

Contract-Production Reliance and the Output Cost of Supply Disruption*

Philip A. Luck[†]

First draft: March 2026
This draft: October 2026

Abstract

Resilience policy treats the vulnerability of production as a question of where inputs come from. We measure a margin this framing misses: contract-production reliance, payments to outside contractors for work on materials the buyer owns, per dollar shipped. Across U.S. manufacturing industries it is nearly orthogonal to import penetration. After COVID-19 hit, industries one standard deviation more reliant lost about 2.4 percent more output over 2020–2023. With controls for their customers' own shocks and for steel tariffs, the 2020 estimate is 1.35 percent. The loss was in physical volume, not prices or margins, and it began as the economy reopened. Employment was still lower in 2024, and the most reliant industries cut contract work by nearly a quarter. European and Japanese panels show the same acute loss; in Europe, as in the United States, it comes mostly from differences between broad sectors. Under less synchronized disruptions we find no comparable loss, though most of those tests have little power.

JEL Classification: L23, L60, F14, D24, E32.

Keywords: contract manufacturing, firm sourcing, supply-chain resilience, production networks, productivity, COVID-19.

*I thank seminar participants at CSIS for comments. This draft is preliminary; comments are welcome, and all errors are my own.

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

1 Introduction

Resilience policy treats the vulnerability of production as a question of geography. Reshoring has become the standard prescription for supply-chain fragility, and its most explicit instrument is the import restriction.¹ The exposure measures that inform it, import penetration, foreign-input content, and dependence on a single source country, all ask where an industry’s inputs come from. Production also has an organizational structure. When a manufacturer needs a production step performed on materials it owns, it can carry the step out in house or pay an outside contractor to do it. That choice is silent about location, and the resilience debate does not measure it. Yet on the shock that motivated the reshoring prescription, this sourcing margin moved U.S. manufacturing output about as much per standard deviation as the geographic one did. The Census records the contracting choice as the cost of contract work—payments made to other establishments for work performed on materials the manufacturer owns. Scaled by shipments, that cost is what we call an industry’s *contract-production reliance* (CPR). It is a specific sourcing margin, distinct both from buying finished inputs and from owning a supplier.

The goal of this paper is to measure that organizational margin and to ask what it cost when the 2020 shock hit. We make two arguments, and they should be kept distinct. One does not depend on the kind of shock: the organizational margin that carries much of the vulnerability is predominantly domestic and nearly orthogonal to import penetration, so a lever aimed at geography, the import restriction that dominates resilience policy, reaches it only incidentally.² The other is specific to the disruption we can identify. That margin carried a measurable output cost when a broad, synchronized shock hit. For the targeted disruptions most reshoring is aimed at, the diversification logic of Section 7.2 runs the other way, and we have no evidence that a vulnerability of the kind found here arises. To test these arguments, we build a panel of 358 six-digit NAICS manufacturing industries observed annually from 2005 to 2024 (188 of them before 2012). It joins contract-work and real-activity data from three Census instruments, the Annual Survey of Manufactures, the 2022 Economic Census, and the 2023 and 2024 Annual Integrated Economic Survey, to NBER-CES productivity measures, the NSF Business Enterprise R&D survey, and an HS10-to-NAICS6 trade concordance that yields a time-varying measure of industry-level foreign-input exposure.

That a cost should appear at all turns on an economic asymmetry between normal and disrupted times. Routing a production step to an outside contractor is an ordinary, cost-saving transaction that adds capacity and flexibility when conditions are calm. When a broad, synchronized shock hits, an industry that produces in house can re-sequence and reprioritize lines within its own walls; an industry that contracts out must rely on outside suppliers doing the same under arrangements that cannot specify every contingency. The COVID-19 shock of 2020 provides a precisely dated test of whether this asymmetry bites.

