 0.014$

four (0.056); that this channel carries the dose gradient below while showing no channel-level effect of its own has a natural reading in dilution, since commercial and other value is 98.5 percent of these codes' trade, so its channel-level estimate approximates the diluted product total. The channel-level STA coefficient is mechanical, since that channel is essentially zero before transfer by construction, and we never read it as an effect. Dollar components in the text are rounded to the nearest \$0.1 billion and sum to the \$4.3 billion total before rounding. Script 102.

²³The Defense Security Cooperation Agency's worldwide deliveries series breaks at FY2017, growing 69 percent year-on-year where every other year of the window falls between -2 and $+10$ percent, so the comparison is stated to FY2016. DSCA reports these aggregates by recipient country and fiscal year with no Munitions-List category breakdown, so they bound the channel's scale but cannot isolate the transferred goods. Source: DSCA Historical Facts Book, FY2017 edition.

for exports and $p < 0.001$ for world share. Dropping exports to Canada leaves the export slope essentially unchanged (+0.46, permutation $p = 0.024$), so the dose is not an artifact of exemption trade dropping out of the license numerator.²⁴ The gradient is monotone across dose quartiles (Figure 4), and its levels matter for the reading: most quartile long differences are negative (exports run from -0.13 in the bottom dose quartile to $+0.09$ in the top; world share from -0.23 to essentially zero), so the gradient says high-dose products held up or gained while low-dose transferred products declined, the same avoided-decline structure as the raw means of Section 8. The cross-sectional analog of a pre-trend test is clean: the dose does not predict pre-period growth ($+0.01$, $t = 0.1$ for exports; -0.10 , $t = -0.8$ for world share). Because this compares levels across products of differing treatment intensity and traces no event-time path, it does not rest on the dynamics that make honest-DiD conservative. Its identifying assumption, that the pre-reform license dose is unrelated to differential post-period demand within sector, differs from the timing exogeneity the reduced form relies on. The channel-by-channel evidence of Section 11.3 bounds how far the licensing reading can be pushed, and the dose-response is consistent with licensing relief without isolating it from allied procurement growth concentrated in the same high-dose, interoperability-central products.

That growth has a leading candidate: an allied-rearmament demand shock concentrated, like the rollout order itself, in the interoperability-central categories that transferred first; Section 10.3 measures that candidate directly with SIPRI delivery data, and the measured version removes between a sixth and two-fifths of the export effect, with the preferred specification significant at one percent. The design bounds it further in two ways. The dose variation is within-sector, so the confound must act through licensing intensity and not through category membership, and the gain is present in the fire-control cohort, whose rule took effect at the end of 2016, after the post-2014 demand shift was under way.

A direct test of the dose-response’s identifying assumption is consistent with it but underpowered: latent allied demand for each transferred product, measured from non-U.S. suppliers’ sales of the same HS6 and so exogenous to U.S. decontrol, does not rise significantly with the dose within chapter, but its confidence interval admits a gradient large enough to generate the observed effect.²⁵

The placebo (dose does not predict pre-period growth) and this latent-demand test are necessary conditions the data pass, neither sufficient. The dose-response draws on the same pre-reform license-intensity variation as the reduced form, so it is that variation viewed cross-sectionally and not a second identification leg; the ex-ante rollout schedule of Section 10 remains the one genuinely independent leg. It adds that the gain scales with the licensing burden the transfer removed, which is also what the dilution pattern of Table 2 implied.

Magnitude and a second intensity margin. Two extensions of the dose-response, developed in Appendix B.10, support the reading. Benchmarked through a trade elasticity measured for these products, the slope implies an ad-valorem equivalent of roughly 8 to 9 percent, within the range estimated for non-tariff measures generally (Kee et al., 2009; Hummels and Schaur, 2013). And a distinct intensity margin, how *often* the licensing apparatus bound per dollar of trade, yields a

²⁴Some traded categories shipped under the ITAR §126.5 exemption and are mechanically under-dosed, since the license gate counts only individually adjudicated authorizations. Category VIII(a) aircraft, the dominant treated category, was excluded from §126.5, so those Canadian exports appear in the license numerator, and the affected non-aircraft portion is roughly ten percent of the pre-reform treated base.

²⁵The point estimates are $+1.03$ ($t = 0.89$) in the export sample and $+0.70$ ($t = 0.59$) in the share sample. Adding the latent-demand measure as a control leaves the dose-response essentially unchanged ($+0.48 \rightarrow +0.47$ on exports, unchanged at $+0.46$ on world share), but that is a weak check: the control is orthogonal to the dose by construction of the first result, so it could not have moved the coefficient.

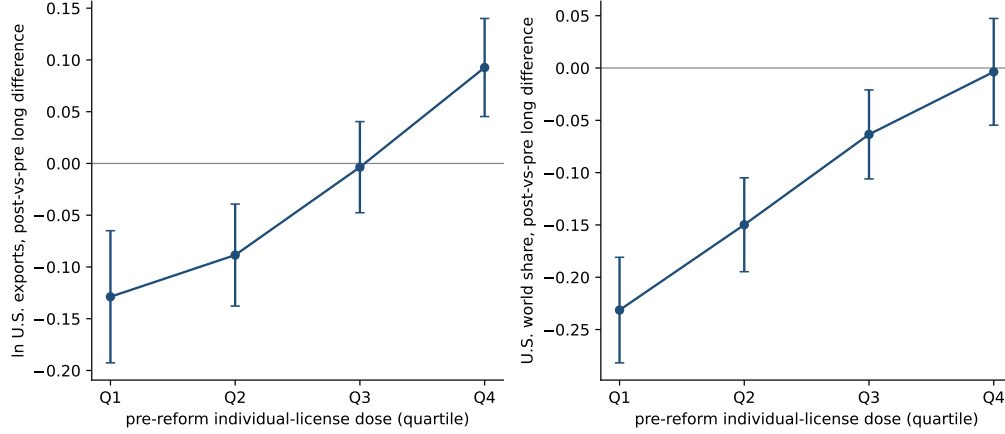


Figure 4. Cross-sectional dose-response, transferred products (ever-600, cohorts 2013–2017). Mean post-versus-pre long difference by quartile of the pre-reform individual-license share (the decontrol dose). Left: ln U.S. exports; right: U.S. world share. Both rise monotonically with dose, with quartile levels mostly negative, so the gradient reads as high-dose products declining less or gaining while low-dose products fell; the dose does not predict pre-period growth (placebo in text). 95% confidence intervals.

second placebo-clean world-share gradient orthogonal to the license dose. That gradient describes which products gained position; it is separate from the question of which transaction margin the gain arrived on, a distinction Appendix B.10 sets out.

12.2 The transaction margin

The Census filing records report a shipment count alongside value, so the estimated gain splits exactly into a transaction-count margin and a transaction-size margin: $\ln(\text{value}) = \ln(\text{shipments}) + \ln(\text{value}/\text{shipments})$, and the same stacked design run on each piece must add back to the total. It does, to six decimals: the +0.1703 on value splits into +0.0435 more shipments (SE 0.035, wild-cluster $p = 0.23$) and +0.1268 larger ones (SE 0.038, $p = 0.002$). Roughly three quarters of the gain arrives through transaction size, and the count margin is both small and imprecisely estimated, with a pre-period deviation (0.064) larger than its post-transfer effect. By the discriminating prediction of Section 4, that favors relief on the per-order authorization over relief on the per-consignment cost: per-shipment administrative costs predict de-bundling into more and smaller shipments (Hornok and Koren, 2015a,b; Kropf and Sauré, 2014), and it is the size margin that carries the gain here. The asymmetry is worth stating precisely. The size margin is significant on every cut we report; the count margin is not signed, and its confidence interval reaches +0.11, which would be two thirds of the total effect. So the data select the per-order reading without rejecting the per-consignment one, and a sharper test would need a margin the count cannot supply—consignments per authorization, or the number of distinct approved consignees.

The pattern is not a reporting artifact. Shipment counts show no break at the effective date (+0.031 at $k = 0$), the size margin builds over the same multi-year horizon as value instead of jumping, and 600-series items require electronic export filing regardless of value exactly as Munitions List items did, so the filing threshold does not change at transfer.²⁶ The same asymmetry appears in the destinations where the gain concentrates: in the STA-eligible allied markets the size margin

²⁶15 CFR 758.1(b). The value margin estimated on this independently constructed panel returns +0.170, reproducing Table 2. Script 101.

is +0.143 against +0.088 elsewhere. It is also consistent with the absence of any entry response in Section 11.1: relief that raises the value a single authorization can carry deepens existing relationships and does not open new ones.

One composition threat to the decomposition is serious. Foreign Military Sales shipments are large and lumpy, and their share of these products' value did rise across the reform, from 15.9 percent before to 20.6 percent after. A shift toward that channel would raise average shipment size arithmetically, with no change in any exporter's transaction behavior, and would reproduce the result above with no per-order content at all. Removing the channel settles it. Estimated on non-FMS value alone, the decomposition returns +0.158 on value, +0.044 on the count margin (SE 0.037, wild-cluster $p = 0.24$) and +0.114 on transaction size (SE 0.035, $p = 0.001$), leaving 72 percent of the gain on size against 74 percent on the full sample, and on a cleaner pre-period than the full sample gives (max $|pre|$ 0.018 on value against 0.028).²⁷ The size margin is therefore not an artifact of the shifting FMS share, and the per-order reading is earned on the commercial trade where an individual license actually bound. Nor is it a property of the aircraft cohort: splitting the treated set at the aircraft and engine families, where the platform programs ran and the lumpy contracts sit, the count margin is a null and the size margin significant in each half separately.²⁸

13 Effects beyond trade

The trade result raises a second question, whether decontrol leaves a mark on the industries that make these products, on their innovation, or on where the goods end up. Nothing beyond trade is detectable, and at this aggregation these are bounds set by data resolution, not demonstrated absences. Mapping the treated and control HS6 to NAICS6 yields 22 treated and 120 never-treated manufacturing industries, in which employment (−0.04), payroll (−0.03), value of shipments (+0.03), value added (+0.04) and patenting (+0.03) are each indistinguishable from zero, on confidence intervals that exclude gains above about three percent for employment, fifteen to seventeen percent for output, and twelve percent for patents. Capital expenditure is the one positive (+0.32) and is not identified, its only pre-period being the 2012 Economic Census at the post-sequestration budget trough. On where the goods went, comparing exports to re-export and sanctions-exposed hubs against exports to non-hub allies finds no diversion signature on either value or the count of hubs served, though a positive control on the 2020 firearms transfer shows the value test is underpowered, and the public end-use-monitoring series cannot attribute checks to the transferred categories at all. Appendix B.14 reports each of these.

The arithmetic explains the pattern and is the reason to read the nulls as bounds. The export gain is roughly \$6 billion a year against some \$400–425 billion in treated-industry shipments, about 1.5 percent, so even if every additional export dollar were new production the implied industry-level response would sit inside the confidence bands above. What the nulls establish is the absence of a

²⁷The FMS channel taken alone is uninformative on every margin, with point estimates that cannot be signed and pre-period deviations of 0.29 to 0.36; the same asymmetry appears in the channel-level estimates of Section 11.3. Within non-FMS value the size margin is +0.126 ($p = 0.002$) in STA-eligible allied markets against +0.069 elsewhere, reproducing the allied concentration. The value effect here, +0.158, differs from the +0.165 of Section 11.3 only in the estimation sample: the decomposition requires cells with positive non-FMS shipments, while the channel estimate retains the few cell-years with none. Script 104.

²⁸On non-FMS value, size carries close to nine tenths of the 40 program-family products' gain (+0.166, $p = 0.012$, against a count margin of +0.020, $p = 0.74$) and about half of the other 48's (+0.069, $p = 0.049$, against +0.064, $p = 0.19$), so lumpiness does inflate the share where the platforms are, but the mechanism holds in both halves. Script 106.

large industry-wide movement in either direction, which bounds the critics' offshoring prediction as well as industry's expansion claim; they cannot separate genuinely additional production too small to register from a reallocation of existing output toward foreign customers. The same arithmetic explains why firm-, county- and country-level versions of this design also fail. The decontrolled products are a narrow slice of every aggregate, and an effect real at the product level washes out once aggregated. Trade is the level at which the variation has power.

