The Trade Effects of Loosening Export Controls*

Philip A. Luck†

This version: July 2026
First draft: February 2026

Abstract

Evidence on the economics of export controls is almost entirely about tightening. The 2013–2017 Export Control Reform did the opposite, moving less sensitive military products from the U.S. Munitions List to the lighter Commerce Control List on a pre-announced, staggered schedule. We use these "600-series" transfers to estimate the effect of loosening, with treatment defined by pre-reform licensing status rather than by realized trade. Decontrol raised the affected products' exports by 15–18 log points relative to not-yet-treated defense products: roughly $4 billion a year of realized growth, and on the order of $6 billion once the decline the treated products escaped is counted. The gain concentrated in allied and arms-customer markets. Against rival allied suppliers (Korea, Japan, Germany, Italy, and the Netherlands), the United States kept ceding position in the products that stayed under ITAR and stopped where control was lifted. That contrast prices the competitive cost of licensing multilaterally listed goods more strictly than rivals' national regimes do, though the bilateral data locate the gain by market without isolating the decontrolled item within the traded code. The gain scales with each product's pre-reform licensing burden and arrives as larger transactions rather than more of them. The exposed industries' output, employment, and patenting show no detectable response. Because the trade data cannot value the screening the license bought, the reform's security cost is stated as a break-even rather than measured. Identification rests on the ex-ante rollout schedule and cannot rule out a category-graded, U.S.-supplier-specific demand shift; the estimates price this configuration of decontrol, not loosening in general.

JEL codes: F13, F14, F52.
Keywords: export controls, export licensing, ITAR, decontrol, defense trade, difference-in-differences.

*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. This is a working paper; comments are welcome.

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

1 Introduction

Export controls have moved to the center of U.S. trade and security policy,[1] yet credible estimates of their economic effects are scarce because the treatment is rarely quasi-random. The rate at which exports are individually reviewed, the natural measure of stringency, cannot serve as an instrument: its denominator is the trade the control itself chills. Firm-level designations such as the Entity List have been studied (Crosignani et al., 2026) but target firms, not products, and the October 2022 advanced-computing controls, though product-based, restrict end-users and destinations rather than reclassifying a product across control regimes. As far as we are aware, then, the literature has not addressed *product-level reclassification across control regimes*. That is the gap this paper fills. The 600-series transfer of the 2013–2017 Export Control Reform (ECR) is the cleanest available instance: a discrete, dated, staggered loosening of stringency on a defined set of military products.

The closest published precedent, Liu et al. (2024), exploits the 2007 U.S. "China Rule" through within-list product inclusion and exclusion. That is a different margin, and the opposite policy: they measure the target country's innovation response to *tightening*, whereas we measure the controller's export response to *loosening*. The 600-series design holds the item fixed and varies the *regime* it sits under, which makes it a distinct and, to our knowledge, unused research design.[2]

More broadly, the paper sits in 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), though the policy here is a licensing regime rather than a tariff. 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 rather than 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—the per-order margin of Section 3 imposed as tightening—and the sanctions-trade literature prices the cost of restriction directly (Crozet and Hinz, 2020; Felbermayr et al., 2020). Export controls differ from those embargoes in kind—a licensing-regime change rather than a bar on trade—and in measuring loosening this paper supplies the decontrol counterpart on the trade margin.

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

[1] On economic networks and chokepoints as instruments of statecraft, see Farrell and Newman (2019); for a formal treatment, Clayton et al. (2023).

[2] Descriptive 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 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.

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. That is a testable claim, and the 2013–2017 transfers are the largest and cleanest product-level decontrol available to test it: unlike the 2014 satellite or 2020 firearms transfers, they rolled out in a staggered, pre-announced sequence that supports a counterfactual. 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 (Section 12).

The same claim was made for the reform measured here, in numbers. A manufacturers'-association study projected $64 billion of added GDP and 340,000 jobs by 2019, and when the House Foreign Affairs Committee asked Commerce for a jobs figure in April 2013, Commerce answered with that study's export number, "up to an additional $60 billion a year in U.S. exports." Critics replied that individually licensed goods were under three percent of U.S. exports and that the United States already held most of the world market for them, so the gains would be negligible while diversion risk rose. The agencies produced no economic estimate of their own, quantifying only paperwork. Thirteen years on, none of it has been checked against outcomes; the estimates below are the first opportunity to do so, and the two sides turn out to have erred in opposite directions. That is the paper's second use, because a decontrol debate is running again on the same two kinds of claim.

We find four things. First, the main result is an avoided decline, and it is large in dollars. Applied to the affected products' export base, the estimates imply on the order of $6 billion a year relative to the no-decontrol counterfactual, of which about $4 billion is realized export growth and the remainder a decline the treated products escaped. Underneath that figure, the decontrolled products held their world-export share while comparable defense products' shares fell about 29 percent, a relative gain of roughly 15 percent, and their exports rose 15–18 log points against a falling control path. Section 12 decomposes the dollar figure and bounds the share attributable to the decontrolled military good.

Second, the gain landed in allied markets and shows up against rival allied suppliers. The differential gain to allied and arms-customer destinations is +0.39 log points on reconciled third-party trade, and roughly double that on the more military-concentrated U.S. filing records among continuously served trade relationships. In a reconciled bilateral panel comparing the United States against every other supplier of the same product in the same market, the allied gain is twice the gain elsewhere. Relative to the products that stayed under ITAR, the United States gained position against Korea, Japan, Germany, Italy, and the Netherlands—in levels, it kept losing ground to them in the still-controlled products and stopped losing it where control was lifted. No new markets were entered. Less favorably for a licensing story, nearly two-thirds of the realized dollar increase moved through government-to-government Foreign Military Sales, where an individual license is not the binding constraint—a share of the dollar change, not of the log effect, which the channel's removal barely moves (Section 9.3).

Third, the effect behaves as a licensing cost would. It scales with each product's pre-reform licensing burden—the dose—which is itself unrelated to pre-period growth, and it arrives on the

[3]Executive 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), whose preamble ties expanded UAV exports to "a strong and secure domestic UAV industry vital to the U.S. defense industrial base"; 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).

transaction margin that an order-based license taxes: three-quarters of the gain is larger shipments rather than more of them. Inverted through a trade elasticity measured for these products, the estimates put the ad-valorem equivalent of the licensing wedge at 8 to 13.5 percent when rescaled to a fully-licensed-product basis and 2 to 3.5 percent without that rescaling, across routes that are not independent of one another. That implies an administrative saving between roughly $0.3 and $1.3 billion a year, most of it cost lifted from trade that would have happened anyway. Because the calculation values the screening the license bought at zero, we use it as a break-even rather than as a welfare estimate: the reform is neutral if decontrol raised expected end-use losses above a threshold between about one and four cents per dollar of these products' trade, one cent taking the measured effects at face value and three to four under the rescaling (Section 12).

Fourth, no domestic footprint is detectable. Output, employment, and patenting in the exposed industries show no response, and the goods show no diversion toward the re-export hubs the data cover, though a positive control shows that diversion test is underpowered (Section 11). These nulls are bounds set by data resolution: the export gain is about 1.5 percent of the exposed industries' shipments, so even the industrial-base rationale working in full would not register at the industry level.

Identification rests on the exogeneity of the ex-ante rollout schedule. The design cannot rule out a demand shift graded by product category and specific to U.S. suppliers, though a supplier-differenced comparison rules out one common to U.S. and foreign suppliers, a limit Section 8 takes up and under which the estimates are read. The residual has a name: the allied fighter, rotary-wing, and Gulf procurement wave of 2014–2018, concentrated in aircraft and gas-turbine engines and supplied only by U.S. firms. Removing those two program families entirely leaves the export effect at +0.137, just below the 15–18 log-point headline range, and raises the world-share effect, so the confound is named and bounded.

Three contributions follow. First, the estimate is the first of its kind, and with it a treatment definition that transfers: building the treated set from the pre-reform licensing record rather than from realized post-reform trade separates dilution from selection, and any study of a reclassification can reuse it. Second, because the transferred goods stayed on the Wassenaar Munitions List throughout, the design isolates a quantity the arms-control literature has modeled without observing: the competitive cost of licensing a multilaterally listed good more strictly than one's rivals do (Levine and Smith, 2000; García-Alonso and Levine, 2007). Third, the paper prices the licensing regime on the same scale as a tariff, an ad-valorem equivalent for the administrative wedge, and turns the reform's contested security cost into a break-even; the same exercise gives the first outcome-based check of the forecasts made when the reform was debated.

Section 2 describes the reform and Section 3 reads the license it removed as a trade cost; Sections 4–8 define the design, present the trade results, and assess identification. Section 9 asks where the gain landed and Section 10 how it arrived; Sections 11–12 bound the domestic footprint, price the license, and score the record against the contemporaneous forecasts, and Appendix B collects the robustness.

2 The 2013–2017 Export Control Reform

The reform began from a judgment that the United States controlled too much, and Robert Gates put the premise in numbers. In an April 2010 address the Secretary of Defense observed that the government approved well over 95 percent of license cases for EU and NATO destinations, so a

system built for the Cold War was spending most of its effort on applications it was going to grant, and proposed to "dispense with the 95 percent of easy cases" so that "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 thus 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 was the instrument that required no bargaining, and it ran on an administrative timetable, not a legislative one.

What the transfers changed. 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.[5] 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.[6] 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.

A sequence set out in advance. 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

[4] Remarks by Secretary of Defense Robert M. Gates to Business Executives for National Security, April 20, 2010; the White House "Fact Sheet on the President's Export Control Reform Initiative," issued the same day, set out the initiative and its phases. The address contains no projection of exports, output or employment, a point we return to below.

[5] White House, "Fact Sheet: Announcing the Revised U.S. Export Control System," October 15, 2013, which states that "our most sensitive items... remain on the USML."

[6] 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 items destined to 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 chose instead to retain the EAR's pre-existing 25 percent 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 the aircraft end items in the largest treated family, its use additionally required a one-time prior BIS eligibility approval (15 CFR 740.20(g)).

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.[7]

One step in that sequence is a documentary anchor rather than a stated intention, and Section 8 builds on it. Removing an item from the Munitions List requires advance notification to Congress under section 38(f)(1) of the Arms Export Control Act. 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.[8] A statutory filing required *before* removal 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, and 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."[9] A customer mix that is mostly civilian is not a property of these products' trade performance, which is what the control logic requires.