14 The cost of the licensing regime

14.1 The size of the gain

The treated products exported about \$32 billion a year before reform and about \$38 billion after (\$32.1 and \$37.7 billion; Table 1). Applying the estimated effect to the post-period base implies on the order of \$6 billion a year in additional U.S. exports in these categories relative to the no-decontrol counterfactual. The figure carries two qualifications. First, it is a difference against a declining counterfactual, not a realized flow: the treated products' own exports rose about eleven log points over five years (Figure 2), roughly \$4 billion a year on the pre-reform base, and the remainder of the estimate is the decline the controls suffered and the treated products avoided. Second, it is the HS6-total figure, and the administrative record caps how much of it can be the decontrolled military good: Commerce-jurisdiction exports of the first-rule items ran about \$4.2 billion a year by 2019 (Online Appendix Table OA.1), and the first rule is roughly half the dollar dose, so the observable 600-series flow across all rules is on the order of \$8–9 billion a year. A \$6 billion gain accruing mostly to the military good would leave an implausibly small no-decontrol counterfactual for items that already shipped in volume under ITAR licenses, so a material share of the HS6-total gain is co-moving civilian trade in the same codes, and the figure is an upper bound on the decontrolled military good. It is also an upper bound on the gain attributable to licensing relief, for the channel reason set out in Section 11.3, which the identified log effect does not inherit. The cleaner statement remains the world-share result: the treated products held their share of the world market while comparable defense products lost roughly 25 percent of theirs by the fifth year, a relative gain of about 15 percent averaged over the post window, a few percentage points on the base share, which nets out global demand.

14.2 The implied wedge and administrative saving

Gross export flows do not measure welfare. The \$6 billion of the previous subsection is a gross trade flow; what decontrol plausibly saved is the real resource cost of the licensing apparatus on the trade that had to cross it, and that can be quantified from the estimates themselves. Inverting the measured effects through a trade elasticity prices the wedge the reform removed—the ITAR regime's stringency over the 600-series Commerce treatment the goods moved to, not the cost of licensing relative to none—from three routes: the cross-sectional license dose, the supplier-differenced triple difference, and the bilateral panel of Section 11.1. As measured, the routes give 2 to 3.5 percent ad valorem. Extrapolated to a fully-licensed basis by the treated set's mean dose of 0.27, a linear-in-dose extrapolation, they give 8 to 13.5 percent, and everything downstream is carried as that range. Section 9 bears on the choice between the two: solved in general equilibrium the measured wedge reproduces the estimated export response, while the extrapolated wedge implies a response two to

four times too large, which is a reason to read the lower range as the operative one.²⁹

Applying that wedge to the \$9.0 billion a year of treated trade that actually moved under individual State licenses before reform, and adding the surplus on the trade the reform induced, gives an annual administrative saving of roughly \$0.3 billion as measured and on the order of \$1.3 billion under the dose extrapolation. The rescaled figure splits into \$1.07 billion of cost no longer incurred on trade that would have happened anyway and \$0.24 billion of surplus on trade that would not have. Most of what decontrol bought was not new commerce but the removal of a cost from existing commerce, which is why the gross export figure and this one answer different questions.

Because the routes into the wedge are not equally exposed to the estimand ambiguity of Section 8, we price that ambiguity as a scenario. Under the drawdown reading of the control path, the supplier-differenced route is the most contaminated of the three—it is precisely the U.S.-specific control decline restated—while the license-dose gradient and the bilateral panel are estimated on the treated products’ own variation and do not rest on the control path. Rebuilding the ledger from those two routes alone, and crediting only the realized \$4 billion component of the gain, gives a wedge of 8.1 to 9.3 percent rescaled (about 2 percent unrescaled) and an annual saving of roughly \$0.8 to \$0.9 billion under the dose rescaling and \$0.2 billion without it, against \$0.3 to \$1.3 billion when the control decline is attributed to continuing control. The measured domestic-demand differential runs against the drawdown reading (Appendix B.4), so we treat the full range as the primary object; the scenario shows the ledger’s floor does not collapse under the less favorable attribution.

Those two components stand on different footing, and Section 11.1 is what separates them. The \$1.07 billion inframarginal term is a real resource saving, administrative effort no longer expended on trade that would have crossed the apparatus in any case, and it holds whoever the trade came from. The induced component does not, and neither does the \$6 billion gross flow. The bilateral panel identifies against whose position the U.S. gain registered, and the answer is allied industrial suppliers operating under a multilateral list that did not move. To that extent the market-position gain is a transfer between allied producers and not a global efficiency gain, though a calculation that stops at the U.S. border counts it as one. Section 9 prices that transfer and finds it second-order in welfare even where it is first-order in market share. The resource saving is the part of this ledger a unilateral reform can claim on efficiency grounds. The share gain is a distributional result, and whether it leaves the alliance better off in aggregate is a question about allied industrial capacity that this design does not answer.

We do not report that as the welfare effect of the reform, because it is not one. It is the saving on one side of a ledger whose other side is unmeasured: the calculation treats the license as pure deadweight, so it values at zero whatever screening the individual review was buying, and the security literature’s whole objection to decontrol is that the screening was worth something (Levine and Smith, 2000; Stricker and Albright, 2017). The defensible use of the number is therefore as a break-even. Set against the treated products’ \$32 billion a year of pre-reform exports, the saving translates into a threshold: the reform is welfare-neutral if decontrol raised expected end-use losses by more than somewhere between one and four cents per dollar of these products’ trade—about one cent taking the measured effects at face value, three to four cents under the dose rescaling. The threshold moves by known amounts within that range: counting only the inframarginal resource

²⁹Appendix B.10 sets out the inversion. The routes are not independent: the supplier-differenced and bilateral estimates are both U.S.-versus-foreign contrasts on the same reconciled data at different levels of aggregation, so their agreement is consistency and not corroboration. We lead with the supplier-differenced figure because it is already a relative quantity. Its elasticity band runs from 6.2 to 25.9 percent, which is dispersion in σ alone and carries no sampling error; the unrescaled 3.5 percent figure is equivalent to a 3.5 percent reduction in iceberg delivery costs on these products in the Armington parameterization. Script 103.

saving, which is what survives the transfer argument above, trims the rescaled figure to about three cents, and carrying the elasticity band instead of its center stretches the upper end to seven; pricing the drawdown scenario above, which strips the routes and components that rest on the control path, puts the floor just under one cent and the rescaled figure near three. In absolute terms the question is whether decontrol raised expected end-use losses by more than roughly \$0.3 billion a year on the unrescaled figure, or \$1.3 billion on the rescaled one. On the wider general-equilibrium ledger of Section 9 the threshold is higher, about nine to sixteen cents per dollar of these products' trade at the measured wedge, so the administrative ledger stated here is the conservative of the two. That is the quantity a decision-maker needs and the one this design cannot supply: whether the screening given up was worth more or less than that is a question about enforcement and end use, not about trade flows. Stating the threshold names the missing term and puts a number on how large it would have to be to overturn the ledger.

The threshold is also a specific decision object, and four assumptions travel with it, each moving it in a known direction. First, it prices losses at their expected value, linear in trade dollars, so a decision-maker with convex losses over tail events—a single diverted subsystem can impose costs far beyond its trade value—should read the figure as a lower bound on the required caution. Second, it compares flows inside the outcome window, while the losses from a transferred capability materialize with lags the satellite episode of Section 3 measures in decades, so the gain side is measured and the loss side right-censored. Third, its two sides fall on different parties—the saving accrues to U.S. exporters and agencies while expected end-use losses fall on allies, third countries, and future U.S. forces—so welfare-neutral is a national accounting statement and not an alliance one. Fourth, end-use exposure concentrates in a thin slice of shipments, so the tolerable loss per dollar actually at risk is a multiple of the per-dollar-of-trade figure stated here.³⁰

14.3 Comparison with the agencies' ex-ante estimates

Neither agency that issued these rules estimated what the relief would be worth, and GAO said so at the time. They produced administrative estimates instead. State's analysis of its own rule put the saving at about 35,000 fewer license submissions a year, scored at an hour of burden each, plus a thousand registrants relieved of registration. BIS valued exporters' saving at \$1.5 to \$5.1 million a year on roughly 2,300 eligible transactions and expected some 20,000 more to become eligible once Munitions List items moved, which carried to that caseload at the agency's own per-transaction rate implies \$15 to \$50 million a year.³¹ The wedge estimated above prices a wider object at \$0.3 to \$1.3 billion a year. Remarkably, even the unrescaled figure is an order of magnitude above the agencies' own valuation; the rescaled one is two.

That gap is not an arithmetic error on either side, and what opens it is the mechanism of Section 12.2. State was answering the Paperwork Reduction Act, which asks what a filing costs to prepare, and BIS was scoring exporters' saving on the transactions one new license exception made eligible. Both priced what they were asked to price. What neither was asked to price is the regime the filing belonged to: the registration a manufacturer owed whether or not it ever shipped, the Part

³⁰The calculation is also a partial-equilibrium accounting on the U.S. exporter side alone, which is why Section 9 solves the same shock in general equilibrium; the trade elasticity is imprecise for the aerospace codes, which is why a band is carried and not a point; and the inversion assumes the whole trade response runs through a cost wedge, which Section 11.3 shows is doubtful for the largest channel. Nothing here rests on a measured producer margin, which the public data do not reveal.

³¹78 FR 22740, at 22752, for the State estimate; GAO-12-613, pp. 25–26, for the BIS cost-benefit analysis and for the finding that the agencies “did not provide an analysis supporting the expected benefits.”

124 agreements with no Commerce analog, the value ceiling each authorization carried, the approval risk on an order not yet placed, and the standing incentive for a foreign manufacturer to design U.S. content out. Exporters respond to the regime, and the relief they took arrived on transaction size, the margin an adjudicated authorization taxes, and not on the filing frequency the agencies counted. Khandelwal et al. (2013) report the same asymmetry in a different apparatus, where the distortion sat in administrative machinery and not in any measured fee. A regulatory impact analysis that scores an export license at an hour of clerical time will understate what removing it does, and by a wide margin.

The estimates also settle a denominator dispute that the current debate is repeating. Measured against U.S. goods exports the gain here is at most about four tenths of one percent, which vindicates the contemporaneous critics who called the prospective gains marginal on that basis. Measured against the affected products' own trade it is about 17 log points, and it reverses a decline their comparators did not escape. Both statements describe the same estimate. The ex-ante argument ran on the national denominator when the policy question turned on the product one, and the case now being made for loosening controls on accelerators and drones is again a national-aggregate case about a product-level instrument. The period's two loudest employment claims, a gain of 340,000 jobs and an offshoring loss of production and technology, both sit outside the range these data support (Section 13).

14.4 The security side of the ledger

What matters for policy is the gain net of any security cost, the proliferation risk formal models of arms-export control weigh against the commercial gain (Levine and Smith, 2000). On the measured margins that cost does not appear: bilateral trade shows no diversion of the decontrolled goods toward re-export hubs, on a test whose power the firearms positive control bounds, and the gain flows to allied and government defense customers. But trade data see reconciliation, not end use. A shipment can clear customs, reconcile across both reporters, and still reach an unauthorized user downstream. The defensible statement is that ECR bought an allied-trade gain at no security cost the trade data can detect, which is weaker than no security cost at all. What the break-even above does is put the missing term on a scale, and that is a magnitude the case-level enforcement record could in principle speak to.

One implication of the paper's own mechanism belongs in the ledger. The substitutability that generates the trade gain also caps the security cost of decontrol. Where a rival supplier working from the same multilateral list can take the sale—the configuration in which the trade estimates apply—a denied U.S. license does not deny the buyer the good, so the marginal screening value of the stricter national regime is small and the break-even threshold is correspondingly easy to meet. Where no substitute exists, the security stakes are largest and the trade estimates do not apply (Section 14.5). To a first approximation, decontrol pays and is safe under the same condition: a substitutable foreign supplier on the same multilateral list.

The ledger is also incomplete on its benefit side, because the reform's declared objectives were themselves security objectives, and the one body that looked directly at whether scrutiny deepened on the items the reform retained did not find that it had. That is a finding about the residual caseload and not a cost of decontrol, and it means the reform's own security rationale should not be assumed to have operated; the fullest outside review of the screening margins reaches a less favorable verdict still (Stricker and Albright, 2017).

14.5 External validity: the 2025–2026 decontrols

The 2025–2026 decisions rest on two premises. That lighter licensing raises exports is affirmed and quantified here: decontrol raised exports of the affected products by about 17 log points, concentrated among allied and defense customers. That export growth then strengthens the industrial base, the premise explicit in the drone orders and the trade-policy counterpart to the subsidy-based industrial policy now common in strategic sectors (Goldberg et al., 2024), is neither confirmed nor refuted. The transfers left no detectable mark on the exposed industries over the five to ten years these data cover, and the arithmetic of Section 13 explains why the rationale working in full would not register at that aggregation. The estimates discipline the rationale’s scale, then, and not its truth, and a decision-maker who wants the industrial-base claim tested needs microdata fine enough to keep the decontrolled lines as their own observation. We do not attempt that here; the conclusion states what such a test would require.

How far any of it carries is bounded by three differences, which together make a rubric for scoring a proposed decontrol before the fact. The ECR moved allied-interoperability military items to allies inside an established monitoring system, so the diversion result is conditional on that destination structure and does not transfer to dual-use sales into adversary or high-risk markets. The ECR was also rule-based, announced years ahead and applied uniformly, where much of the current decontrol is transactional: case-by-case licenses, negotiated revenue shares, suspensions traded inside wider deals. A discretionary regime preserves exactly the per-order adjudication and approval uncertainty whose removal generated the gains measured here, so the trade response to it should be weaker than the estimates above. And decontrol permits trade without compelling it. The January 2026 chip liberalization had produced licenses and almost no deliveries by mid-2026, with the counterparty government withholding demand.³² Table 6 scores the four 2025–2026 actions on those margins, together with the two the estimates identify as the sources of the gain: how much individual licensing the affected trade bore before the change, and whether a substitutable foreign supplier exists on the same multilateral list.

³²Licenses: Bloomberg, “Nvidia Gets US License for Small Amount of H200 Exports to China,” February 26, 2026. Withheld demand and blocked shipments: Chinese customs blocked H200 shipments days after the U.S. authorization and Beijing directed firms not to buy them (Financial Times and Reuters, January 2026).

Table 6. The 2013 configuration and the 2025–2026 decontrols, on the margins the estimates identify

	Pre-change individual licensing	Substitutable foreign supply	Allied, monitored destinations	monitored destinations	Rule-based and pre-announced
600-series transfers, 2013–2017	High: 27% of value individually licensed	Yes: Korea, Japan, Germany, Italy, Netherlands	Predominantly, with retained MDE notification	Yes	
Civil-drone export expedition, 2025–26	Low: an EAR-side caseload of tens of licenses a year	Yes, and the fringe is Chinese	Mixed		Yes
Military UAS review standard, 2025	High	Partly, on MTCR-class systems	Mixed		Partly: a review-policy change, not a list change
AI-stack export financing, 2025	Not a licensing change	Limited	Mixed		No: package by package
H200 accelerators to China, 2026	Denial policy, so licensing was not the margin	No: dominant U.S. supplier	No		No: case-by-case, inside a wider negotiation

The first two columns are the conditions the estimates identify as generating the trade gain: a licensing burden large enough to price, and a foreign fringe able to take the sale if the burden is not removed. The third and fourth are the conditions under which no security cost appeared. The 2013 row is the configuration these estimates describe; the rows below it are scored from the rules and orders cited in Section 3 and above, not estimated. A decontrol that scores low on the first two should not be expected to reproduce the trade estimate; one that scores low on the second two should not be expected to reproduce the security record.

15 Conclusion

This paper asked whether easier export strengthens the industries that make the good, at an acceptable security cost. For the 2013–2017 decontrol the answer is a qualified yes, and the qualification matters: what the estimates support is a particular configuration of decontrol, not loosening in general. The measured gains and the undetected costs are joint products of the 2013 design—allied-heavy destinations, government end-users under Strategic Trade Authorization conditions, retained congressional notification for major defense equipment, and a rule-based, pre-announced rollout—and the one decontrol that departed from that configuration on destination and end-user structure, the 2020 firearms transfer, is also the one where GAO found post-transfer shipments rising fastest to high-risk resellers (GAO-25-106849; Appendix B.14). That is a single episode, so it motivates the configurational reading without establishing it. Decontrols that abandon the 2013 destination and end-user features should expect to inherit neither the trade estimate nor the security record.

The evidence rests on one identifying assumption: that the pre-announced rollout order is unrelated to a demand shift graded by product category and specific to U.S. suppliers. The measured allied-demand controls, the supplier-differenced bound, and the dose-response evidence are consistent with that assumption. The leading candidate for such a shift, the allied fighter and rotary-wing delivery wave of 2014–2018 concentrated in aircraft and gas-turbine engines, is

bounded by dropping those two families outright, which leaves the export effect at +0.137 and raises the world-share effect; the control side's counterpart, a U.S.-specific procurement drawdown, runs against the measured domestic-demand differential and is the reason the estimates are stated as the policy wedge of Section 8 (Appendix B.4).

Under that assumption, the 600-series transfers raised affected products' exports by about 17 log points relative to not-yet-treated defense products, and by 15 to 18 across heterogeneity-robust estimators. They held their world-export share while comparable defense products' shares fell roughly 25 percent by the fifth year, a relative gain of about 15 percent averaged over the post window, concentrated in allied and defense-customer destinations and registered against the market position of allied industrial suppliers. The exposed industries' output, employment, and patenting show no detectable response, on bounds set by data resolution, and no diversion appears in bilateral trade.

Set against what the reform was promised to deliver, those estimates contradict both sides of the original argument, in opposite directions (Section 14.3). What the record shows is a product-level instrument debated at the national aggregate, and a licensing regime costed as though its cost were its paperwork.

Three of those results carry beyond the episode. Because these goods stayed on the Wassenaar Munitions List throughout, the relative position recovered against Korea, Japan, Germany, Italy, and the Netherlands prices what a supplier gives up by licensing a shared list more strictly than its competitors' national regimes do. Because the gain scales with the licensing burden each product bore and arrives as larger authorized transactions and not more frequent ones, the friction removed can also be priced: an ad-valorem equivalent of 2 to 3.5 percent for the ITAR wedge, 8 to 13.5 percent on a fully-licensed basis, falling on the per-order authorization. And the administrative saving those numbers imply turns the security question from an impasse into a threshold: the reform paid for itself unless decontrol raised expected end-use losses above one to four cents per dollar of the affected trade. Whether it did is a question about enforcement records and not about trade flows, and it is the one this design hands on.

What these data leave open would be sharpened by two kinds of record. The domestic nulls are bounds set by data resolution, because the decontrolled goods are too small a share of any public aggregate for even the industrial-base rationale working in full to register. Isolating them requires microdata fine enough to keep them as their own observation, and the Census Longitudinal Firm Trade Transactions Database would let the question be asked within the firm: whether a firm's decontrolled lines grew relative to its others, whether any expansion reached hiring or investment, and how they moved against the firm's own retained-USML lines, which public product data cannot compare within category. Closing the security side needs the other kind of record. The break-even above is a threshold on expected end-use losses, and the case-level enforcement and end-use-check record is where a number for it would come from. That record is not published in a form that can be attributed to the transferred goods, a gap that is itself informative about how export-control rules are evaluated.

References

- Aghion, P., A. Bergeaud, M. Lequien, and M. J. Melitz. 2024. “The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports.” *Review of Economics and Statistics* 106(3): 608–626.
- Autor, D., D. Dorn, G. H. Hanson, G. Pisano, and P. Shu. 2020. “Foreign Competition and Domestic Innovation: Evidence from U.S. Patents.” *American Economic Review: Insights* 2(3): 357–374.
- Bloom, N., M. Draca, and J. Van Reenen. 2016. “Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity.” *Review of Economic Studies* 83(1): 87–117.
- Borusyak, K., X. Jaravel, and J. Spiess. 2024. “Revisiting Event-Study Designs: Robust and Efficient Estimation.” *Review of Economic Studies* 91(6): 3253–3285.
- Broda, C., and D. E. Weinstein. 2006. “Globalization and the Gains from Variety.” *Quarterly Journal of Economics* 121(2): 541–585.
- Callaway, B., and P. Sant’Anna. 2021. “Difference-in-Differences with Multiple Time Periods.” *Journal of Econometrics* 225(2): 200–230.
- Cengiz, D., A. Dube, A. Lindner, and B. Zipperer. 2019. “The Effect of Minimum Wages on Low-Wage Jobs.” *Quarterly Journal of Economics* 134(3): 1405–1454.
- Clayton, C., M. Maggiori, and J. Schreger. 2023. “A Framework for Geoeconomics.” NBER Working Paper No. 31852.
- Costinot, A., and A. Rodríguez-Clare. 2014. “Trade Theory with Numbers: Quantifying the Consequences of Globalization.” In *Handbook of International Economics*, Vol. 4, ed. G. Gopinath, E. Helpman, and K. Rogoff, 197–261. Amsterdam: Elsevier.
- Crosignani, M., L. Han, M. Macchiavelli, and A. F. Silva. 2026. “Securing Technological Leadership? The Cost of Export Controls on Firms.” *Journal of Financial Economics* 175: 104192.
- Crozet, M., and J. Hinz. 2020. “Friendly Fire: The Trade Impact of the Russia Sanctions and Counter-Sanctions.” *Economic Policy* 35(101): 97–146.
- de Chaisemartin, C., and X. D’Haultfoeuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review* 110(9): 2964–2996.
- DellaVigna, S., and E. La Ferrara. 2010. “Detecting Illegal Arms Trade.” *American Economic Journal: Economic Policy* 2(4): 26–57.
- DeVol, R., et al. 2010. *Jobs for America: Investments and Policies for Economic Growth and Competitiveness*. Santa Monica: Milken Institute, with the National Association of Manufacturers, January.
- Djankov, S., C. Freund, and C. S. Pham. 2010. “Trading on Time.” *Review of Economics and Statistics* 92(1): 166–173.
- Fajgelbaum, P. D., P. K. Goldberg, P. J. Kennedy, and A. K. Khandelwal. 2020. “The Return to Protectionism.” *Quarterly Journal of Economics* 135(1): 1–55.
- Fan, J., K. Handley, and N. Limão. 2026. “Multilateral Export Control Regimes and Aggregate Exports.” *AEA Papers and Proceedings* 116: 109–113.
- Farrell, H., and A. L. Newman. 2019. “Weaponized Interdependence: How Global Economic Networks Shape State Coercion.” *International Security* 44(1): 42–79.

- Feenstra, R. C. 1994. “New Product Varieties and the Measurement of International Prices.” *American Economic Review* 84(1): 157–177.
- Felbermayr, G., A. Kirilakha, C. Syropoulos, E. Yalcin, and Y. V. Yotov. 2020. “The Global Sanctions Data Base.” *European Economic Review* 129: 103561.
- Fergusson, I. F., and P. K. Kerr. 2020. “The U.S. Export Control System and the Export Control Reform Initiative.” Congressional Research Service Report R41916, January 28.
- Flaaen, A., and J. Pierce. 2024. “Disentangling the Effects of the 2018–2019 Tariffs on a Globally Connected U.S. Manufacturing Sector.” *Review of Economics and Statistics*, published online 2024 (DOI: 10.1162/rest.a_01498).
- García-Alonso, M. D. C., and P. Levine. 2007. “Arms Trade and Arms Races: A Strategic Analysis.” In *Handbook of Defense Economics*, Vol. 2, ed. T. Sandler and K. Hartley, 941–971. Amsterdam: Elsevier.
- Goldberg, P. K., R. Juhász, N. J. Lane, G. Lo Forte, and J. Thurk. 2024. “Industrial Policy in the Global Semiconductor Sector.” NBER Working Paper No. 32651.
- Goodman-Bacon, A. 2021. “Difference-in-Differences with Variation in Treatment Timing.” *Journal of Econometrics* 225(2): 254–277.
- Haidar, J. I. 2017. “Sanctions and Export Deflection: Evidence from Iran.” *Economic Policy* 32(90): 319–355.
- Handley, K., and N. Limão. 2015. “Trade and Investment under Policy Uncertainty: Theory and Firm Evidence.” *American Economic Journal: Economic Policy* 7(4): 189–222.
- Handley, K., and N. Limão. 2017. “Policy Uncertainty, Trade, and Welfare: Theory and Evidence for China and the United States.” *American Economic Review* 107(9): 2731–2783.
- Hartung, W. D. 2013. *Risks and Returns: The Economic Illogic of the Obama Administration’s Arms Export Reform*. Washington, DC: Center for International Policy, August.
- Hornok, C., and M. Koren. 2015a. “Administrative Barriers to Trade.” *Journal of International Economics* 96: S110–S122.
- Hornok, C., and M. Koren. 2015b. “Per-Shipment Costs and the Lumpiness of International Trade.” *Review of Economics and Statistics* 97(2): 525–530.
- Hummels, D. L., and G. Schaur. 2013. “Time as a Trade Barrier.” *American Economic Review* 103(7): 2935–2959.
- 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.
- Kee, H. L., A. Nicita, and M. Olarreaga. 2009. “Estimating Trade Restrictiveness Indices.” *Economic Journal* 119(534): 172–199.
- Kropf, A., and P. Sauré. 2014. “Fixed Costs per Shipment.” *Journal of International Economics* 92(1): 166–184.
- Levine, P., and R. Smith. 2000. “Arms Export Controls and Proliferation.” *Journal of Conflict Resolution* 44(6): 885–895.
- Lileeva, A., and D. Treffer. 2010. “Improved Access to Foreign Markets Raises Plant-Level Productivity... For Some Plants.” *Quarterly Journal of Economics* 125(3): 1051–1099.
- Liu, X., Y. Liu, A. Makarin, and J. Y. Wen. 2024. “Export Controls and Innovation in Sanctioned Countries.” Harvard Business School Working Paper 25-004, August 2024, revised September 2025.