What industry and its critics predicted, and what the government did not. Industry had pressed for the reform for a decade, and the satellite case was its exhibit. After a 1999 statute moved commercial communications satellites onto the Munitions List, the U.S. share of the world market for satellite manufacturing fell from roughly two-thirds in the late 1990s to about 40 percent by the mid-2000s, as foreign suppliers advertised "ITAR-free" hardware designed to avoid U.S. components.[10] Industry was not unanimous—the electronics-interconnect association opposed decontrolling printed-circuit designs, and organized labor predicted that lighter controls would move production offshore[11]—and the critics' economic prediction ran the other way on the aggregate (Hartung, 2013), while the government produced no economic estimate of its own. Section 12.3 collects the dated forecasts from each side and scores them against the estimates, so the specific numbers appear once, where they are checked.

[7]Effective 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 rather than 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, the fire-control category from a May 2015 proposal and Categories XIV and XVIII from a June 2015 proposal (80 FR 34572; 80 FR 34562), both to a December 31, 2016 effective date. The slippages had administrative and political causes unrelated to the products' trade, which is the property the not-yet-treated control logic requires.

[8]Executive 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 22740 (State), April 16, 2013, effective October 15, 2013.

[9]Testimony 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.

[10]The move to the Munitions List was directed by Section 1513 of the FY1999 National Defense Authorization Act; Section 1261 of the FY2013 Act restored the President's authority to move satellites off it, and most of Category XV transferred in 2014 to the parallel `9x515` Commerce classifications created for the purpose (79 FR 27418). The market-share figures are industry and U.S. Air Force industrial-base estimates; the recovery was under way before the 2014 rule, so they motivate the reform rather than measure its effect.

[11]Testimony of the Association Connecting Electronics Industries (IPC), House Foreign Affairs Committee, February 7, 2012; letter of Thomas Buffenbarger, International Association of Machinists, March 2013.

Oversight bodies raised a different objection, about the government's own capacity rather than exporters' costs. GAO found that State and Commerce had not fully assessed what list reform would do to the resources their compliance and end-use-monitoring activities would need, and reported that against as many as 30,000 transferred license applications Commerce planned 850 end-use checks in each of FY2013 and FY2014, unchanged from the year before (GAO-12-613). Enforcement officials told GAO separately that shrinking the pre-export information stream would move checking downstream toward post-shipment verification and would require better pre-export targeting (GAO-12-246, p. 29). The fullest outside evaluation of the reform's implementation concluded that it tilted the balance toward trade at the expense of controls (Stricker and Albright, 2017). The industrial-base and competitiveness rationale industry advanced is close to the one the current drone and AI-export orders invoke, and it is the claim the estimates below assess; the domestic results of Section 11 bound the offshoring prediction in the same exercise.

The multilateral baseline held, and the goods stayed on it. One feature of the reform decides how its competitive effects should be read, and it is easy to state backwards. 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."[12] 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. Meanwhile the baseline itself held. 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.[13] One distinction keeps the reading precise. The Wassenaar Arrangement harmonizes what its members *list*, 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. A shared baseline here means a shared list of controlled goods, not a shared licensing threshold the United States exceeded. 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.

[12] 78 FR 22660, at 22661 and 22668.

[13] Wassenaar Arrangement, Lists of Dual-Use Goods and Technologies and Munitions List, documents WA-LIST (13)1 through WA-LIST (17)1, with the annual Summaries of Changes; the munitions-list edits in the window are sub-entry technical revisions concentrated in energetic materials (ML8), leaving ML5, ML6, ML9, ML10, and ML11 in place. 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). That the transferred goods remained on the WAML throughout is confirmed independently by the transfer rules, which describe 600-series items as "WAML and formerly USML items subject to the jurisdiction of the EAR" (78 FR 22660, pp. 22661, 22668).

3 A licensing regime 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 and the oversight record. 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. Adjudication took time even for the government's own agencies: responses to enforcement bodies' requests for license determinations on defense items ran 10 and 93 days against goals of 3 and 45 by FY2010, an indirect but contemporaneous measure of the apparatus's pace.[14] 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 sits above both and is easy to miss, because it is not a shipment license. 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.[15] 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 10.2 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.

[14] GAO-12-246, Table 5. The requests support seizures and pre-trial certifications, and the goals were set by the DHS/ICE-led Exodus Command Center rather than agreed by the licensing agencies, so these are interagency response times, not exporter license-processing times. Registration is a separate standing cost: fees ran in three tiers, $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 (testimony of Marion Blakey, House Foreign Affairs Committee, February 7, 2012).

[15] GAO-13-119R, November 16, 2012, which enumerates seven compliance and three enforcement activities. 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. The report describes the rules as proposed in June 2012, so it establishes the regulatory anatomy rather than the implemented safeguards. In its first twelve months STA carried 74 exporters, 604 transactions and $33.9 million.

What the logic predicts. 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. 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. 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 makes the mechanism testable rather than assumed, 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 10.2 reports that decomposition, which the Census filing records support directly because they count consignments as well as valuing them.[16] Either version is intensive-margin in 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.[17] 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 10.2 and 9.1 report all three, so the extensive-margin null there is what the regulation requires rather than a behavioral finding that happens to fit.

What the logic does not predict. 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 rather than 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 12 returns to. One channel runs through third-country manufacturers rather than U.S. exporters. 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

[16]The 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 of Section 10.2 rejected it, and this framework was rewritten around the per-order structure. We state the pair as a discriminating prediction because either outcome would have been informative and the one the data selected was not our prior. The regulatory support in the next sentences was found afterwards and does not depend on the estimate.

[17]Note 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.

standing reason to design that content out, and the "ITAR-free" hardware of the satellite episode is the visible form of it. A 25 percent *de minimis* removes the penalty below the threshold, which predicts a shift back toward U.S. supply against the allied industrial suppliers capable of building the substitute. Section 9.1 finds that pattern, though the cross-sectional test in Appendix B.9 keying on exposure to this channel returns a placebo-clean null, so the *de minimis* route is consistent with the bilateral result rather than demonstrated by it. Finally, the framework makes no prediction about the channel that carries most of the money. Government-to-government procurement through the Defense Security Cooperation Agency is routed exclusively to U.S. firms and negotiated between governments, so neither the competitive-fringe logic nor the per-order adjudication binds it, yet Foreign Military Sales account for most of the realized dollar increase. Section 9.3 sets out that decomposition; what drove the FMS rise is a question these data inform without settling.

4 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 6 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 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 rather than creating a 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

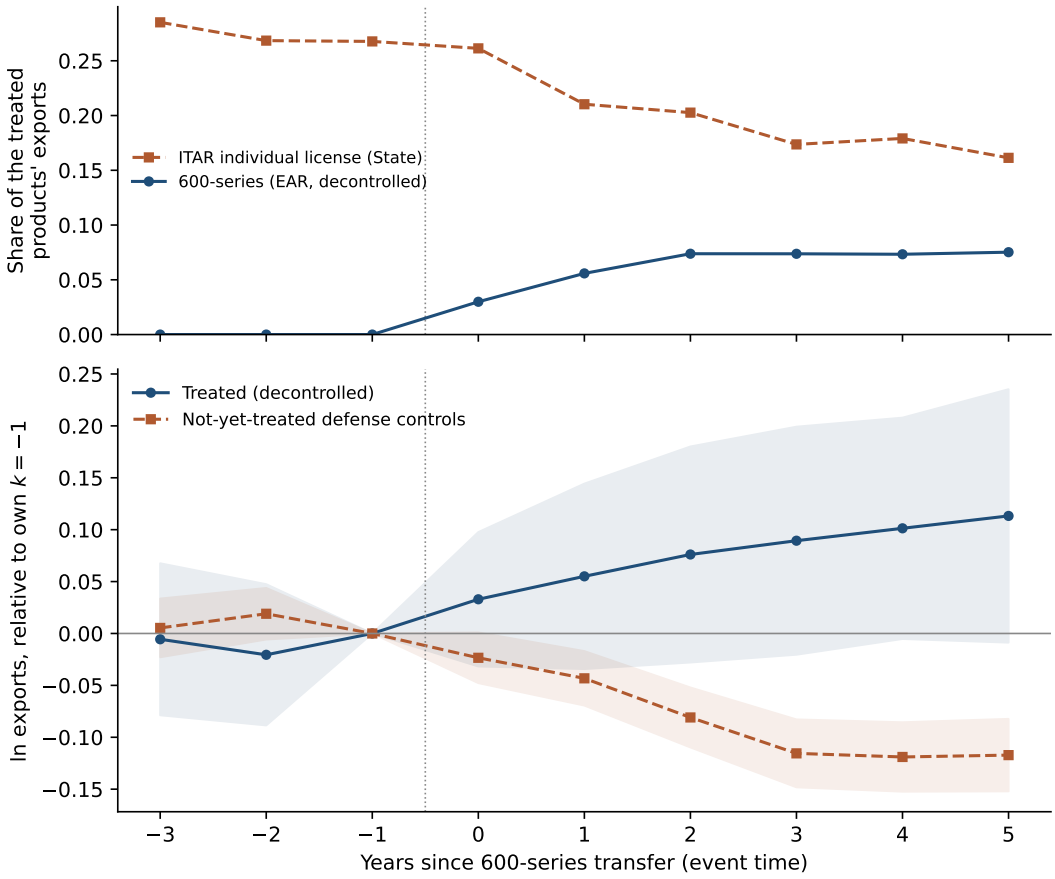

Figure 1. The decontrol experiment. Treated products are the decontrolled defense goods (the regime-gated treated set, Section 4); controls are defense products not yet transferred at the same event time. *Top:* averaged across the treated products and by event time, the share of their exports moving 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.

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.[18] 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

[18]The 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 U.S. Census Bureau, *Automated Export System Trade Interface Requirements* (AESTIR), Part III, Appendix F ("License and License Exemption Type Codes and Reporting Guidelines"); `C59` designates the Strategic Trade Authorization exception and `S94` designates Foreign Military Sales in the same appendix.

cumulative transferred value is distinct from the annual export base of these products (about $32 billion *a year*, Section 12), which the two figures coincidentally approximate.

5 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.11 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 rather than 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 6.

6 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).[19] Late-transferring categories, including firearms (USML Categories I–III, transferred in 2020), serve as clean not-yet-treated controls for the early cohorts.[20] 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

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

[20] The 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.