- Makioka, R., and H. Zhang. 2024. “The Impact of Export Controls on International Trade: Evidence from the Japan–Korea Trade Dispute in the Semiconductor Industry.” *Journal of the Japanese and International Economies* 74: 101336.
- Moretti, E., C. Steinwender, and J. Van Reenen. 2025. “The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers.” *Review of Economics and Statistics* 107(1): 14–27.
- Pierce, J. R., and P. K. Schott. 2012. “A Concordance between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries.” *Journal of Official Statistics* 28(1): 53–68.
- Pierce, J. R., and P. K. Schott. 2016. “The Surprisingly Swift Decline of US Manufacturing Employment.” *American Economic Review* 106(7): 1632–1662.
- Rambachan, A., and J. Roth. 2023. “A More Credible Approach to Parallel Trends.” *Review of Economic Studies* 90(5): 2555–2591.
- Roth, J., P. H. C. Sant’Anna, A. Bilinski, and J. Poe. 2023. “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature.” *Journal of Econometrics* 235(2): 2218–2244.
- Soderbery, A. 2015. “Estimating Import Supply and Demand Elasticities: Analysis and Implications.” *Journal of International Economics* 96(1): 1–17.
- Stricker, A., and D. Albright. 2017. *U.S. Export Control Reform: Impacts and Implications*. Washington, DC: Institute for Science and International Security, May 17.
- Sun, L., and S. Abraham. 2021. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics* 225(2): 175–199.
- Volpe Martincus, C., J. Carballo, and A. Graziano. 2015. “Customs.” *Journal of International Economics* 96(1): 119–137.

A Data

Table 7. Principal datasets

Dataset	Source	Level and coverage	Role	Main limitation
Census ECCN export panel	U.S. Census AES (restricted-access filing records)	HS6 $\times$ destination $\times$ license type, quarterly, 2009Q1–2026Q1 (estimation uses complete years 2009–2025)	Treatment (600-series flag; pre-reform license-code gate), primary export-volume outcome, license dose, destination test	Restricted-access; the ECCN field over-populates, so treatment is gated on the license-code field; the HS6 total mixes co-moving civilian trade
BACI	CEPII, rec- onciling UN Comtrade	HS6 bilateral, annual; HS92 2009–2011, HS12 2012–2024; defense chapters 84,85,87,88,89,90,93	U.S. share of world exports (differences out global demand); product-level world-excl.-U.S. demand control	HS6 only, so it cannot separate the decontrolled military variant from civilian trade; HS92/HS12 vintage seam at 2011/12
County Business Patterns	U.S. Census	NAICS6, national, annual 2008–2023	Industry employment, establishments, payroll (second stage)	HS6→NAICS6 mapping dilutes the shock; ≈ 22 treated industries limit precision
Economic Census	U.S. Census	NAICS6, quinquennial 2012/2017/2022	Value of shipments, value added, capital expenditure (second stage)	One pre-period (2012), the defense-budget trough, confounds the capital-expenditure result; wide output bounds
PatentsView	USPTO PatentsView	/ Granted patents mapped to NAICS6 by primary technology class, grant year	Innovation second stage	CPC→NAICS6 cross-walk reallocates mass; no military technology class for aircraft or defense electronics
SIPRI Arms Transfers Database	SIPRI	Weapon-category $\times$ year delivery volumes (TIV), allied and world	Measured category-level demand control	Major-weapons scope only (misses small arms and components); a volume index, not dollars
BIS/OTE USML-to-CCL statistics	U.S. Commerce, BIS Office of Technology Evaluation, from AES	600-series export and licensing value, by transfer rule	Administrative cross-check on scale, destinations, and STA composition	No pre-period; a level, not a difference; drawn from the same AES as the Census panel, so consistent but not independent
Section 655 reports	U.S. State, Directorate of Defense Trade Controls	Munitions-List DCS authorization value, by USML category (and country), annual FY2010–2019	State-side pre-period cross-check on the State→Commerce migration of the transferred categories	Authorizations, not shipments; excludes Foreign Military Sales; USML-category level, not HS6; FY2010/2012 OCR-recovered from scanned images

Datasets that carry the identification and the main results. Crosswalks (regulation-text USML→HS, Pierce-Schott and FGKK HS→NAICS6) and confound checks are described where used.

B Additional robustness

B.1 Sensitivity to the aircraft family

Because the dollar dose is aircraft-heavy (Herfindahl 3,047), we ask whether the estimate is an aircraft artifact. It is not. Estimated family by family, the ATT is positive for every family on both outcomes (9x610 +0.170/ + 0.087, 9x619 +0.400/ + 0.082, 3x611 +0.115/ + 0.241, the remaining families +0.104/ + 0.123; small-count families are noisier but positive). In a leave-one-out exercise (Table 8), dropping the entire 9x610 aircraft family—half the dose—leaves the volume ATT essentially unchanged (+0.170 $\rightarrow$ +0.176) and *raises* the share ATT (+0.147 $\rightarrow$ +0.174). Dropping the whole HS88 aircraft chapter moves neither outcome materially (+0.146/ + 0.148).

Those cells are taken one at a time, which is not quite the right test for the confounder that motivates them. The allied fighter and rotary-wing delivery wave of 2014–2018 ran through airframes and engines together, and dropping the HS88 chapter does not remove the engines, whose turbojets sit in chapter 84. Removing both program families at once leaves 48 treated products and gives +0.137 (SE 0.047) on exports and +0.193 (SE 0.057) on world share, with pre-period deviations of 0.016 and 0.020, cleaner than the full sample’s. The chapter-based version of the same cut, dropping HS88 and the HS8411 engine heading, gives +0.120 (SE 0.041) and +0.149 (SE 0.052). The export estimate attenuates by about a fifth and the share estimate rises; neither collapses. On the 48 products with no exposure to the fighter and engine programs the effect is still there.³³ The effect is also positive across cohorts (volume +0.257, +0.160, +0.074 and share +0.095, +0.118, +0.286 for 2013–2015). On the realized set, by contrast, dropping aircraft attenuated the volume ATT from +0.18 to +0.12; the regime gate removes that fragility. The Herfindahl bears on dose-weighting and inference, not on identification.

B.2 Sensitivity to the regime gate

The 5 percent ITAR-individual gate is one cut. Table 9 sweeps it from 2 percent to 50 percent. The volume ATT is stable-to-rising as the gate tightens (+0.156 at 2% to +0.256 at 50%), and the share ATT holds in the +0.12–+0.19 range. Tightening the gate keeps the most clearly treated products, and the effect does not weaken, the opposite of what a survivorship or dilution artifact would produce. Precision and pre-period quality do fall, however, and the export pre-trend is not flat across the sweep: it stays at or below 0.028 through the 10 percent gate and then deteriorates as the treated set thins, to 0.061 at 25 percent and 0.136 at 50 percent, where 17 products remain. The tightest cuts are the most clearly treated and the least well identified at once, so we read them as a direction of movement rather than as point estimates.

The gate sweep also bounds the civilian-contamination concern of Section 10.4. The pre-reform ITAR individual-license share sets the minimum military content of the HS6 baseline: at the 50 percent gate, more than half of each treated product’s pre-reform exports moved under individually adjudicated State licenses, so the HS6 baseline is majority military by construction. At that gate the export gain rises to +0.256 ($t = 2.5$) and the world-share gain to +0.189 ($t = 2.8$). The two outcomes’ pre-periods diverge sharply there: the share pre-period coefficient is 0.032, the cleanest in the table, while the export pre-period coefficient is 0.136, the worst, so the bound below is stated on the export estimate whose own pre-period is the weakest of the sweep. A civilian co-movement large enough to fully account for the +0.256 estimate at this threshold would require civilian exports of the treated

³³Script 106. Estimated on the same 88-product regime-gated set as Table 8, whose full-sample values are +0.170 and +0.147.

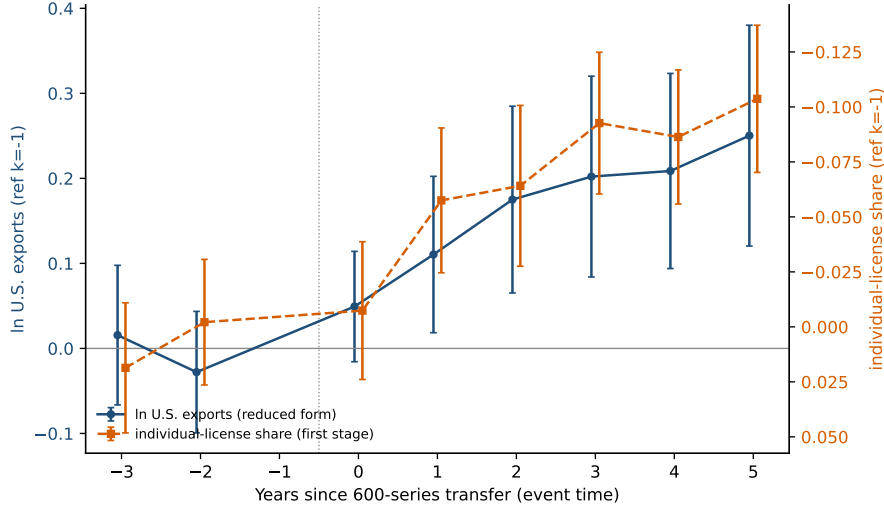


Figure 5. The licensing channel. Individual-license share (right axis, inverted) and ln U.S. exports (left axis), regime-gated stacked event studies. Both are flat before the effective date and build after it: trade rises as the individual-license burden falls. 95% confidence intervals, clustered on HS6.

Table 8. Leave-one-out pooled ATT, regime-gated treatment

Dropped	<i>n</i> HS6	ln U.S. exports	U.S. world share
(none) full	88	+0.170 (.045)	+0.147 (.048)
9x610 aircraft family (48% of dose)	56	+0.176 (.058)	+0.174 (.053)
9x619 engines	80	+0.147 (.041)	+0.154 (.049)
3x611 electronics	69	+0.185 (.053)	+0.116 (.052)
0x606 vehicles	86	+0.160 (.045)	+0.138 (.048)
HS88 aircraft chapter	78	+0.147 (.048)	+0.148 (.050)

Pooled ATT (post) re-estimated on the regime-gated treated set with the indicated family or chapter removed. SEs clustered on HS6 in parentheses.

products to grow more than 40 log points above those of the controls, a gap the supplier-differenced bound of Section 10.3 rules out for any demand shift common to U.S. and non-U.S. suppliers of the same HS6. The residual civilian-contamination concern is therefore a differential shock specific to U.S. civilian exports of the early-transferred product categories, narrower than the general confound and inconsistent with the allied-destination concentration of Section 11.1.

Its pre-trend on the share outcome degrades at the 25 percent gate (0.120) and then recovers at 50 percent (0.032). The intermediate degradation reflects small-sample noise as treated products with 25 to 50 percent ITAR share are added: most of the 17 products at the 50 percent cut are larger, longer-traded platforms with stable pre-periods. The default 5 percent gate sits in the range where both outcomes and pre-trends are stable, and the gate-robustness results confirm the civilian-contamination bound does not weaken as the baseline becomes more clearly military.

Table 9. Regime-gate threshold sensitivity

ITAR _{ind} gate	n treated	ln exports ATT	max pre	world share ATT	max pre
2%	160	+0.156 (.034)	.027	+0.119 (.041)	.037
5%	88	+0.170 (.045)	.028	+0.147 (.048)	.049
10%	62	+0.216 (.055)	.023	+0.145 (.049)	.074
25%	38	+0.191 (.063)	.061	+0.116 (.055)	.120
50%	17	+0.256 (.103)	.136	+0.189 (.068)	.032

Pooled ATT (post), standard errors clustered on HS6 in parentheses, and largest absolute pre-period coefficient at each gate. Treated set = ever-600 $\cap$ pre-2013 ITAR-individual share $\geq$ gate, cohorts 2013–2017. At the 50% gate more than half of each treated product’s pre-reform exports moved under individually adjudicated State Department licenses; the HS6 baseline is majority military. Expanding the treated set to 101 codes by excluding Canada from the ITAR-individual share denominator (the §126.5 exemption covers most categories but not Category VIII(a) aircraft, which required individual licenses and already appear in the numerator) changes the 5% gate export ATT by less than 2 percent (+0.168) and the share ATT by about 6 percent (+0.138). Scripts 73 (gate sweep) and 83 (Canada exclusion).

B.3 Control-set robustness and share scaling

The baseline control pool is the not-yet-treated defense universe, mostly later- and never-transferred defense products. Two variants confirm the ATT does not depend on it. Dropping every transferred-but-sub-gate product from the universe leaves the estimate numerically unchanged, because such a product’s own 2013–2017 cohort falls inside every sub-experiment’s window and it never enters a control group. Restricting the controls to the never-600 set alone, defense products with pre-reform ITAR individual licensing that never transferred, leaves the export ATT if anything larger (+0.185), though the pool is thin (29 HS6), so the standard error and the pre-trend both widen and the world-share estimate loses precision on nine control products. The effect is not an artifact of using later-transferred products as not-yet-treated controls. A treated-set restriction provides complementary evidence: excluding the 2013 cohort (25 HS6 across electronics, engines, aircraft, and optics — the largest cohort and the one most exposed to sequestration-recovery trends) leaves the export ATT at +0.139 ($t = 2.7$) and the world-share ATT at +0.170 ($t = 3.0$), both significant at five percent. The world-share estimate rises modestly, consistent with the 2013 cohort contributing large export volume but below-average share gains. On the share scaling, the outcome $\text{asinh}(\text{share} \times 100)$ approximates a log because every treated product’s pre-reform world share exceeds one percent (median 17 percent); re-estimating with the log of the share on the positive-share subsample returns +0.169, against the +0.147 baseline, so the proportional reading is not an artifact of the transform.