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 rather than 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

$$y_{i,t,g} \;=\; \alpha_{i,g} \;+\; \gamma_{t,g} \;+\; \sum_{k \neq -1} \beta_k \, \mathbf{1}\{t - g = k\} \cdot \mathrm{Treat}_i \;+\; \varepsilon_{i,t,g}, \tag{1}$$

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.

7 The export gain and the licensing relief

Table 1 reports 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).

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.7 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.7, B.3). As the data section notes, the arcsinh share coefficient reads as a proportional change, and the log of the share on the positive-share subsample returns the same +0.169; the tables report the arcsinh 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

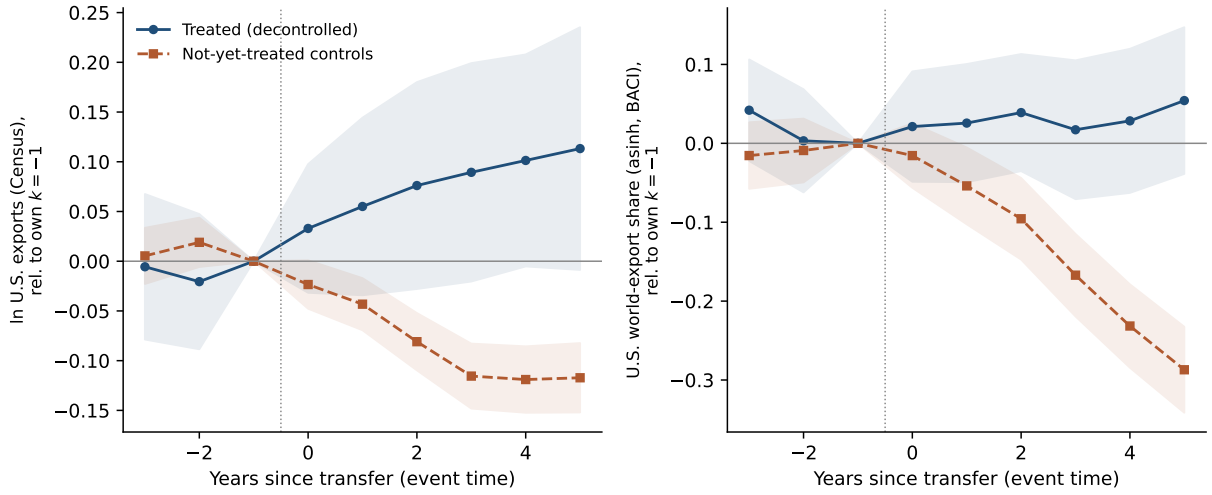

Figure 2. The difference-in-differences in group means. Each series is a within-product change relative to the product's own value the year before transfer ($k = -1$), averaged across the treated (decontrolled) products and across the not-yet-treated defense controls; bands are 95 percent intervals across 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.

while comparable defense products lose about twenty-nine 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-nine-percent endpoint because it averages 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, 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—rather than the effect of decontrol against a stable counterfactual; the two coincide only to the extent that the controls' decline is itself the cost of continuing control, an attribution Appendix B.4 tests where it can and otherwise leaves as an interpretation. The substitution reading fails where its prediction is sharpest, on the closest substitutes (Appendix B.4), but the distinction is why we rely on export volume as the primary outcome and state the share result in relative form: it is real but rests on the control group's trajectory.

What the estimate survives. Table 2 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 8.3, 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, the two categories where the allied fighter programs of these years ran, leaves +0.137 on

Table 1. 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)	146	+0.181 (.033)	+0.106 (.039)	.082 / .021
Full-eligible (ever 600, any value)	1,179	+0.061 (.023)	+0.047 (.031)	.019 / .013
Regime-gated (ever 600 $\cap$ $\text{ITAR}_{\text{ind}} \geq 5\%$)	90	+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 nominal treated-set definition. For the regime-gated set, two Chapter-98 special-provision lines ("U.S. goods returned") are excluded from estimation as reporting artifacts, leaving an 88-code export sample; a further eleven lack a BACI world-share panel, leaving 77 for that outcome (and correspondingly for the other definitions). The reported ATTs use the 88-code sample throughout (+0.170/+0.147); retaining the two Chapter-98 lines (the nominal 90-code set) gives +0.169, an immaterial difference shown in the artifact check of Table 12. All estimates are significant at the 1 percent level except the full-eligible set, whose export estimate is significant at 5 percent and whose world-share estimate is not significant at conventional levels.

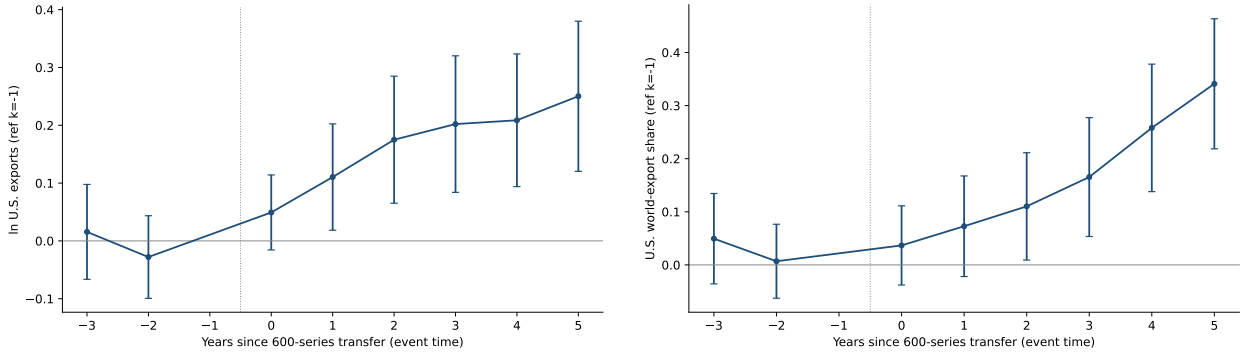

Figure 3. Stacked event studies, regime-gated treatment. Left: ln U.S. exports (primary). Right: U.S. world-export share (demand-differenced). 95% confidence intervals, clustered on HS6. Both pre-trends are flat.

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.7). Appendix B collects the 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 (Figure 4): trade rises as the individual-license burden falls. Neither margin is a sharp jump at the effective date; the licensing

Table 2. The regime-gated effect across the specifications that challenge it

Specification	ln U.S. exports	U.S. world share
Baseline stacked estimator (Table 1)	+0.170 (.045)	+0.147 (.048)
Estimator		
Callaway–Sant'Anna	+0.184 (.048)	+0.163 (.037)
Sun–Abraham	+0.150 (.042)	+0.135 (.036)
Demand		
+ 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)	—
Treated set		
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)
Window		
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; Appendix B.6 reports the estimator comparison and Table 12 the Chapter-98 sensitivity. Sections 8.3, B.10, B.1 and B.5 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.10).

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.

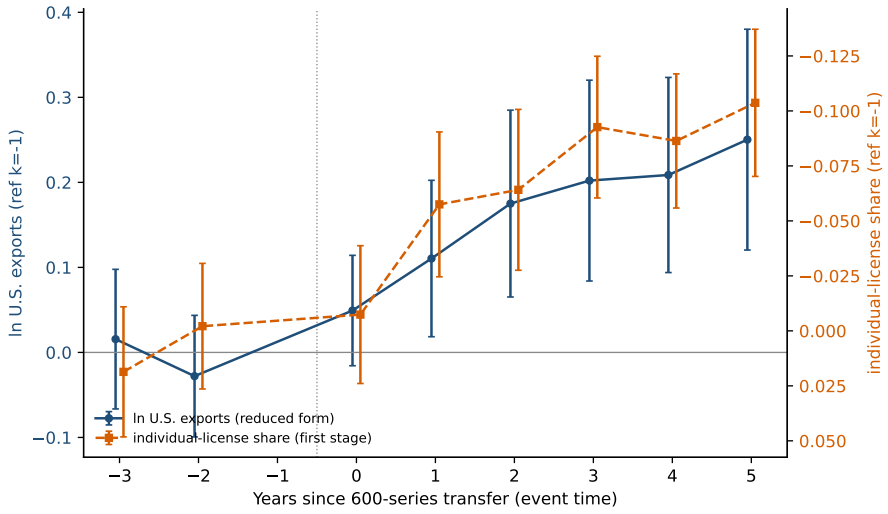

Figure 4. The licensing channel. Individual-license share (right axis, inverted) and ln U.S. exports (left axis), regime-gated stacked event studies. Both are flat before the effective date and build after it: trade rises as the individual-license burden falls. 95% confidence intervals, clustered on HS6.

8 What the timing was set by

8.1 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 $t = 0.87$, $p = 0.39$), though the differential's confidence interval is wide, and Appendix B.4 prints the implied-bias arithmetic rather than resting 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 it is worth separating what the contemporaneous documents say from what was said about them afterwards. The order was fixed years before the trade it might have tracked (Section 2). Two determinants appear in the record at the time, and neither concerns trade. 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 (Category VII) first because it was the drafting template the rest would follow. The contemporaneous account of what governed progress thereafter is procedural: testifying in April 2013, officials described the remaining categories as either undergoing or awaiting interagency review. The strongest documentary anchor is the statutory one described in Section 2: 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; the Congressional Research Service account the literature relies on picks the sentence up from that hearing (Fergusson and Kerr, 2020). 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 rather than a fact that helps it. Two things keep 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 8.3 then confronts directly with measured allied demand.

The cleanest illustration is the category whose timing was set by something unrelated to the review queue: firearms and ammunition (Categories I–III) were held back inside the reform window for the customer-mix reason set out in Section 2, a property of who buys them rather than 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 rather than 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 being 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).

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.

8.2 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.8). This is a modest breakdown, and we read it as quantifying the identifying assumption rather than 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 8.3, is more favorable rather than less. Its pre-period deviations are somewhat larger (max |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.[21] 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 10.1 corroborates it without using the event-time path at all.

8.3 Category-level demand and same-tier controls

The confound the cross-sectional evidence cannot fully rule out is a demand shock concentrated in the categories that transferred first, an allied-rearmament wave in the interoperability-central items after 2014 being the leading candidate. 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, close to baseline, and the world-share ATT roughly halves to +0.08; neither is 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, leaving the specification to rely on the mixed chapters (electronics, engines, optics) that carry fewer treated products.

The second test measures the confound directly rather than 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; the mapping covers 93.8 percent of the dollar dose, and the small-arms control categories, outside SIPRI's major-weapons scope, are covered by a product-level BACI variant. Two facts emerge (Table 3). 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); the early-versus-late gradient exists, but it is driven by a collapse of allied procurement in the late and control categories, it is not monotone in transfer order (armoured vehicles, a 2014 cohort, 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. The collapse of measured demand for the control categories also corroborates the reading of the control path in Section 7, with the non-U.S.-supplier evidence of Appendix B.4 indicating that U.S. suppliers bore that decline.

Second, controlling for the measured demand path leaves the effect standing. The preferred SIPRI 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, and the demand-controlled event study keeps flat pre-coefficients (max |pre| = 0.042 exports, 0.028 share). Two features of the SIPRI control qualify how far it reaches. Its delivery covariate is defined only on the roughly sixty-four major-weapons codes the crosswalk maps, so the full-sample specifications carry the unmapped codes on the baseline, and on the mapped subsample alone

[21] Identical at every fixed-effect rung, including product-by-year absorption, which is the numerical signature of the orthogonality discussed in Section 8.3. Script `100`.

Table 3. 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 × 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 1; 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`.

the export effect is too imprecisely estimated to sign (-0.10, $t = -0.4$, $N = 964$), though the share stays weakly positive. 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 (final row, Table 3), both significant at one percent, and it is the cleaner of the two demand controls. 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 ($t = 2.4$). The fully flexible variant, demand-quartile-by-year fixed effects, restates the power concession above: with aircraft always in the top demand quartile it approximates category-by-year absorption, and the export effect is unidentified there while the share holds marginally. What remains of the demand objection is therefore a category-graded confound invisible both in SIPRI deliveries, from any supplier, and in world trade of the products; the measurable version is measured here and does not overturn the headline.

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, $t = 3.4$), and the estimate is numerically unchanged whether the specification absorbs demand at the cohort-by-year, the chapter-by-year, or the product-by-year (HS6-by-year) level; restricting the supplier-differenced comparison to non-aircraft treated codes (removing the ten HS88 codes) leaves the estimate at +0.188 ($t = 3.3$), ruling out the concern that Airbus and Embraer civilian-aviation comparators drive the result. The identifying variation is how U.S. exports of the treated products diverge from foreign exports of the same products, so any demand path common to a product's suppliers is orthogonal to it, and the fixed effects that would absorb such a path leave it intact (Appendix B.4).

This states the world-share logic of Section 7 as an explicit difference-in-difference-in-differences, and it avoids the two limits of the tests above: it keeps the within-category power that chapter-

by-year absorption gives up, and it does not depend on SIPRI's product coverage. Its content is the same avoided-decline structure, a U.S.-specific decline in the control products rather than an absolute rise in the treated ones (Appendix B.4), so it is not an independent identification leg but the world-share result read against demand. Its own pre-period tolerates violations up to $\bar{M} = 0.4$ before the robust confidence set admits zero, above the breakdown of either headline estimate (Section 8.2). What it cannot difference 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.

The concrete form of that residual is the Foreign Military Sales channel, supplied only by U.S. firms and so invisible to the supplier difference; Section 9.3 takes it up.

A separate concern, the treated-set contamination and the absence of a clean same-tier control, is taken up in Appendix B.5: 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, so the pooled estimate's precision rests partly on cross-category comparisons—the reason the within-category intensity evidence of Section 10 carries weight.

8.4 What the design still cannot do

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.

The bilateral evidence of Section 9.1 indicates the gain tracks defense-trade destinations rather than spreading as civilian trade would, and the supplier- differenced bound of Section 8.3 rules out demand shocks common to U.S. and non-U.S. suppliers of the same product. The destination triple difference of Section 9.2 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.

9 Where the gain landed

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 8.4 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 that follows depends on the pre-window sample rule the market comparison also uses, and we set it out where it appears in the argument, not at the end. The third cut bears on whether licensing relief is the operative mechanism, and it is the least favorable evidence in the paper.

9.1 Market position against rival suppliers

The Census panel behind Section 7 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.[22]

Table 4 collects 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. It is stable dropping aerospace (+0.116), restricting controls to the same

[22]Weighting 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, and the coefficient it delivers is the top one percent of cells. We do not use it. Scripts 97–99.

HS4 (+0.106), and dropping the Chapter-93 contamination (+0.105). Because treatment is assigned at the product-category level, we subject it to the same cluster coarsening the headline faces in Appendix B.7: 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.[23] 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 9.2 reaches the same allied concentration on a different contrast, comparing destinations rather than suppliers, and on a treated set the same sample rule thins considerably further, so the two are corroborating rather than 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 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. What the coefficients record is therefore not the United States taking sales from these suppliers in the decontrolled goods, but the United States continuing to cede position to them in the goods it kept licensing above its rivals' national standards, and ceasing to 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 2, 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. Three qualifications attach. The comparison assumes rival suppliers' own regimes and export postures were stationary, which the design does not test, and the failures of that assumption do not share a direction: European small-arms and Middle East licensing tightened in these years, which inflates the affected rival coefficients, but Japan's April 2014 Three Principles on Transfer of Defense Equipment replaced a near-total export ban mid-window, and Korea's state-backed export expansion ran through the same years, changes that work the opposite way on exactly the two largest coefficients; the rival coefficients are therefore not uniformly upper bounds, and we read the cross-rival ranking at its sign only. They also rest on this one source, because the military-only

[23] Script `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.

comparison that would corroborate them independently turns out not to be available, for reasons set out below. And the estimate is a private competitive return: whether the stricter treatment the United States gave up was buying security the multilateral baseline does not is a question Section 12 states as a threshold rather than 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 with a breakdown of 0.40), 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 United States did not begin serving markets it had been absent from. The gain is deeper sales into relationships that already existed, the margin the per-order reading of Section 3 implies, and the absence of an entry response is also why conditioning on positive flows does not bias the estimates above.

Two limits bind. 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 that limit 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.[24]

9.2 Destination evidence on the treated good

A second cut asks the same question across destinations rather than 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 9.1 is imposed or not. The U.S. filing records give

[24]No cut is significant, the signs are mixed across rivals and negative pooled, and in every specification the largest pre-treatment deviation is as large as the estimate or larger, the same failure the dose test above exhibits. Script `110`, documented in `findings_sipri_competitor_2026-07-26.md`. The SIPRI register for 2005–2024 is mapped to Munitions List categories at the weapon-description level and stacked on this paper's cohorts, with Category IV munitions and Category II artillery as controls, since fifteen of the twenty-one Munitions List categories transferred inside the window and little else is both visible to SIPRI and untransferred. Japan cannot be tested at all: it supplies four deals in the entire register, its export restrictions having been eased only in 2014, which is a reminder that the position gained against Japan is in components rather than in platforms.

Table 4. U.S. market position against rival suppliers of the same product

	Estimate	(SE)	max|pre|	Breakdown
Panel A. Pooled and by destination				
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
Panel B. Robustness of the pooled estimate				
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)	—	—
Panel C. Fixed-cell rivals[a]				
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
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
Panel D. Extensive margin[b]				
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 × HS6 × 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. Korea, Japan, the Netherlands, Italy and Germany survive the wild cluster bootstrap at the one percent level with a breakdown of at least 0.20; the remaining rows do not survive 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.

+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 rather than 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.12 reports both sources, the sample rule, and what the rule selects on.

9.3 The channels the gain ran through

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 increase largely does.[25]

Is the FMS rise itself a consequence of the reform? The evidence here does not close the question. What the data do establish is 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 that a general expansion of allied procurement does not predict.[26] The test that would settle it lacks the power to. Restricting to bilateral FMS flows and absorbing destination-by-year fixed effects compares

[25]The 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 four (0.056); that this channel carries the dose gradient below while showing no channel-level effect of its own is a tension we do not resolve. 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`.

[26]The 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; carried to FY2017 the ranking reverses on the denominator alone (worldwide deliveries +100 percent, treated +50 percent, controls +10 percent). 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.

treated against control defense products bought by the same government in the same year, which no country-level rearmament path survives. On a panel of 7,683 product-destination-year cells, however, 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 rather than as a null.

The concentration also sharpens a structural limit of the demand evidence in Section 8.3. The supplier-differenced triple difference removes demand common to all suppliers of a product, but FMS is U.S.-supplier-specific by construction: a rise in allied FMS procurement concentrated in the decontrolled categories would register as a positive U.S.-versus-foreign export gap and would be indistinguishable from a supply-side response to decontrol. The SIPRI evidence that allied purchases in the early-transferred categories tilted toward non-U.S. suppliers constrains that reading without closing it.

One further pattern qualifies the licensing mechanism rather than the effect. Running the dose-response of Section 10.1 channel by channel, the gradient in the total (+0.48, $t = 2.6$) is carried by commercial and other EAR-authorized value (+0.79, $t = 3.2$), while the two channels in which an individual license actually bound are estimated far too imprecisely to sign. The benign reading is arithmetic, since commercial and other value is 98.5 percent of these products' trade; the less favorable reading is that the gradient sits where the license did not bind. These data do not adjudicate between the two.

10 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 8.2 constrains. Together they say whether the mechanism the framework names is the one visible in the data.

10.1 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 gives a design that does not use the path at all. 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 rather than widens 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 $p = 0.014$ 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.[27] The gradient is monotone across dose quartiles (Figure 5), 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 7. 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 rather than tracing an 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 9.3 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.

The leading candidate is an allied-rearmament demand shock concentrated, like the rollout order itself, in the interoperability-central categories that transferred first; Section 8.3 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. Two features bound it further: the dose variation is within-sector, so the confound must act through licensing intensity rather than 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 assumption is consistent with it but underpowered, and we do not read it as decisive. Latent allied demand for each transferred product, measured as the allied and arms-customer destinations' imports of the same HS6 from non-U.S. suppliers and so exogenous to U.S. decontrol, does not rise significantly with the dose within chapter, but the point estimates are large and imprecise ($+1.03$, $t = 0.89$ in the export sample; $+0.70$, $t = 0.59$ in the share sample): the confidence intervals admit a latent-demand gradient comfortably large enough to generate the observed effect, so the test fails to reject a confound rather than ruling one out. 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 enters on a mismatched scale and is orthogonal to the dose by construction of the first result, so it could not have moved the coefficient.