B.4 Counterfactual validity: anticipation, spillovers, and the control path

Three threats to the not-yet-treated counterfactual are tested here: anticipation by the controls, substitution from control to treated products, and a control trajectory driven by something other than the treated products’ counterfactual.

Firearms coding and control-side anticipation. The 2020 rule moved Categories I–III to a 500-series classification, so the data-driven 600-series cohort rule codes firearms products without stray 600-series codes as never-treated: six firearms HS6 sit in the control pools through the panel windows, alongside the other never-600 defense controls. Because every control category

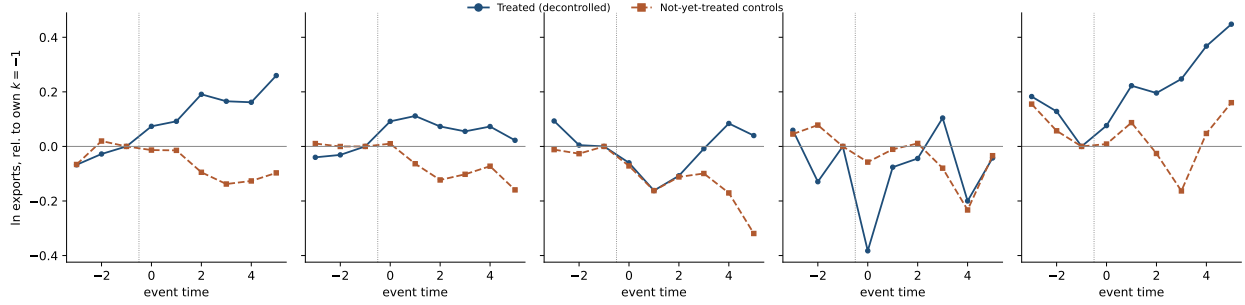


Figure 6. Staggered divergence, by transfer cohort (discussed in Section 10). Each panel plots $\ln$ exports relative to each product’s own value the year before its cohort’s transfer ($k = -1$), averaged across that cohort’s treated products and across the defense controls not yet treated within its window. The treated series turns up at each cohort’s own event zero. The two large early cohorts, aircraft and engines in 2013 and ground vehicles and vessels in 2014, show the clearest separation; the later cohorts hold a handful of products each and are noisier.

Table 10. Censoring and firearms-coding variants of the pooled ATT

Variant	$\ln$ U.S. exports	U.S. world share
Baseline	+0.170 (.045)	+0.147 (.048)
Censor firearms controls from 2018 (proposed rule)	+0.170 (.045)	+0.150 (.048)
Censor firearms controls from 2020 (effective)	+0.170 (.045)	+0.147 (.048)
Censor every control at its own proposed date	+0.207 (.064)	+0.132 (.072)
Drop firearms controls entirely	+0.168 (.045)	+0.154 (.048)
Firearms-clean (censor 2018 + drop 11 contaminated treated)	+0.161 (.049)	+0.163 (.050)

Regime-gated pooled ATT under firearms-coding and announcement-censoring corrections; SEs clustered on HS6 in parentheses. Full own-date censoring removes 69 percent of stacked control rows (most stray-late-600 controls have 2010–2012 proposed dates), so its estimates are bounds on a much thinner pool, and the share variant is significant at ten percent ($p = 0.067$).

eventually received its own proposed rule, announcement effects on controls would contaminate the post-windows of the early cohorts. Event studies centered on each control’s *own* announcement show no downward break: firearms exports move insignificantly after their 2018 proposed rule (-0.15 , $p = 0.44$; label-permutation $p = 0.63$) and after the 2020 effective date (-0.16 , $p = 0.51$), and the pooled announcement-clock event study across all announced controls is flat, the one marginal coefficient being a *positive* share shift ($+0.10$, $p = 0.07$) that would bias the estimate toward zero. Censoring corrections confirm the headline does not depend on these observations (Table 10): censoring firearms controls at 2018 or 2020, or dropping them, moves either ATT by no more than 0.016; censoring *every* control at its own proposed date, which deletes two-thirds of the stacked control rows, raises the export ATT to $+0.207$ and leaves the share estimate near baseline at reduced precision.

An out-of-sample positive control: the 2020 firearms transfer. The 2020 transfer of Categories I–III to a 500-series classification is a decontrol outside the estimation sample, on which the design can be re-run from scratch. Coding the nine firearms HS6 that clear the same pre-reform licensing gate as treated (cohort 2020) and comparing them to still-USML defense products, the stacked estimator returns an export ATT of $+0.280$ (SE 0.110, $t = 2.5$; wild-cluster $p = 0.02$,

randomization $p = 0.06$), building only after the transfer and stable across COVID windows (+0.34 to +0.35 dropping 2020 or 2020–21). The exercise is consistent with the design recovering a decontrol gain of the headline magnitude on a cohort it never saw, but it is not an independent validation: its lead coefficient of +0.30 two years out is as large as the effect, the criterion on which estimates are discarded elsewhere in the paper. The world-share outcome is null (+0.01, $p = 0.97$): firearms are fungible, competitively supplied goods on which the United States neither holds nor gains share, so the avoided-decline share result does not generalize to them. The exercise also supplies the diversion power check of Section 13; both are in script 67.

The control decline is U.S.-specific. A within-product comparison of U.S. and non-U.S. exports of the same control HS6 in BACI shows the control decline of Figure 2 is specific to U.S. suppliers: post-transfer, U.S. exports of the control products fall 17 log points relative to non-U.S. exports of the same products (post $\times$ U.S. source -0.167 , SE 0.034), the same contrast for treated products is zero (+0.02, $p = 0.63$), and the triple difference is +0.188 (SE 0.055), a figure numerically invariant to absorbing demand at the chapter-by-year and product-by-year levels (script 71), which Section 10.3 reads as the supplier-differenced demand bound. Non-U.S. suppliers' exports of the control products did not fall at all. A worldwide demand shock to these categories therefore cannot explain the control path. Three readings remain. One is substitution: decontrol diverted U.S. sales effort and buyers' budgets from still-ITAR lines toward the decontrolled lines, making part of the control decline treatment-caused and the difference-in-differences an overstatement of the net gain. The second is the framework's own mechanism operating on the controls: still-controlled U.S. lines kept paying the licensing cost and lost ground to unconstrained foreign suppliers, the ITAR-free displacement the satellite episode exemplifies, in which case the control path is exactly the policy-relevant counterfactual (what continued control costs), and no assumption is violated. The third is a U.S.-specific cycle with no licensing content. The post-2012 procurement drawdown and sequestration fell on the retained and later-transferred categories—munitions, ordnance, legacy platforms—largely through pipelines served by U.S. incumbents, and export streams tied to U.S. force posture contracted with them. A decline of that origin is U.S.-supplier-specific by incumbency, so the supplier-differenced contrast above cannot separate it from the cost of continuing control; it is invisible to SIPRI, which records complete platforms rather than the components and consumables where a drawdown bites; and the same-HS4 test below does not bear on it either. The drawdown itself can be measured. BEA's national accounts split federal defense purchases of equipment by product type (NIPA Table 3.11.5: aircraft, missiles, ships, vehicles, electronics, ammunition, other), a purely domestic series that contains no export flow and has a full pre-period. Mapping the seven types to transferred and retained Munitions List categories, the confound requires domestic demand to have fallen more in the control-mapped types. It fell more in the treated-mapped ones: mean log change 2009–2012 to 2014–2018 of -0.34 treated-mapped against -0.14 control-mapped (-0.26 and -0.13 weighting by pre-period value), a differential of the wrong sign for the confound by 13 to 20 log points (script 115). Two qualifications travel with it: aircraft, half the dollar dose, is flat instead of contracting, so the unweighted differential is carried by vehicles and electronics; and BEA's seven types are coarser than twenty-one Munitions List categories, with the residual "other" type certainly containing some transferred content—a misallocation that pulls the control-mapped average toward the treated categories' behavior and so works against the finding. The measurement leaves a sharper cousin of the objection live: a collapsing domestic order book in the treated categories could push their producers toward export markets, a pivot-to-export response with no licensing content, which works with rather than against the measured sign and which the license-dose gradient of Section 12.1 addresses only within chapter. The margin where substitution has its sharpest prediction rejects

it: crowd-out should be strongest for the closest substitutes, but the 140 control HS6 that share an HS4 heading with a treated product declined *less* than distant controls, and excluding them *raises* the ATT (exports $+0.170 \rightarrow +0.179$; share $+0.147 \rightarrow +0.156$). A totals test, whether the treated-plus-control pie grew, is structurally uninformative: the treated gain of roughly \$4 billion a year is about one percent of the ~\$430 billion control pool, far below detectability (permutation $p = 0.50$). Firm-level reallocation across distant product lines remains untestable in public product data and belongs to the transaction-level data agenda the conclusion sets out. Between the two readings that survive—continuing control costing the still-controlled lines ground, and a U.S.-specific drawdown doing so—the trade data do not adjudicate: what separates them is whose decision drove the control decline, and no test here observes that. The domestic-demand measurement above bounds the drawdown reading without closing it. This is why Section 8 states the estimand as the decontrol-versus-continued-control policy wedge. The design identifies the treated products’ gain relative to the path comparable U.S. defense products actually followed; the attribution of that path to continuing control is an interpretation the supplier-specificity and same-HS4 evidence support but do not prove. Under the drawdown reading the avoided-decline component of the dollar magnitude weakens accordingly, while the realized-growth component, roughly \$4 billion a year, and the export effect net of the program families ($+0.137$) do not depend on it.

Pre-trend balance. The raw 2009–2012 balance comparison behind Section 10 deserves its numbers in print. The treated-versus-never-treated slope differential is $+0.075$ per year, Welch 95 percent CI $[-0.102, +0.252]$; extrapolated over a five-year post-window it implies a cumulative bias of $+0.376$ $[-0.507, +1.259]$, an interval that straddles zero but whose point estimate exceeds the ATT, so the failure to reject is not evidence of parallel paths. The stacked event-study leads against the not-yet-treated pool are an order of magnitude smaller (exports $+0.016 / -0.028$; share $+0.049 / +0.007$), reflecting the different control pool and conditioning. With two clean leads and a monotone post-period build, pre-trend flatness alone cannot carry the design; that is what the honest-DiD bounds of Section 10.2 quantify, and it is why identification rests on the ex-ante institutional timing argument. A treated-specific trend adjustment would not inform this design—against a flat pre-period and a monotone post-period build, such a trend is nearly collinear with the treatment path itself—whereas for the patents outcome of Section 13, whose pooled coefficient rides a rise already under way before decontrol, industry-specific trends legitimately remove it. The diagnostic is the pre-period shape.

B.5 Measured category-level demand controls in detail

Table 11 reports the full ladder of SIPRI-based demand-control specifications discussed in Section 10.3; the preferred SIPRI control and the product-level BACI control also appear among the headline challenges of Table 3.

Two features of the SIPRI control qualify how far it reaches. The first is coverage. The regulation-text crosswalk maps 93.8 percent of the dollar dose to a SIPRI category, but not every estimation unit: the demand-controlled rows of Table 11 are estimated on the full sample with the category covariate imputed for unmapped units, and on the SIPRI-*mapped* subsample alone the same-sample export baseline is -0.098 (SE 0.234, $t = -0.42$, $N = 964$, 64 treated codes; script 66, row `S0_map`)—an uninformative estimate on a sample that drops the small-arms controls and much of the identifying contrast. The mapped-sample comparison therefore cannot re-establish the headline; what the SIPRI rows show is that adding the measured demand path to the design that does identify the headline moves it little. The second is object. SIPRI records deliveries of complete

Table 11. Measured category-level demand controls (SIPRI arms-transfer deliveries)

Specification	ln U.S. exports	U.S. world share
Baseline	+0.170 (.045)	+0.147 (.048)
+ world deliveries of own category	+0.129 (.051)	+0.119 (.047)
+ allied (A:5) deliveries, non-U.S. suppliers	+0.141 (.048)	+0.131 (.048)
+ world and allied non-U.S. deliveries (preferred)	+0.142 (.048)	+0.134 (.048)
+ world and allied deliveries on order dates	+0.100 (.060)	+0.119 (.050)
+ allied deliveries, all suppliers (over-control)	+0.099 (.057)	+0.117 (.050)
Demand-quartile $\times$ year fixed effects	+0.045 (.107)	+0.163 (.087)
+ product-level world-excl.-U.S. exports (BACI)	+0.229 (.055)	+0.149 (.043)

Regime-gated pooled ATT with category-by-year demand covariates built from SIPRI arms-transfer delivery volumes, mapped to 600-series families through the regulation-text crosswalk (93.8 percent of the dollar dose). Same estimator, windows, fixed effects, and HS6-clustered inference as Table 2; SEs in parentheses. The preferred specification keeps wild-cluster $p \leq 0.005$ (exports) and ≤ 0.009 (share). The all-supplier variant includes U.S. deliveries to allies, part of the outcome, and over-controls by construction. The BACI product-level variant covers every unit including the small-arms controls; its covered-sample baseline is +0.231/+0.147. Script 66.

major weapons—91.2 percent of the transferred-category register value is whole aircraft, vehicles, and ships—while the reform moved components, parts, and less sensitive items and retained the platforms. A category-graded demand shock operating through components and consumables is therefore imperfectly visible to this control, which is why the product-level BACI control, available for every unit and defined at the HS6 itself, is the demand rebuttal the paper leans on.