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 rather than a second identification leg; the ex-ante rollout schedule of Section 8 remains the one genuinely independent leg. What it adds is that the gain scales with the licensing burden the transfer removed, which is also what the dilution pattern of Table 1 implied.

Magnitude and a second intensity margin. Two extensions of the dose-response, developed in Appendix B.9, 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

[27] Some 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.

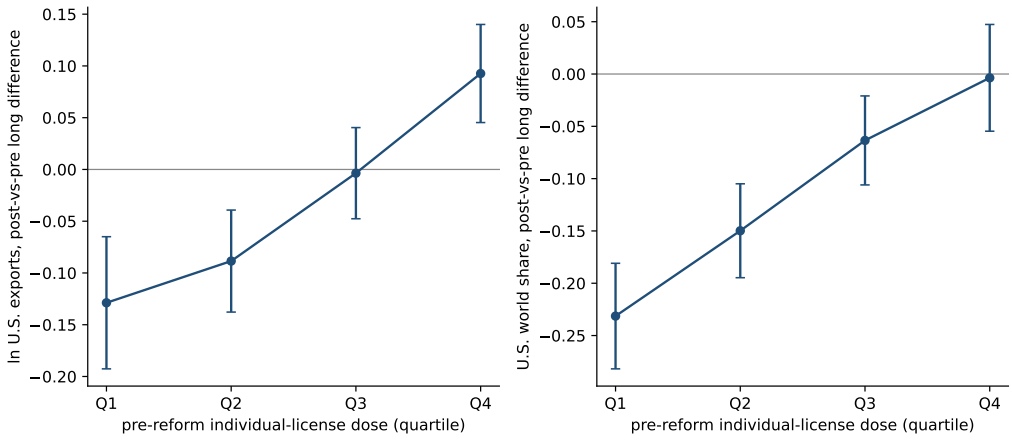

Figure 5. Cross-sectional dose-response, transferred products (ever-600, cohorts 2013–2017). Mean post-versus-pre long difference by quartile of the pre-reform individual-license share (the decontrol dose). Left: ln U.S. exports; right: U.S. world share. Both rise monotonically with dose, with quartile 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.

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 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.9 sets out.

10.2 Which transaction margin the gain arrives on

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: $\ln(\text{value}) = \ln(\text{shipments}) + \ln(\text{value/shipments})$, 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 as large as its post-transfer effect. By the discriminating prediction of Section 3, 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 the decomposition rejects that published prediction rather than merely failing to support it.

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 rather than 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.[28] The same asymmetry appears in the destinations where the gain concentrates: in the STA-eligible allied markets the size margin is +0.143 against +0.088 elsewhere. It is also consistent with the absence of any entry response in Section 9.1: relief that raises the value a single authorization can carry deepens existing relationships

[28] 15 CFR 758.1(b). The value margin estimated on this independently constructed panel returns +0.170, reproducing Table 1. Script `101`.

rather than opening new ones.

The decomposition has one composition threat, and it is the serious one. 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).[29] 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.

A second version of the same worry is worth ruling out, because the aircraft and gas-turbine engine families are both where the platform programs of these years ran and where the lumpy contracts are. Splitting the treated set at those two families and estimating on non-FMS value, the decomposition holds on each half separately. For the 40 program-family products the gain is +0.185 on value, +0.020 on the count margin (wild-cluster $p = 0.74$) and +0.166 on transaction size ($p = 0.012$). For the other 48 it is +0.132, +0.064 on the count ($p = 0.19$) and +0.069 on size ($p = 0.049$). Size carries close to nine tenths of the program families' gain and about half of the rest, so lumpiness does inflate the share where platforms are, but the count margin is a null in both halves and the size margin is significant in both. The mechanism is not a property of the aircraft cohort.[30]

11 No detectable footprint beyond trade

The trade result raises a second question: does decontrol leave 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 rather than 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.13 reports each of these.

The arithmetic explains the pattern and is the reason to read the nulls as bounds. The export

[29] The 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 9.3. 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 9.3 only in sample: the decomposition requires cells with positive non-FMS shipments, while the channel estimate retains the few cell-years with none. Script `104`.

[30] Script `106`.

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 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.

12 Magnitude and the security tradeoff

12.1 The size of the gain

The treated products exported about $32 billion a year before reform and $38 billion after. 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, a range of $5.5 to $6.7 billion across specifications. Two decompositions discipline that figure. 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 (Table 14), 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: roughly two-thirds of the realized dollar increase ran through Foreign Military Sales (Section 9.3), where an individual license did not bind, a caveat the identified log effect does not carry. The cleaner statement remains the world-share result: the treated products held their share of the world market while comparable defense products lost roughly 29 percent of theirs, a relative gain of about 15 percent, a few percentage points on the base share, which nets out global demand.

12.2 What the license cost

A dollar of exports is also not a dollar of welfare, and the two are different objects. The $6 billion 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 9.1. One assumption in that number belongs in the text. Each route is rescaled to a fully-licensed basis by the treated set's mean dose of 0.27, a linear-in-dose extrapolation the direct dose tests support only weakly; with the rescaling the routes

give 8 to 13.5 percent ad valorem, without it 2 to 3.5 percent, and the two figures bracket what the reform's own evidence supports, so everything downstream is carried as a range.[31]

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 on the order of $1.3 billion under the dose rescaling and roughly $0.3 billion without it. 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.

Those two components stand on different footing, and Section 9.1 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 rather than a global efficiency gain, and a calculation that stops at the U.S. border counts it as the latter. 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 saving, which is what survives the transfer argument above, trims the rescaled figure to about three cents, and carrying the elasticity band rather than its center stretches the upper end to seven. 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. We report the range rather than a point because its two ends rest on different assumptions, and the assumption separating them—that the estimated effect scales linearly in the licensing dose—is one the paper's own dose tests support only weakly. 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, and Section 11 is explicit that the trade data cannot see it. Stating the threshold at least names the missing term and puts a number on how large it would have to be to overturn the ledger.

[31] Appendix B.9 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 rather than 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`.

The threshold is also a specific decision object, and four of its assumptions deserve statement. It prices losses at their expected value, linear in trade dollars; end-use losses are threshold events, and a single diverted subsystem can impose costs far beyond its trade value, so a decision-maker with convex losses over tail events should read the figure as a lower bound on the required caution rather than a neutral midpoint. It compares flows inside the outcome window, while the losses from a transferred capability materialize with lags the satellite episode of Section 2 measures in decades, so the gain side is measured and the loss side right-censored. Its two sides do not fall on the same 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 rather than an alliance one. And the risk is not uniform across the base: 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.

Three further caveats attach. The calculation is partial equilibrium and covers the U.S. exporter side alone. The trade elasticity is measured on broad HS6 import data and is imprecise for the aerospace codes, which is why a band is carried rather than a point. And the inversion assumes the whole trade response runs through a cost wedge, which Section 9.3 shows is doubtful for the largest channel. Nothing here rests on a measured producer margin, which the public data do not reveal and the industry results cannot supply: they detect neither a price nor an output response at the aggregation available.

12.3 Scored against what the reform was promised to do

The economic case for the reform was argued in specific numbers on both sides (Section 2), the agencies themselves produced none—GAO found in April 2012 that they "did not provide an analysis supporting the expected benefits," and BIS reported it did not collect data on the size of applicant firms[32]—and none of the numbers has been checked against outcomes. Table 5 does that. The record divides in a way neither side of the original argument anticipated, and the two sides are wrong in opposite directions.

On the trade flow, the reform's advocates had the sign right and the scale wrong by an order of magnitude. The one quantified projection in circulation was industry-financed and about a different policy: a January 2010 Milken Institute study for the National Association of Manufacturers projected $64 billion of additional real GDP and 340,000 jobs by 2019 (DeVol et al., 2010), for a largely commercial decontrol that never happened—its subject is commercially available technology products, it exempts sensitive military technologies, and its mechanism is an assumed closing of half the U.S. share gap in five markets. It became the administration's answer all the same: asked for a jobs figure in April 2013, Commerce answered with the study's export number, "up to an additional $60 billion a year in U.S. exports," and characterized the economic gain as "ancillary."[33] We estimate about $6 billion a year in the categories the reform actually moved, and Section 12.1 treats even that as an upper bound on the decontrolled good itself. The comparison is not like-for-like and we do not offer it as a refutation of the underlying study. What it establishes is that the only export number the reform's sponsors gave Congress was ten times the gain the executed policy delivered, and was never a number about the executed policy.

[32] GAO-12-613, p. 25; House Foreign Affairs Committee, *Export Control Reform: The Agenda Ahead*, April 24, 2013, printed pp. 29 and 75.

[33] Same hearing, question for the record, printed p. 98. The restatement shifts the study's object from GDP to exports, its 2019 terminal-year gap to an annual flow, and its scope from dual-use technology products to the reform as a whole.

The critics were closer on the aggregate and wrong on the product, and the arithmetic that separates the two readings is worth stating because the current debate repeats it. Individually licensed goods were about three percent of U.S. exports, and on that basis the prospective gains were called marginal at the time. Measured against U.S. goods exports the gain estimated here is about four tenths of one percent, which vindicates that reading exactly. Measured against the affected products' own trade it is close to a fifth, and it reverses a decline their comparators did not escape. Both statements describe the same estimate. The ex-ante debate 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.

Where the record is furthest off, the sign of the error reverses and the source is official. State's analysis of the rule it was issuing put the saving at 35,000 fewer individual license applications a year, scored at one hour of burden each, plus 1,000 registrants relieved of registration. The BIS cost-benefit analysis for the license exception valued exporters' saving at $1.5 to $5.1 million a year on roughly 2,300 eligible transactions, and expected about 20,000 further transactions to become eligible once Munitions List items moved; carried to that fuller caseload at the agency's own per-transaction rate, the projection is $15 to $50 million a year. The trade response prices the same bundle at 8 to 13.5 percent ad valorem under the dose rescaling and 2 to 3.5 percent without it, an administrative saving of $0.3 to $1.3 billion a year, one to two orders of magnitude higher.

Both figures are right about what they measure, which is why the gap is informative and not merely an error. Agencies priced the form. The trade response prices the regime, which included the registration whether or not a firm ever shipped, the Part 124 agreements that have 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. Khandelwal et al. (2013) report the same asymmetry in a different apparatus, where the administrative machinery rather than any measured fee carried the distortion, and Section 10.2 says it again inside these data: relief arrived on transaction size, which is the margin an adjudicated authorization taxes, and not on the filing frequency the agencies counted. 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.