B.6 Treated-set contamination and the same-tier control

A separate concern is the treated set itself. A hand-matched crosswalk of the transferred and retained Munitions List categories to HS codes, built from the regulation rather than the export-control field, shows that eleven Chapter-93 arms and ammunition codes enter the regime-gated treated set through ECCN-field over-population: Categories I–III firearms and ammunition, which moved to a separate “500-series” Commerce classification in 2020 and not to the 2013–2017 600-series, and Category IV munitions of war, which stayed on the Munitions List.³⁴ Dropping all eleven—eleven percent of the treated set’s pre-reform trade, dominated by the \$9-billion munitions-of-war line—leaves the export ATT at +0.161 and the world-share ATT at +0.160, both significant at one percent. The headline is not an artifact of miscoded arms; if anything the world-share effect strengthens.

The same crosswalk shows why a clean same-strategic-tier control is hard to build from public data, and corrects a tempting shortcut. The never-transferred (“never-600”) defense products are not a stayed-on-USML set: of the twenty-nine, only a handful are genuinely retained defense goods (Category IV munitions and Category XVI nuclear and energetic materials), while six are the 2020-transferred firearms and eleven are non-defense codes that entered through mis-classified shipments. Restricting the controls to the hand-verified stayed-on-USML items leaves seven, in different product spaces from the treated aircraft, engines, and electronics, and the resulting estimate is uninformative (wild-cluster $p > 0.75$). The sensitive items retained *within* the transferred categories (fighters within

³⁴The crosswalk, `data/external/usml_stayed_hs_crosswalk.csv`, classifies each candidate HS6 by its USML category and transfer status from the rule text. The eleven codes carry a stray 600-series ECCN on a minority of shipments, which assigns them spurious 2013–2015 first-appearance cohorts.

Table 12. Pooled ATT by estimator (same regime-gated treated set and defense universe)

Estimator	ln U.S. exports	U.S. world share
Stacked (Cengiz et al., 2019) (baseline)	+0.170 (.045)	+0.147 (.048)
Callaway–Sant’Anna (Callaway and Sant’Anna, 2021)	+0.184 (.048)	+0.163 (.037)
Sun–Abraham (Sun and Abraham, 2021)	+0.150 (.042)	+0.135 (.036)

Pooled ATT, SEs in parentheses. Callaway–Sant’Anna is doubly-robust with not-yet-treated controls and an entity-clustered multiplier bootstrap (999 draws); Sun–Abraham is the interaction-weighted estimator with unit and year fixed effects, clustered on HS6 (delta-method SEs on the cohort-share aggregation). The alternative estimators are run on the 2009–2022 window of the defense-universe panel (script 25); the stacked baseline uses complete calendar years 2009–2025. All six coefficients are significant at the 1 percent level.

the aircraft heading, the most capable sensors within the optics heading) share HS6 codes with the transferred goods, so no public-data control isolates them. Read together, the effect is stable in sign and magnitude and robust to the treated-set contamination, but its statistical precision rests partly on cross-category comparisons; identifying it strictly within category would need the firm- or transaction-level data the conclusion describes. This is why the within-category intensity evidence in the main text—the license dose-response and the per-shipment-friction gradient, both estimated with chapter fixed effects—carries weight: those supply the within-category variation the pooled ATT is thin on, and both are placebo-clean.

B.7 Robustness to the estimator

The headline uses the stacked estimator of Cengiz et al. (2019), which already rules out the already-treated comparisons that bias two-way fixed effects under heterogeneous timing. To confirm the effect is not an artifact of that choice, we re-estimate it on the same regime-gated treated set and defense universe with two other heterogeneity-robust estimators: the doubly-robust group-time estimator of Callaway and Sant’Anna (2021) with not-yet-treated controls and an entity-clustered multiplier bootstrap, and the interaction-weighted estimator of Sun and Abraham (2021).³⁵ All three agree (Table 12): the export effect is +0.15 to +0.18 and the world-share effect +0.13 to +0.16, each significant at one percent, with flat near-pre-trends and a common post-transfer build in every event study. The estimate is invariant to the estimator.

B.8 Robust inference under dose concentration

The headline uses analytic cluster-robust (CRV1) standard errors on HS6. Because the dollar dose concentrates in the aircraft family (dominant-family Herfindahl 3,047), a natural worry is that the effective number of independent clusters is smaller than the HS6 count suggests. Two checks address it (Table 13). First, a restricted (null-imposed) wild-cluster bootstrap with Rademacher weights, the standard few-cluster correction, is run at three clustering levels, HS6, HS4, and the HS2 chapter; the coarsest collapses the entire aircraft mass into one cluster. Second, an assignment-based randomization-inference test draws placebo treated sets of the same per-cohort size from the defense-products universe and rebuilds the design 500 times. The export ATT keeps a wild-cluster $p \leq 0.025$ and a randomization $p < 0.001$ across all levels. The world-share ATT is comparably robust down to HS4; at chapter-level clustering the BACI share panel spans only seven defense

³⁵The imputation estimator of Borusyak et al. (2024) is a further heterogeneity-robust option in this family.

Table 13. Robust inference for the regime-gated pooled ATT

Outcome	Inference	clusters G	CRV1 p	wild-cluster p
ln U.S. exports (+0.170)	cluster HS6	558	0.000	0.001
	cluster HS4	351	0.000	0.001
	cluster HS2 chapter	71	0.003	0.023
	randomization inference	—	—	0.000
U.S. world share (+0.147)	cluster HS6	191	0.002	0.002
	cluster HS4	99	0.002	0.005
	cluster HS2 chapter	7	0.007	0.101
	randomization inference	—	—	0.032

Regime-gated pooled ATT ($y \sim \text{post} \mid \text{uid} + \text{tid}$ on the stacked data). Wild-cluster p from a restricted (null-imposed) Rademacher bootstrap, $B = 1,999$; randomization-inference p (two-sided) from 500 placebo treated draws of the same per-cohort size from the defense-products universe. The world-share panel (BACI) spans seven defense chapters, so HS2 clustering leaves only seven clusters.

chapters, so the wild-cluster p rises to 0.10 on seven clusters, while randomization inference on the same estimate holds at 0.032. The point estimates are unchanged throughout, reproducing +0.170 and +0.147 by construction.

A related worry is that treatment timing was set by a handful of Federal Register rules, so the effective number of treatment *events* could be smaller still. The premise undercounts the design: the treated set spans fourteen 600-series ECCN families across all five cohorts, and cohorts vary within families. Re-clustering the export ATT on a rule-level construction, HS2 chapter $\times$ cohort for treated units and HS2 chapter for controls (102 clusters), returns +0.170 (SE 0.045, $t = 3.76$).

Two further checks address the dose concentration directly. A dose-weighted effective-cluster count, the inverse Herfindahl of the transferred-value shares, is 8.9 at HS6, 6.2 at HS4, and 3.8 at the chapter level, so the aircraft family pulls the effective number of clusters well below the nominal count. And because whole USML categories were assigned together, we re-run the randomization inference drawing placebo treated sets as whole HS4 headings and HS2 chapters rather than individual HS6. The export ATT keeps a block-randomization p of 0.008 at both levels. The world-share ATT is at the margin under HS4 blocks ($p = 0.05$) and not significant under the chapter block ($p = 0.20$), where the seven-chapter share panel leaves too few blocks to separate, the same soft spot the wild-cluster test shows. The export reduced form is the more robustly inferred of the two outcomes, and the world share the corroborating one.

B.9 Honest-DiD sensitivity in detail

Figure 7 traces the robust confidence set for the average post-period effect as the allowed relative-magnitude violation $\bar{M}$ grows. Two features of the design explain why the breakdown values are modest. The stacked window offers only two clean pre-periods, so the yardstick for a violation, the largest pre-period deviation, is imprecisely estimated; and the post-window is six periods over which the effect builds, so even a small per-period differential trend can accumulate enough to overturn the average effect. A smoothness restriction, which bounds the curvature of the differential trend instead of its magnitude, is if anything less favorable: the effect is significant only while the differential trend is held close to linear, and the robust confidence set includes zero once the allowed curvature reaches roughly a tenth of the curvature already present in the pre-periods (a smoothness

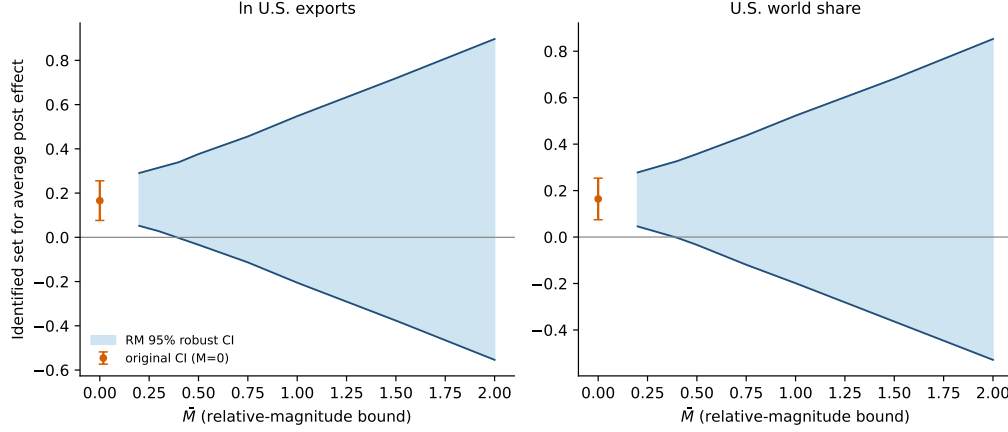


Figure 7. Honest-DiD relative-magnitude sensitivity (Rambachan and Roth, 2023), regime-gated treatment. Shaded band: the 95% robust confidence set for the average post-period effect as the allowed post-period trend violation $\bar{M}$, measured relative to the largest pre-period deviation, grows; the marked point at $\bar{M} = 0$ is the original confidence interval. The set includes zero beyond $\bar{M} \approx 0.3$ on both outcomes.

breakdown near 0.005 against an observed pre-period second difference of 0.067). Extending the pre-window sharpens neither restriction, because the deeper leads are thinly identified and add noise. These are properties of a smoothly adjusting response to a staggered policy, and they bound what any pre-trend-based sensitivity can certify in this setting.

B.10 Bindingness gradients: an ad-valorem benchmark and per-shipment friction

The cross-sectional dose-response of Section 12.1 admits two extensions summarized in the main text.