Two further predictions can be bounded rather than scored. The employment gain, 340,000 jobs in the study Commerce quoted, does not appear: employment in the exposed industries is flat, with a confidence interval that excludes a rise above about three percent (Section 11). The offshoring loss that organized labor predicted from the same policy does not appear either, on the same aggregates and with the same limited resolution. Neither null identifies a mechanism, and Section 11 explains why an industrial-base response operating in full would not register at that aggregation. What can be said is that the two loudest employment claims of the period pointed in opposite directions and that neither is inside the range these data support. One forecast in the record was directly about license counts, the 3,000 low-risk applications the exception was projected to eliminate (GAO-12-613, p. 26), and these data cannot check it; the exception carried 11.1 percent of first-rule 600-series value (Appendix B.14), which is a value share rather than an application count.

12.4 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

Table 5. Contemporaneous forecasts and the estimates

Forecast	Source	Estimate here	Reading
Up to $60bn a year of additional U.S. exports	Commerce, HFAC question for the record, April 2013, quoting DeVol et al. (2010)	About $6bn a year in the transferred categories	Right in sign, high by an order of magnitude; the figure described a wider decontrol that did not occur
Gains negligible: licensed goods are ∼3% of U.S. exports and the U.S. already holds most of the covered market	Hartung (2013)	0.4% of U.S. goods exports; ∼19% of the affected products' own exports	Right on the national aggregate, wrong on the product
+340,000 jobs by 2019, of which 160,000 in manufacturing	DeVol et al. (2010)	No detectable employment response; interval excludes a rise above ∼3%	Not supported at the available aggregation
Lighter Commerce controls move technology and production offshore	International Association of Machinists, March 2013, in Hartung (2013)	No detectable response in output, value added, employment or patenting	Not supported at the available aggregation
35,000 fewer license applications a year at one hour each; $1.5–5.1m a year saved on 2,300 eligible transactions	State, 78 FR 22740 at 22752; BIS, in GAO-12-613, p. 26	Ad-valorem equivalent 2–3.5% (8–13.5% dose-rescaled); $0.3–1.3bn a year	Low by one to two orders of magnitude; the burden was scored as clerical time[a]

Sources and their standing are set out in Section 2; the rule for inclusion is that the statement was made before the outcome it concerns. Estimates are those of Sections 7, 11 and 12.1–12.2. The national share in row two uses U.S. goods exports of about $1.5 trillion a year across the reform window; the product share uses the treated set's $32 billion a year of pre-reform exports.

[a] BIS's own per-transaction figure of $650 to $2,200 carried to the roughly 22,300 transactions it expected to become eligible gives $15 to $50 million a year, against the $0.3 to $1.3 billion the trade response implies.

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 (Sections 11, 9.3). 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—the question is whether the end-use risk decontrol added exceeded one to four cents per dollar of the affected trade—and that is a magnitude the case-level enforcement record could in principle speak to.

The ledger is also incomplete on the benefit side, because ECR's declared objectives were security objectives: concentrate scrutiny on the most sensitive items, and improve allied interoperability. The destination evidence is consistent with the second operating, though it is not a causal measure of it. The first, whether State's scrutiny of the retained items deepened as its caseload fell, is the reform's central security promise, and the one body that has looked directly did not find deepening: auditing the residual Munitions List caseload in 2019, the State Department's Inspector General found approvals granted on incomplete applications, required congressional certifications missed, and end-use checks run by untrained rotating officers.[34] That is a finding about the residual caseload,

[34] U.S. Department of State, Office of Inspector General, *Audit of the Department of State Direct Commercial Sales Program*, AUD-SI-19-07, February 2019: twenty of twenty-one sampled applications approved despite missing required information; seventeen licenses that should have been certified to Congress and were not; Blue Lantern checks run by

not a cost of decontrol; it means the reform's own security rationale should not be assumed to have operated. Evaluations that weigh the screening margins directly are less favorable still—the fullest outside review concluded the reform tilted toward trade at the expense of controls (Stricker and Albright, 2017), and the 600-series channel has been found to carry lighter anti-corruption vetting than the government-to-government alternative (Goodman, 2020)—and the diversion null does not answer them, since they concern screening intensity rather than trade flows.

Reading the estimates into the current decontrol turn. 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 roughly a sixth, 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 11 explains why the rationale working in full would not register at that aggregation. What the estimates discipline is the rationale's scale rather than its truth.

Three differences bound how far any of it carries. 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.[35]

13 Conclusion

We began by asking 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 is the point: what the estimates vindicate is a configuration, 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. The one decontrol that departed from that configuration on destination and end-user structure—the 2020 transfer of firearms, sold heavily through civilian resellers rather than government customers—generated the diversion and oversight risks GAO documented (GAO-25-106849); the staffing and coverage limitations GAO found at Commerce are generic to its end-use-check apparatus and apply

rotating officers without dedicated training, several delayed by more than a year. The audit covers the post-transfer Munitions List caseload rather than the transferred goods, so it speaks to whether retained scrutiny deepened, not to the security cost of the transfer itself.

[35]Licenses: Bloomberg, "Nvidia Gets US License for Small Amount of H200 Exports to China," February 26, 2026; CNBC, May 14, 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), with further reports through June 2026.

to 600-series goods as well. 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; none is sufficient, and the honest-DiD sensitivity makes the residual reliance on the institutional timing argument explicit. Named rather than left abstract, the leading candidate for such a shift is the allied fighter and rotary-wing delivery wave of 2014–2018 together with the post-2015 Gulf build-up, both concentrated in aircraft and gas-turbine engines. Dropping those two families outright leaves the export effect at +0.137 and the world-share effect higher than on the full sample, which bounds that particular story without excluding a version of it operating through the remaining categories. The control side has a counterpart candidate: the post-2012 procurement drawdown fell on the retained categories through pipelines served largely by U.S. incumbents, and a decline of that origin cannot be separated from the cost of continuing control with these data (Appendix B.4), which is why the estimates are stated as the decontrol-versus-continued-control policy wedge rather than as an effect against a stable counterfactual.

Under that assumption, the 600-series transfers raised affected products' exports by 15–18 log points relative to not-yet-treated defense products and held their world-export share while comparable defense products' shares fell roughly 29 percent, a relative gain of about 15 percent concentrated in allied and defense-customer destinations and registered against the market position of allied industrial suppliers—though most of the realized *dollar* increase ran through Foreign Military Sales, where licensing is not the proximate mechanism (Section 9.3). The exposed industries' output, employment, and patenting show no detectable response, bounds set by data resolution rather than evidence of absence, and no diversion appears in bilateral trade, on a test the firearms positive control shows is underpowered on the value margin.

Set against what the reform was promised to deliver, those estimates convict both sides of the original argument, in opposite directions (Section 12.3). 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 rather than more frequent ones, the friction removed can also be priced: an ad-valorem equivalent of 8 to 13.5 percent for the ITAR wedge on a fully-licensed basis, 2 to 3.5 percent without that rescaling, 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, its position in that range set by a dose-rescaling assumption the paper's own tests support only weakly. Whether it did is a question about enforcement records rather than about trade flows, and it is the one this design hands on.

We close with the agendas that would sharpen what these data leave open. The domestic nulls are bounds set by data resolution: the decontrolled goods are too small a share of any public aggregate for even the industrial-base rationale working in full to register, so isolating them requires microdata fine enough to keep them as their own observation. The Census Longitudinal Firm Trade

Transactions Database would let the washout 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. Hardening the exogeneity case would take the internal interagency record, to tie the queue order to contemporaneous decisions rather than to the published criterion, and a pre-sequestration base period, to separate the one suggestive domestic signal, capital expenditure, from the defense-budget cycle.

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.

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'Haultfœuille. 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, C. 2020. *Holes in the Net: U.S. Arms Export Control Gaps in Combatting Corruption.* London: Transparency International Defence & Security, January.

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. Trefler. 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.

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: 101312.

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 6. Principal datasets

Dataset	Source	Level and coverage	Role	Main limitation
Census ECCN export panel	U.S. Census AES (restricted-access filing records)	HS6 × destination × 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, reconciling 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 crosswalk reallocates mass; no military technology class for aircraft or defense electronics
SIPRI Arms Transfers Database	SIPRI	Weapon-category × 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 7), dropping the entire `9x610` aircraft family—half the dose—leaves the volume ATT essentially unchanged (+0.170 → +0.176) and *raises* the share ATT (+0.147 → +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.[36] 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 8 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 8.4. 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 and should be

[36] Script `106`. Estimated on the same 88-product regime-gated set as Table 7, whose full-sample values are +0.170 and +0.147; the two Chapter-98 special-provision codes are excluded from estimation throughout and are added back only in the artifact check of Table 12, which reports the nominal 90-code +0.169.

Table 7. Leave-one-out pooled ATT, regime-gated treatment

Dropped	n 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.

Table 8. Regime-gate threshold sensitivity

ITAR_{ind} gate	n treated	ln exports ATT (t)	max|pre|	world share ATT (t)	max|pre|
2%	160	+0.156 (4.6)	.027	+0.119 (2.9)	.037
5%	88	+0.170 (3.8)	.028	+0.147 (3.1)	.049
10%	62	+0.216 (4.0)	.023	+0.145 (3.0)	.074
25%	38	+0.191 (3.1)	.061	+0.116 (2.1)	.120
50%	17	+0.256 (2.5)	.136	+0.189 (2.8)	.032

Pooled ATT (post) and largest absolute pre-period coefficient at each gate. Treated set = ever-600 $\cap$ pre-2013 ITAR-individual share $\geq$ gate, cohorts 2013–2017. Unlike the other tables, parentheses in the ATT columns report t-statistics rather than standard errors. 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 ATT by less than 2 percent (+0.168 on exports, +0.138 on share). Script `73`.

read with that in mind. A civilian co-movement large enough to fully account for the +0.256 estimate at this threshold would require civilian exports of the treated products to grow more than 40 log points above those of the controls, a gap the supplier-differenced bound of Section 8.3 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 9.1.

The share pre-trend 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.

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-

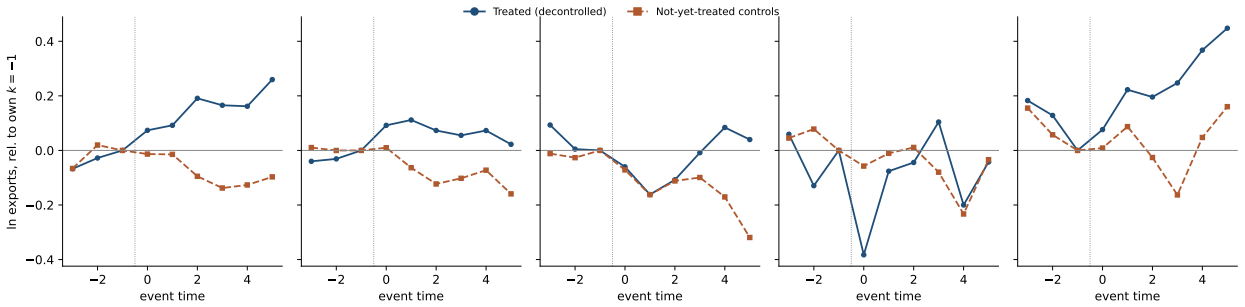

Figure 6. Staggered divergence, by transfer cohort (discussed in Section 8). Each panel plots ln exports relative to each product's own value the year before its cohort's transfer ($k = -1$), averaged across that cohort's treated products and across the defense controls not yet treated within its window. 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.

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 asinh(share × 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 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

Table 9. 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 ($p = 0.067$).

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 9): 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 11; 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 × 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 8.3 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 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—these data do not adjudicate: what separates them is whose decision drove the control decline, and no test here observes that. This is why Section 7 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, printed rather than asserted. The raw 2009–2012 balance comparison behind Section 8 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 8.2 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 11, 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 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 rather than the 2013–2017 600-series, and Category IV munitions of war, which stayed on the Munitions List.[37] 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 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.6 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).[38] All three agree (Table 10): 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.

[37] The 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.

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

Table 10. 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). Panel 2009–2022 over the defense universe. All six coefficients are significant at the 1 percent level.

B.7 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 11). 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 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.

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.8 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

Table 11. 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 ($y \sim$ post $|$ uid + 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.

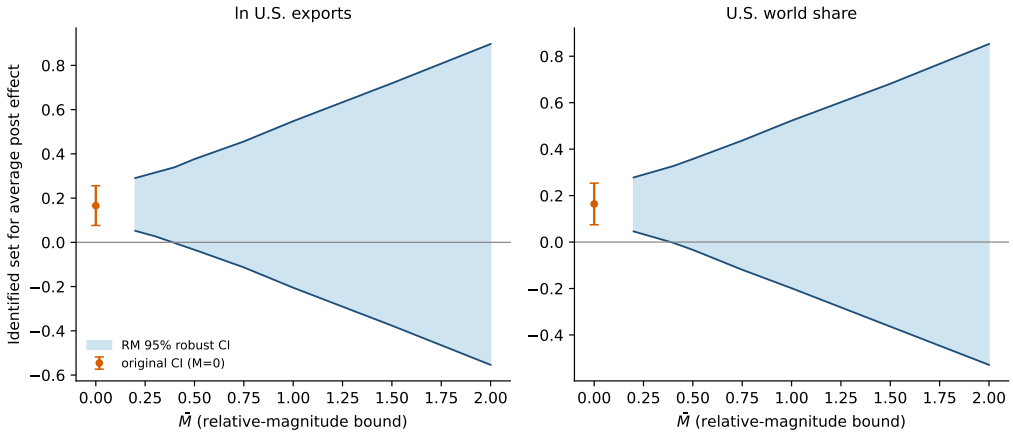

Figure 7. Honest-DiD relative-magnitude sensitivity (Rambachan and Roth, 2023), regime-gated treatment. Shaded band: the 95% robust confidence set for the average post-period effect as the allowed post-period trend violation $\bar{M}$, measured relative to the largest pre-period deviation, grows; the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

curvature reaches roughly a tenth of the curvature already present in the pre-periods (a smoothness 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.9 Bindingness gradients: an ad-valorem benchmark and per-shipment friction

The cross-sectional dose-response of Section 10.1 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).[39] At that measured elasticity the 0.48 response rationalizes an ad-valorem equivalent for the licensing 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 8.3 gives 13.5 percent and the bilateral panel of Section 9.1 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 unsignable in the licensed channels)—are stated where the numbers are used, in Section 12.2, 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,

[39] Estimated 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.

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 10.2) 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.10 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 10.1, 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.11 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.[40] The panel does use as-reported HS vintages, so the

[40]The 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.

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 $408 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 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.12 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.12). 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 12). The trade response is growth on an existing base, not the appearance of a code.

B.12 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

Table 12. The headline is not a coding or reporting artifact

Treated set	n 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. "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.

is not. The U.S. filing records are nonetheless the source closer to the decontrolled good: within the treated codes they are far more concentrated on the military variant—aircraft and engines are 30 percent of their treated value against 75 percent in reconciled trade, whose aircraft series is largely the civilian catch-all documented in Appendix B.11—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 13). 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 13, 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 of the pre-trend is the one Section 9.1 sets out for its own panel: a deep-lead coefficient estimated on a thinner set of cells than the reference period generates an apparent pre-trend. Requiring each importer-by-product cell to report a positive flow in all three pre-treatment years of its own sub-experiment, a restriction on pre-treatment presence that selects on nothing realized after transfer, 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, and excluding the Gulf states alone barely moves it.

The two sources agree. On the cells they share the reconciled panel gives +0.400 and the filing records +0.468, statistically indistinguishable, and dropping aircraft and engines leaves the reconciled coefficient at +0.386 and the filing-record coefficient at +0.319, so the pattern is not a property of the civil-aviation classification difference between the series.[41] The full-sample estimates differ—+0.726 on the filing records against +0.389 reconciled—not because the two disagree about the trade they both record but because of which cells each contains: reconciliation drops small non-reciprocated flows and reallocates the civil-aviation catch-all, while the filing records retain the U.S. export declarations as filed. Because the two agree on the cells they share, the larger filing-record coefficient is a property of the additional U.S.-declared cells it keeps rather than of the values it reports for trade both sources see, which is what a mechanical filing-side explanation would require. 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 corroboration on the source closer to the decontrolled good.

What neither source does is resolve 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 Appendix B.14—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 9.1, the destination evidence supports reading the gain as the decontrolled military good rather than civilian co-movement.

B.13 The second stage in detail

This appendix reports the outcomes summarized in Section 11.

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),

[41] Scripts `108`, `109`. The shared cells are about two-thirds of reconciled cells and 94 percent of filing-record cells; the drop-aircraft estimates are on the all-cells sample.

Table 13. Differential gain to allied and arms-customer destinations, by data source

	Estimate	(SE)	Wild p	max|pre|	Treated HS6
Panel A. Reconciled bilateral trade					
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
Panel B. U.S. filing records, same specification					
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.

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 found[42]—and the industrial-base rationale is untestable at this aggregation rather than rejected.

[42]Export-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 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).

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. Two limits bound this result: 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 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).

End-use monitoring after transfer. The diversion test above compares what two trading partners report; it is silent on end use, whether a delivered good reached an authorized user. The public monitoring series speak to that margin, and they do not support a simple story in either direction. The formal authority barely moved: of the ten compliance and enforcement activities enumerated in Section 3, end-use monitoring, selection for post-shipment verification, border inspection, seizure, investigation and penalties all apply whether or not a shipment needs an individual license. The volumes moved a good deal. State's Blue Lantern checks initiated fell from 820 in FY2012 to 570 in FY2015 and 281 in FY2021 as its licensing caseload shrank, while its unfavorable rate rose from about a fifth to about three tenths over the same span; the check rate per remaining license stayed near one percent. Commerce, which received the transferred goods, completed on the order of 700 to 1,200 checks a year and screened a higher share of its applications than State did of its, about 3.3 percent against 1.3 percent over FY2013–2017, with roughly nine in ten of its checks conducted after shipment.[43] The receiving regime was therefore not the less active of the two. It was differently weighted, toward a check type that establishes location and end use but has been documented as largely not confirming compliance with the conditions a

[43] Directorate of Defense Trade Controls, Blue Lantern End-Use Monitoring reports to Congress; Bureau of Industry and Security end-use checks completed by fiscal year (708 in 2010 rising to 1,171 in 2019); GAO-19-307 for the FY2013–2017 rate comparison and the post-shipment share; GAO-04-357 for the finding that post-shipment verifications largely do not confirm compliance with license conditions; GAO-23-106379 for the FY2013–2021 Blue Lantern totals (4,361 checks, 85 referred as unfavorable) and the FY2021 detail (281 checks initiated against 19,125 authorizations, 256 closed, 175 favorable). GAO withheld precisely the step that converts monitoring volume into realized risk: whether unfavorable determinations implied statutory violations, and the law-enforcement referral count. Two traps in these series: checks initiated and checks closed within a fiscal year are different populations and must not be differenced, and DDTC changed its tabulation method from FY2019, so figures on either side of that break sit on different bases.

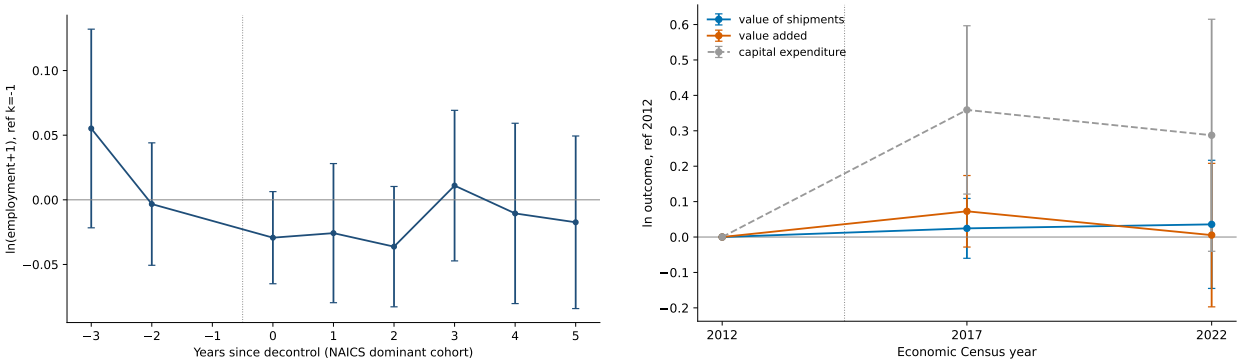

Figure 8. Industry second stage, regime-gated treatment mapped to NAICS6. Left: employment (Census County Business Patterns), flat after transfer. Right: value of shipments and value added (flat) and capital expenditure (rises, but not identified; see text), from the Economic Census. The left panel is 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.

license carried, so continuity in the count of checks is compatible with a change in what a check establishes. What neither series can do is attribute checks or unfavorable findings to the transferred categories, a consequence of the same absence of a common item categorization across licensing, export and enforcement systems that prevents any public series from following these goods across the jurisdictional boundary. Whether end-use risk rose for these specific goods is not identified here. The trade-discrepancy null bounds visible re-routing; a direct test of the security cost would need the case-level end-use record itself.