An ad-valorem benchmark. The slope permits a magnitude check against the trade-cost literature. Taken at face value, moving a product from zero to fully individually licensed trade is associated with a long-difference response of about 0.48 log points. Rather than assume a trade elasticity, we use one estimated for these very products: applying the across-source estimator of Feenstra (1994) to HS6 import data gives a median Armington elasticity of about 6.5 across the treated defense products (interquartile range 4.0 to 12.6; trade-weighted median about 10).³⁶ At that measured elasticity the 0.48 response rationalizes an ad-valorem equivalent for the licensing

³⁶Estimated on BACI import data following Feenstra (1994), matched to the regime-gated treated set: 61 of the 88 treated products carry a measured elasticity, covering 76 percent of their pre-reform export value. The treated-product median sits at the top of the conventional four-to-eight range used in the trade-cost literature, so the benchmark does not rest on an assumed value. The aircraft family, whose elasticity is estimated off few source countries and imprecisely, pulls the trade-weighted figure up; the unweighted median is the cleaner central value. Elasticities estimated on the wider set of products that ever carried a 600-series code, rather than on the treated set the design estimates, are materially lower (median about 4.5), so the set matters for this benchmark. One caution belongs with any elasticity from this family. The Feenstra (1994) estimator, as implemented at scale by Broda and Weinstein (2006), is subject to small-sample bias when few source countries are available, and Soderbery (2015) supplies a corrected hybrid estimator for that case. The aircraft codes sit squarely in it, which is a further reason to prefer the unweighted median over the trade-weighted one and to carry the elasticity band through to the welfare arithmetic rather than a point. We have not re-estimated with the corrected estimator, so the elasticities here should be read as the standard construction rather than the state of the art.

bundle of roughly 8 to 9 percent, depending on whether the elasticity enters as σ or as $\sigma - 1$, with the interquartile range on the measured elasticity carrying it from about 4 to 18 percent. That is a tariff-equivalent comfortably inside the range estimated for non-tariff measures generally (Kee et al., 2009). Two other estimates in the paper invert to the same wedge: the supplier-differenced triple difference of Section 10.3 gives 13.5 percent and the bilateral panel of Section 11.1 gives 8.1 percent, both after rescaling to a fully-licensed-product basis by the treated set’s mean dose of 0.27. The cautions on reading that agreement as corroboration—the routes are not independent, and the dose rescaling is what carries the figures to this range (unrescaled they give 3.5 and 2.1 percent; the linearity the rescaling assumes is non-monotone in the bilateral dose bins and unobservable in the licensed channels)—are stated where the numbers are used, in Section 14.2, which carries the supplier-differenced figure into the welfare arithmetic because it is already a relative quantity. The wedge is within the range time costs alone can support—each day in process is worth about 0.6 to 2.1 percent ad valorem (Hummels and Schaur, 2013), so one to two incremental weeks of adjudication prices at 4 to 30 percent—before counting registration, agreement provisos, and approval uncertainty, the components the policy-uncertainty literature prices separately (Handley and Limão, 2017). The magnitude is large for a friction whose adjudications were mostly approved, but it is not outside what the administrative bundle can carry.

A second intensity margin: per-shipment friction. The license dose measures how much of a product’s trade bore an individual license. A distinct dimension is how *often* the licensing apparatus bound: a product that shipped a given value in many small consignments incurred the per-shipment costs of licensed trade—filing against a license, decrementation of its authorized value, and fresh applications as orders and end users turned over—more often per dollar, so its constraint was more binding on that margin, the kind of per-shipment administrative friction that customs and border-delay studies find depresses trade (Volpe Martincus et al., 2015). Measuring this as the pre-reform number of shipments per dollar of value, from the same Census records and only weakly (negatively) correlated with the license dose ($\rho = -0.30$), the world-share gain rises with it: a one-standard-deviation increase in shipment friction raises the post-versus-pre change in world share by +0.083 ($t = 4.4$; chapter-clustered $t = 3.1$, within-chapter permutation $p < 0.001$) within chapter and *net of* the license dose, which itself stays at +0.078. The gradient is placebo-clean on world share (friction does not predict pre-period growth), robust to dropping aircraft and engines, and dominates a horse race against the license dose, destination breadth, and foreign-supplier count. It is clean only on the world-share outcome: on raw exports the same measure predicts pre-period growth, because small-consignment products such as parts and components sit on their own demand trends, a further reason to read the world-share result as the cleaner one. Two other candidate intensity measures do not generate comparable heterogeneity. The strength of foreign competition for the product (the “ITAR-free” substitution channel the satellite case suggests) and the share of a product’s latent demand sitting in STA-eligible allies both return placebo-clean nulls, so the gain does not visibly scale with foreign substitutability or with the STA extensive margin. This heterogeneity is a cross-sectional result; in the stacked difference-in-differences the analogous treated-group interaction has the right sign but does not survive a label permutation, so we rest the friction evidence on the cross-section, as a second bindingness gradient orthogonal to the license dose rather than a separate identification leg. A further reason for that restraint is that the export response did not arrive on the shipment-frequency margin at all. Decomposing the export gain into transaction count and transaction size (Section 12.2) leaves three quarters of it on size, and splitting the treated set at the median of this same friction measure puts the larger export gain in the *less* granular half (+0.20 against +0.14). We therefore read the friction gradient as describing which

products improved their world-market position, not the margin on which the gain materialized, and the two are consistent: a product that shipped granularly bore the licensed-trade apparatus more often per dollar, while the relief it received still showed up as larger authorized transactions.

B.11 Section 232 and 301 tariffs

The treated products’ post-transfer windows reach into 2018, when the Section 232 and 301 tariffs took effect, raising the possibility of a tariff confound. Three things rule it out. The effect predates the tariffs: restricting the stacked design to years through 2017 leaves the export ATT at +0.14 ($t = 3.1$), and the dose-response of Section 12.1, re-estimated on a strictly pre-2018 window, remains +0.47 ($t = 2.4$) on exports and +0.30 ($t = 2.0$) on world share. The measures are also the wrong object for the outcome: Sections 232 and 301 are U.S. import tariffs, and the only export-relevant channel, foreign retaliation, is minimal for goods whose market is allied and defense-customer concentrated. And no treated product sits on the Section 232 steel or aluminum chapters. The one qualification is that the world-share effect matures late, so the through-2017 stacked estimate of it is imprecise (+0.06); the pre-2018 dose-response, which has more power, confirms the share gain scales with the decontrol dose before any tariff.

B.12 No coding or reporting artifact

A distinct worry is that the measured export rise is a classification artifact rather than real trade: that the transfer, or a coincident nomenclature change, created a new HS6 code into which trade suddenly “appears.” It is not. An HS6 code is World Customs Organization product nomenclature, assigned by what a good is; an item’s export-control jurisdiction (Munitions List versus Commerce list) is an orthogonal classification and does not create, split, or renumber its HS6, so a good keeps the same HS6 across a USML→CCL transfer.³⁷ The panel does use as-reported HS vintages, so the WCO revisions of 2012, 2017, and 2022 add and retire HS6 codes, and a treated code born at a revision could in principle manufacture a jump. It does not happen here, for a structural reason and an empirical one. Structurally, the regime gate requires at least five percent of a code’s 2009–2012 export value to have shipped under ITAR individual licenses, so a code absent before 2013 cannot enter the treated set. Empirically, all 90 treated codes trade throughout the pre-period, with a median pre-reform export base of \$404 million; none first appears at or after its cohort year; and none is new in the 2017 revision that coincides with the fire-control rule, where all ten codes of the 2016 and 2017 cohorts are already present in 2016; the WCO correlation tables confirm that every chapter-90 fire-control and optical subheading is unchanged one-to-one from the 2012 to the 2017 nomenclature.

The world-share outcome adds a further guard. It is built from BACI, which reconciles U.S.-reported exports against partners’ mirror imports and holds each HS revision in a single constant-nomenclature release, so a purely U.S.-side reporting change would not move it. Because the effect appears in both the Census export series and the BACI share, a U.S.-reporting artifact is ruled out. The one nomenclature seam in the design is self-inflicted: the splice of HS-1992-vintage BACI (2009–2011) onto HS-2012-vintage BACI (2012 on) at the 2011/2012 boundary. Dropping

³⁷The ECCN and the Schedule B (HS) product code are separate fields on the same export record: the ECCN determines whether a license is required, the Schedule B classifies the good. The Census AES implementation bulletins for the reform (effective October 15, 2013 and January 6, 2014) added only new 600-series ECCN values and license types, and no new Schedule B or HTS commodity codes; the statistical-code committee that maintains U.S. reporting numbers has authority only over ten-digit suffixes, which do not change the six-digit aggregate used here.

Table 14. The headline is not a coding or reporting artifact

Treated set	<i>n</i> HS6	ln U.S. exports	U.S. world share
Baseline (regime-gated)	90	+0.169 (.045)	+0.147 (.048)
Drop Chapter 98 (“U.S. goods returned”)	88	+0.170 (.045)	+0.147 (.048)
Drop HS-2012-vintage codes	87	+0.156 (.045)	+0.138 (.048)
Core defense chapters only (84/85/87/88/89/90/93)	77	+0.193 (.049)	+0.147 (.048)
Drop all flagged artifacts	85	+0.157 (.045)	+0.138 (.048)

Pooled ATT re-estimated on the regime-gated treated set with the indicated codes removed; same universe, controls, windows, and fixed effects; SEs clustered on HS6. This table is the single home for the code-count accounting: the nominal regime-gated set is 90 codes, the two Chapter 98 special-provision lines are excluded from estimation everywhere else in the paper, and retaining them moves the ATT from +0.170 to +0.169. “HS-2012-vintage” are the three treated codes (850760, 930110, 930520) that first appear at the 2012 WCO revision. Every variant is significant at one percent.

the pre-2012 segment entirely, which removes the seam, leaves the ATT at +0.151 on exports and +0.172 on world share; dropping 2009 or 2014, the years of the Census FMS-coverage and temporary-export-filing changes, leaves it unchanged.

One classification difference between the two bilateral sources belongs on the record here, because the destination evidence of Appendix B.13 turns on it. U.S. exporters may file civil aircraft, engines, and parts under the Schedule B catch-all 8800.00.0000 rather than under the specific commodity code; the catch-all carried about \$552 billion over 2012–2016, 99 percent of it under civil-aviation license codes, and reconciled trade data redistribute that value back to the underlying HS6 headings while the U.S. filing records leave it out of them. That single provision is why aircraft and engines are 75 percent of treated value in the reconciled panel against 30 percent in the filing records, and why the filing-record HS6 aero series is far more military-concentrated than the reconciled one: the reconciled aircraft series absorbs the civilian catch-all, and the filing-record series is net of it. The catch-all grows smoothly across 2009–2019 with no break at any cohort year, so it generates no spurious discontinuity, and the destination estimates survive dropping the aircraft and engine chapters outright (Appendix B.13). Script 90.

Two minor artifacts survive the gate: two Chapter 98 special-classification lines (“U.S. goods returned,” 2.0 percent of the treated export base) and three codes created in the 2012 revision (0.3 percent). Dropping them, or restricting to the seven core defense chapters, leaves the ATT between +0.157 and +0.193 on exports and +0.138 and +0.147 on world share, each significant at one percent (Table 14). The trade response is growth on an existing base, not the appearance of a code.

B.13 Destination evidence in detail

The product-level outcome is the HS6 total, which mixes the decontrolled military variant with any co-moving civilian trade in the same code. Where the gain landed ought to indicate which of the two it is. A gain concentrated in the destinations that buy U.S. military goods is hard to reconcile with civilian co-movement; a gain spread evenly across trading partners is not.

The test runs on two bilateral sources, and we lead with reconciled third-party trade, because its estimate is insensitive to the sample rule this design needs (below) while the filing-record estimate is not. The U.S. filing records are nonetheless the source closer to the decontrolled good, because

within the treated codes they are far more concentrated on the military variant, for the Schedule B classification reason set out in Appendix B.12; and they corroborate on the same specification. Neither source resolves trade below the HS6.

In a triple difference with product-by-year and destination-by-year fixed effects and not-yet-treated defense controls, the differential gain to allied and major-arms-customer destinations is +0.389 (SE 0.107, $t = 3.64$) on reconciled trade and +0.726 (SE 0.140, $t = 5.20$) on the filing records, each against a largest pre-treatment coefficient below 0.03 (Table 15). Because treatment is assigned at the product-category level and clustered on a treated set concentrated in a few chapters, the setting where CRV1 standard errors are least reliable here, both differentials are put to a restricted wild cluster bootstrap: each clears the one-percent level on HS6 clusters and survives coarser clustering on HS4 and on chapter (Table 15, notes; Script 111), so the precision is not an artifact of the standard error. Narrowing the allied group to the STA-eligible A:5 list alone gives +0.202 ($t = 2.09$) and +0.453 ($t = 3.66$) respectively, muted because A:5 excludes several of the largest U.S. arms customers, among them Israel, the Gulf states, and Taiwan, which the broader grouping returns to the treated side.

That pre-period rests on a sample rule that both sources use, and its effects on the two sources differ in a way that decides which estimate should lead. Estimated on all cells the reconciled coefficient is +0.365 (SE 0.139) against a pre-treatment deviation of 0.345, a ratio near one; the filing-record coefficient is +0.266 (SE 0.123) on a 0.178 pre-trend. The diagnosis is the one Section 11.1 sets out for its own panel, and the rule is the same one: imposing pre-window completeness removes the pre-trend on both sources. On the reconciled source it leaves the point estimate essentially unchanged, +0.365 to +0.389, the behavior of a rule that cleans composition. On the filing records it nearly triples the coefficient, from +0.27 to +0.73, so there the rule does estimand work as well as cleaning: the filing-record figure is an estimate for continuously served trade relationships, not for the full panel. The rule is also not free. It retains about 42 percent of observations, cuts the reconciled treated set from 77 products to 64 (the filing records retain 74, because they carry U.S.-declared cells the reconciliation drops), and raises the allied share of surviving importer-by-product pairs from about a third to about half, because continuously served relationships are disproportionately allied ones. The estimates under the rule are therefore cleaner and narrower at once: they describe the differential among relationships the United States was already serving throughout the pre-period. The direction of the selection is worth naming, since the coefficient is itself an allied differential. What the rule drops from the comparison group is the intermittently served non-allied cells, leaving the most established of them as the counterfactual, which should work against finding an allied differential rather than for it—though that is an argument about the comparison group’s composition, not a test, and the rule-dependence of the filing-record point is the reason the reconciled estimate leads.