B.14 An administrative cross-check on the treated-product gain

The Commerce Department's Bureau of Industry and Security publishes export and licensing values for the 600-series items themselves, compiled from the Automated Export System and reported separately for each USML-to-CCL rule, and the rules correspond to this paper's cohorts.[44] This series is not an identification strategy. It has no pre-period, because before the transfer these goods shipped under State Department authority and generated no Commerce record, and it is a level rather than a difference. It is instead a public cross-check on the treated group, compiled from the same Automated Export System as the Census panel but assembled by the government from the control-field ECCN rather than from the license-code gate used here; the two are consistent but not independent.

It corroborates the estimates on three points. First, the scale of the treated-product flow. Commerce-jurisdiction exports of the first-rule items, aircraft and gas-turbine engines, rose from $88 million in the last quarter of 2013 to about $4.2 billion a year by 2019 (Table 14), in a series this paper did not construct. Part of that ramp records the licensing handoff rather than trade growth: as grandfathered State licenses expired and contracts cycled onto Commerce authorizations, recorded Commerce-jurisdiction value had to rise from zero whatever the trade path, and the State-side

[44] Bureau of Industry and Security, Office of Technology Evaluation, "Statistics on U.S. Exports and BIS Licensing under the USML-to-CCL Regulatory Changes," from the Commerce U.S. Exports Exporter Support System and the Automated Export System; the series and the rule-by-rule reports are at `bis.gov/about-bis/bis-leadership-and-offices/OTE/usml-ccl-regulatory-changes`. Figures cited here are from the first-rule (aircraft and gas-turbine engines) report through August 2020 and the 600-series licensing summary through December 2022.

Table 14. Commerce-jurisdiction exports of the first-rule 600-series items, an independent administrative series

Period	U.S. exports under Commerce jurisdiction ($ million)
2013 Q4	88.1
2014	2,030.0
2015	3,495.2
2016	4,052.5
2017	4,150.8
2018	3,980.9
2019	4,263.8
2020 (first half)	1,996.3
Cumulative, Oct. 2013–Aug. 2020	24,792.9

Actual U.S. exports of the first-rule 600-series items (aircraft ECCN family `9x610` and gas-turbine engines `9x619`) under Commerce (Export Administration Regulations) jurisdiction, from the Automated Export System as compiled by the BIS Office of Technology Evaluation. 2013 is the fourth quarter only, as the rule took effect October 15, 2013; 2020 is the first half. The series has no pre-transfer values, because these goods shipped under State Department authority before the transfer; it corroborates the level and post-transfer build of the treated-product gain, not its identification. Annual values are digitized from the report's chart; the published cumulative total ($24,792.9 million) matches exactly. Source: BIS/OTE first-rule statistics through August 2020 (see footnote in text).

mirror shows the same migration: in the State Department's Section 655 reports, Munitions-List authorizations for the transferred categories fell far faster than for the categories that stayed on the list (Table 15). The corroboration is therefore the level the series settles at, the scale on which the bounds of Section 12 are built, not the ramp itself.

The BIS series also closes the measurement loop for the 2013 cohort in a way the DiD alone cannot. As Section 8.4 notes, the pre-reform measurement of the treated good is the HS6's ITAR individual-license value (S-code exports), and the DiD outcome is the HS6 total, which includes co-moving civilian trade. The BIS/OTE `9x610`/`9x619` series provides the post-reform, military-specific counterpart—AES exports under Commerce jurisdiction for those ECCNs are, by construction, the decontrolled military good. Their settlement at about $4.2 billion a year by 2019 is consistent with the Census-based DiD estimate, and is drawn from the same Automated Export System but assembled from the control-field ECCN rather than from the license-code gate. The two series agree on the level, confirming the DiD captures the military good rather than co-moving civilian trade in the aircraft HS6 codes.

Second, the gain is allied and defense-customer concentrated. The top fifteen destinations, all allies or established arms customers and led by Japan, Canada, the United Kingdom, and South Korea, take 77 percent of first-rule export value, matching the destination result of Section 9.1. Third, the licensing channel shows up in the composition. Eleven percent of first-rule export value moved under the Strategic Trade Authorization exception, the allied license exception the framework of Section 3 identifies as the mechanism, and BIS approved 90 percent of 600-series applications by count and 92.5 percent by value, with average monthly approvals across all items 77 percent above their pre-reform benchmark, in large part the licensing workload that migrated from State to Commerce with the transfers. None of this identifies the effect, which the difference-in-differences does; it confirms that the object the design isolates exists at the scale, in the destinations, and under the licensing composition the design says it does.

The State-side mirror. The Commerce series has no pre-period; its State-side counterpart does. The State Department's Section 655 reports to Congress give the annual value of Direct Commercial Sales authorizations under the Munitions List by USML category, so the transferred categories can be followed as they leave State's jurisdiction.[45] Authorizations under the transferred categories fell far more than under those that did not transfer. Aircraft and the gas-turbine engines the reform split into the new Category XIX—followed jointly, since before FY2014 the engines sat inside Category VIII—dropped from a pre-transfer average of $15.2 billion a year (FY2010–2012) to $6.1 billion over FY2015–2017, military electronics (Category XI) from $10.8 billion to $3.0 billion, and the main 600-series tranches as a group by about three-fifths, against a decline of 6 percent for firearms and ammunition, whose own transfer did not take effect until March 2020 and which are the not-yet-treated controls of the trade design (Table 15).

The comparison group here needs care, and it is not the one a reader would first reach for. Fifteen of the twenty-one Munitions List categories were subject to a Commerce transfer rule effective inside this window, so "categories that stayed on the list" is a much smaller set than it sounds: only Categories XVI, XVII and XXI had no transfer rule at all, and their combined authorizations are two orders of magnitude smaller than the transferred group's and swing by orders of magnitude from year to year. The firearms categories are the defensible comparison because they are large, they are stable, and their exclusion from the window is the same institutional fact the trade design already relies on. The contrast is not sensitive to how the pre-period is measured: against an FY2011-only base the main tranches fall 69 percent and firearms 24 percent. Because Section 655 records authorizations rather than shipments, and because authorizations drift down across the board, neither the level nor the common decline is itself a decontrol effect; the differential decline of the transferred categories is the object that corroborates the design.

Reproducibility. All results come from two public sources—the U.S. Census ECCN export panel and BACI (CEPII's reconciliation of UN Comtrade)—and are produced by numbered scripts in the `ECCN_IV` project. The regime-gated treatment and the headline stacked estimates (Table 1) 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` (`findings_ecr_exante_schedule_2026-06-30.md`). A script-by-script map from each table and figure to the code and findings memo that produces it is given in the project's replication README (`REPLICATION.md`).

[45] U.S. Department of State, Directorate of Defense Trade Controls, Section 655 Annual Military Assistance Reports, FY2010–FY2019. These count authorizations, not shipments, and cover direct commercial sales only, excluding Foreign Military Sales (reported separately by the Defense Security Cooperation Agency). Fiscal years 2010 and 2012 are published only as scanned images; their category totals are recovered by OCR, corrected against the printed category ordering, and are consistent with the machine-readable FY2011 and FY2013 figures.

Table 15. State-side Direct Commercial Sales authorizations, by transfer status within the window

USML category	Pre, FY2010–12 ($B)	Post, FY2015–17 ($B)	Change
Transferred to the Commerce list, main 600-series tranches			
VIII+XIX Aircraft and engines	15.2	6.1	−60%
XI Military electronics	10.8	3.0	−72%
XII Fire control/optics	2.5	1.2	−50%
VI Surface vessels	0.4	0.2	−45%
VII Ground vehicles	1.9	1.2	−38%
Subtotal[a]	30.8	11.7	−62%
Transferred within the window, remaining categories			
Subtotal[b]	4.8	2.9	−39%
Not transferred within the window			
I–III Firearms and ammunition[c]	1.6	1.5	−6%
XVI, XVII, XXI[d]	0.20	0.03	−86%

Fiscal-year average of Direct Commercial Sales authorization value under the U.S. Munitions List, from the State Department Section 655 reports. Pre-period FY2010–2012 is fully before the first transfer (aircraft, effective October 2013); post-period FY2015–2017 follows the main tranches. FY2011 and FY2013+ are machine-readable; FY2010 and FY2012 are OCR-recovered (sequence-corrected). Authorizations, not shipments; excludes Foreign Military Sales. The four blocks partition all twenty-one Munitions List categories; transfer status is assigned from the effective date of each Commerce transfer rule.

[a] Categories VI, VII, VIII, XI, XII, XIX, and XX. Aircraft (Category VIII) and gas-turbine engines are shown jointly as VIII+XIX because the reform split the engines into a new Category XIX that appears on the State side only from FY2014; before then the engines sat inside Category VIII, so the two share a single pre-transfer figure. Category XX (submersibles) is negligible and included here. The displayed rows sum to the subtotal.

[b] Categories XIII (effective January 6, 2014), IV, V, IX and X (July 1, 2014), XIV and XVIII (December 31, 2016), and XV, which moved to the 9x515 series rather than the 600 series (June 27, 2014). These carry no treated products in the trade design but are not retained categories either.

[c] Transfer effective March 9, 2020, after the window; these are the not-yet-treated controls of the trade design. On an FY2011-only pre-period base the decline is 24 percent, against 69 percent for the main tranches.

[d] The only categories with no transfer rule. Their combined authorizations are two orders of magnitude smaller than the transferred group's and move by orders of magnitude between adjacent years, so they cannot carry the comparison.