Two demand shocks would confound this estimate if either drove it, and both are concentrated in exactly these products: the Gulf procurement step-up after March 2015 and the allied fighter and rotary-wing deliveries that ran through 2014–2018. Neither does. Excluding the Gulf states leaves +0.376 ($t = 3.43$); excluding the F-35 partner nations and the large fighter and rotary-wing FMS customers leaves +0.499 ($t = 4.63$); excluding both leaves +0.497 ($t = 4.53$), all with pre-treatment deviations below 0.05. On the filing records the same exclusions give +0.685, +0.872, and +0.843. The differential is larger, not smaller, on both sources once the fighter-program destinations are removed; excluding the Gulf states alone barely moves it.

The two sources agree. Holding the cell set fixed to the cells both contain, so that only the reported values vary, the reconciled panel gives +0.748 (SE 0.136) and the filing records +0.784

(SE 0.140), a difference a quarter the size of either standard error, and both on pre-treatment deviations below 0.05.³⁸ The direction of that agreement is informative. On common cells both sources land near the filing records’ full-sample +0.73 rather than near the reconciled +0.39, so the lower reconciled figure comes from the third of its cells the filing records do not contain—the redistributed civil-aviation catch-all and small non-reciprocated flows—and not from the filing records overstating the trade both sources see, which is what a mechanical filing-side explanation would require. The full-sample gap between +0.726 and +0.389 is therefore about coverage, not about values. Which source better measures defense trade is not something this paper can settle; we lead with the reconciled source because its estimate does not depend on the sample rule, and read the filing records—more military-concentrated and more precise, but rule-dependent—as corroborating the direction on the source closer to the decontrolled good. Their magnitude is not corroborated: a coefficient that nearly triples under the sample rule carries the rule’s estimand with it, which is why no dollar claim in the paper rests on the +0.73.

Neither source resolves trade below the HS6, so neither cleanly separates the decontrolled military variant from co-moving civilian trade in the same code; the filing records come much closer, for the composition reasons above. Read with the pre-reform licensing gate, the administrative series of the Online Appendix (summarized in Appendix B.15)—where the Commerce-jurisdiction 600-series flow is separately observable and its top fifteen destinations, all allied or major defense customers, take 77 percent of first-rule export value—and the supplier-differenced panel of Section 11.1, the destination evidence supports reading the gain as the decontrolled military good rather than civilian co-movement.

B.14 The second stage in detail

This appendix reports the outcomes summarized in Section 13.

Production, employment, and innovation. Mapping the treated and control HS6 to NAICS6 through a trade-weighted concordance (Pierce and Schott, 2012) yields 22 treated and 120 never-treated manufacturing industries. Employment (−0.04), payroll (−0.03), value of shipments (+0.03), and value added (+0.04) are each indistinguishable from zero; confidence intervals exclude gains above about three percent for employment and fifteen to seventeen percent for output (Figure 8, left). Capital expenditure is the one positive (+0.32, significant under cluster-robust, wild-cluster, and randomization inference), but not identified: the only pre-period is the 2012 Economic Census at the post-sequestration defense-budget trough, so the comparison cannot separate decontrol from the procurement recovery; it is a flagged signal, not a result. Patents show no response (+0.03, flat event study; confidence interval rules out effects above 12 percent), and a pooled positive is a pre-trend that industry-specific linear trends absorb. The shock dilutes to about 1.5 percent of industry shipments—an order of magnitude thinner than the firm-level demand expansions in which trade-and-innovation links are found³⁹—and the industrial-base rationale is untestable at this

³⁸Scripts 108, 114. The shared cells are about two-thirds of reconciled cells and 94 percent of filing-record cells. These rows impose the pre-window rule, which is what makes them clean: without it the same comparison gives +0.400 and +0.468 against pre-treatment deviations of 0.39 and 0.43, larger than the coefficients, so the agreement was visible there but could not be read as a level. Dropping aircraft and engines on the all-cells sample leaves +0.386 reconciled and +0.319 on the filing records, so the pattern is not a property of the civil-aviation classification difference between the series.

³⁹Export-demand expansions raise firm patenting, concentrated in initially more productive firms and realized at a two-to-five-year lag (Aghion et al., 2024), and import-competition shocks have been shown to move domestic innovation in both directions (Autor et al., 2020; Bloom et al., 2016). Defense R&D shocks do move measurable

Table 15. Differential gain to allied and arms-customer destinations, by data source

	Estimate	(SE)	Wild p	max pre	Treated HS6
<i>Panel A. Reconciled bilateral trade</i>					
Allied and arms-customer destinations	+0.389	(0.107)	< 0.001	0.008	64
A:5 destinations only	+0.202	(0.097)	0.043	0.032	64
Excluding the Gulf states	+0.376	(0.109)	< 0.001	0.009	64
Excluding fighter programs ^a	+0.499	(0.108)	< 0.001	0.045	64
Excluding both	+0.497	(0.110)	< 0.001	0.032	64
<i>Panel B. U.S. filing records, same specification</i>					
Allied and arms-customer destinations	+0.726	(0.140)	< 0.001	0.026	74
A:5 destinations only	+0.453	(0.124)	< 0.001	0.035	74
Excluding the Gulf states	+0.685	(0.142)	< 0.001	0.018	74
Excluding fighter programs ^a	+0.872	(0.142)	< 0.001	0.030	72
Excluding both	+0.843	(0.145)	< 0.001	0.017	72

Stacked triple difference of decontrol, destination group, and post-transfer timing, with importer-by-product, product-by-year and importer-by-year fixed effects within each cohort sub-experiment, on arcsinh of bilateral U.S. export value; standard errors clustered on HS6. Both panels impose pre-window completeness: an importer-by-product cell must report a positive flow in all three pre-treatment years of its sub-experiment. Panel A is reconciled third-party trade, whose estimate is insensitive to that rule; Panel B is the U.S. filing records, which within the treated codes are far more concentrated on the military variant but whose estimate rises from +0.27 without the rule, so it reads as an estimate for continuously served relationships. “Allied and arms-customer” is Country Group A:5 plus the major non-A:5 U.S. arms customers. Scripts 106, 108, 109.

Wild p is a restricted wild cluster bootstrap- t (Rademacher weights, $B = 9,999$, null imposed), clustered on HS6 to match the standard errors. The allied-and-arms-customer differential also clears conventional significance under HS4 and under chapter (HS2) clustering on both sources; among the narrower cuts only the reconciled A:5-only estimate weakens, to $p = 0.11$ under HS4. Script 111.

^a F-35 partner nations and the large fighter and rotary-wing FMS customers of 2014–2018.

Without the pre-window rule the estimates are +0.365 (max|pre| 0.345) reconciled and +0.266 (max|pre| 0.178) on the filing records, on 77 treated products; the rule’s effects on each source are discussed in the text.

aggregation rather than rejected.

Diversion. Comparing U.S. exports of the affected products to re-export and sanctions-exposed hubs against exports to non-hub allies finds no diversion signature; goods flow toward clean allied destinations if anything. The null holds on both value and the extensive margin: the treated products served no more re-export hubs after decontrol (+0.06 additional hub destinations, $t = 0.5$), though that count carries a largest pre-period deviation of 0.326, five times its effect, so it is weakly identified. The result is bounded on two sides: it is a trade-discrepancy test, not an end-use test, and the hub panel omits the United Arab Emirates and Kyrgyzstan. The 2020 firearms transfer, where GAO documented post-transfer shipments rising fastest to high-risk resellers (GAO-25-106849), serves as a positive control: the same design recovers the extensive-margin signal on firearms (+1.9 high-risk destinations, $t = 3.0$) while returning null on value (+0.25, $p = 0.68$). The ECR null on both margins—the detector that registers firearms diversion finds none for the platforms—is a low-power bound on value and, given the count margin’s own pre-period, a weak one on destination count as well. Detecting arms-trade violations in market data has a methodological ancestor in

innovation at feasible industry aggregations (Moretti et al., 2025), which sharpens the reading: a market-access shock of this size is not such a shock. Tightening imposes large costs on the U.S. firms that supply designated entities (Crosignani et al., 2026).

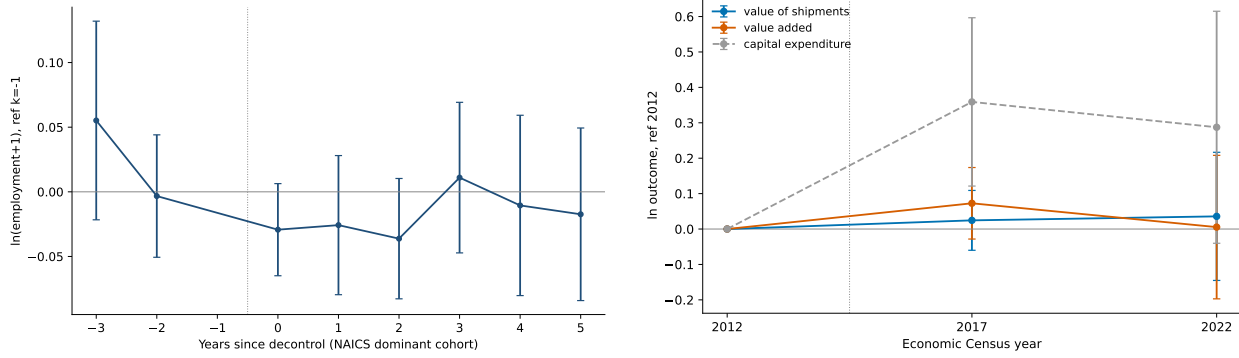


Figure 8. Industry second stage, regime-gated treatment mapped to NAICS6. Left: employment (Census County Business Patterns), flat after transfer. Right: value of shipments and value added (flat) and capital expenditure (rises, but not identified; see text), from the Economic Census. The left panel is in event time ($k = -1$ base); the right is in calendar time (Economic Census years 2012/2017/2022, 2012 base). 95% confidence intervals, clustered on NAICS6.

DellaVigna and La Ferrara (2010), who infer embargo violations from arms-maker stock prices around conflict events; trade-discrepancy tests like this one are the product-level counterpart, with less power against small flows. The comparison worth having in mind is Haidar (2017), who documents substantial deflection of Iranian exports toward non-sanctioning destinations after the 2008–2012 sanctions and shows it at the exporter level. That is the response a restriction can produce and a well-powered test can find. Ours is the mirror question, whether relaxing a restriction sends goods toward the destinations enforcement worries about, and the design would detect a deflection of that magnitude and does not detect one, while remaining uninformative about much smaller responses. The same exercise returns a +0.28 export gain (wild-cluster $p = 0.02$) on a cohort held out of estimation, though its own pre-period is not clean (Appendix B.4).

The post-transfer end-use monitoring record—Blue Lantern and BIS check volumes, rates, and what they can and cannot attribute to the transferred categories—is reviewed in the Online Appendix; its bottom line is that whether end-use risk rose for these specific goods is not identified in any public series.

B.15 Administrative cross-checks (Online Appendix)

The Online Appendix reports two administrative cross-checks on the treated-product gain: the BIS Office of Technology Evaluation series of Commerce-jurisdiction 600-series exports, compiled rule by rule from the Automated Export System, whose first-rule flow settles at about \$4.2 billion a year by 2019 with 77 percent of value going to the top fifteen allied and arms-customer destinations; and its State-side mirror, the Section 655 authorization series, in which the transferred categories' Munitions-List authorizations fall about 62 percent against 6 percent for the not-yet-transferred firearms controls. Neither series is an identification strategy; both corroborate the scale, destinations, and licensing composition of the object the design isolates.

Data availability and reproducibility. The two datasets that carry the results differ in their access conditions, and the difference bounds what replication can be offered. The Census ECCN export panel is built from Automated Export System filing records, which are restricted-access: the commodity-by-destination-by-license-type cells are not redistributable, so the archive deposits the

code that constructs the panel, the treatment and dose definitions, and the aggregated intermediate files underlying each reported table, but not the underlying records. BACI is obtained from CEPII under its terms rather than from the archive. Every other source is public and cited where used: SIPRI arms transfers, County Business Patterns, the Economic Census, PatentsView, the BIS Office of Technology Evaluation USML-to-CCL statistics, and the State Department Section 655 reports. Within those limits the pipeline is fully scripted. The regime-gated treatment and the headline stacked estimates (Table 2) are built in scripts `04b/05b` from the Census license-code field rather than the ECCN field; the ex-ante rollout schedule, with Federal Register citations verified against primary sources, is in `output/tables/ecr_sequence.csv`. A script-by-script map from each table and figure to the code that produces it, with instructions for obtaining the restricted component, is deposited with the replication archive.