We set the level of analysis deliberately, and it bounds everything that follows. An industry panel identifies the output response of the contract-reliant margin. It cannot cleanly separate the organizational frictions of the contracting relationship from the baseline fragility of the commodity production typically routed to contractors, because the industries that contract out cluster at the standardized, low-R&D end of manufacturing. Separating them requires within-industry variation

¹Most reshoring *spending* runs through subsidy and procurement rather than trade policy, and our design speaks to neither; Section 8 returns to what that leaves open. The import restriction is the instrument on which the resilience case has been argued most directly, and it is the one whose reach we can assess.

²By resilience, we mean the ability to maintain output through an acute disruption; we do not measure recovery-time or ex-ante-investment dimensions emphasized in the theory of resilience as an investment good (Grossman et al., 2024).

across establishments: a Census Research Data Center design that compares plants with different contract-work shares inside the same six-digit industry and year. We recognize that this is a strong limit, and we do not attempt the establishment-level design here. What the industry data can do is bound how far the response survives inside sectors and families, which Section 4.4 reports; the sign and the timing do not depend on the separation. The estimand is bounded in a second way. Least squares concentrates its weight on a few reliant industries (a Kish effective sample of 31 of 358), so the estimate describes a substantially reliant industry, not the average manufacturing dollar, and the acute response does not carry over to dollar weights.

Section 7.2 writes that asymmetry down. Contracting a production step out diversifies idiosyncratic supplier risk but loads on a common supply factor, so the effect of reliance on output flips sign with the correlation of supplier shocks: positive when disruptions are weakly correlated, negative when a common factor hits all contractors at once. An acute, synchronized collapse like COVID-19 should therefore bite hard, while diffuse supply-chain pressure or the demand-driven 2008–09 recession should not. The framework is an interpretive aid: its correlation parameter is not measured, so it orders the shocks we observe, and no result depends on it.

Four results follow. The first carries the shock-independent argument; the second, third, and fourth carry the shock-specific one. First, the margin is predominantly domestic and structurally separate from the import channel. Across industries CPR is nearly orthogonal to import penetration, accounting for one percent of the cross-sectional variance and virtually none of the within-industry variation over time. The only directly measured component of offshore contracting (the customs provision recording offshore processing of U.S.-owned goods, HTS 9802.00) runs 5–9 percent of total contract-work spending. Because the filing is optional and duty-driven, that share is a floor; the partner-mirror and receipts checks point the same way. Did foreign sourcing displace the margin? A shift-share design built on foreign suppliers’ export capacity to other high-income countries finds no evidence that it did. Reshoring policy that targets imports therefore does little about the vulnerability domestic contract manufacturing carries.

Second, contract-production reliance carried a real output cost when the 2020 shock hit. Taking pre-pandemic reliance as a pre-determined treatment, industries one standard deviation more reliant (1.6 cents of contract work per dollar shipped, against a 0.7-cent median) lost 2.4 percent more output over 2020–2023 and 2.2 percent in 2020 alone, and the more reliant half lost 6.4 percent more.³ The comparison behind this number is between industries such as forgings, powder-metallurgy parts, and foundries, where contract work runs above the 2.8-cent ninetieth percentile, and industries near the median. Composition controls move it to 2.8 percent. The main threat to that reading is a customer shock. Many of the industries carrying the estimate are metalworkers whose customers in oil and gas, aerospace, and autos took their own shocks in 2020, and whose input costs Section 232 steel tariffs had raised. Controlling for both leaves 1.35 to 1.6 percent in 2020 and 1.7 to 2.1 percent over 2020–2023, depending on how customer exposure is measured, and we read the lowest figure as a bound. The loss runs mostly across sectors. Inside three-digit sectors the 2020 estimate is under 1 percent and not significant. Over 2018–2023 it is 1.8 percent; it is borderline with a six-digit measure of customer exposure ($p = 0.054$) and insignificant with the coarser commodity-level one, and it falls further inside four-digit families (Section 4.4).

The contraction ran through physical volume. Profit margins and per-worker productivity did not move, and relative prices show no response through 2023 and a marginal one in 2024 that

³Primary inference clusters at the six-digit level at which reliance varies ($p = 0.002$; the above-median contrast has $p < 0.001$); a conservative wild bootstrap on the 21 three-digit sectors gives $p = 0.022$. The ladder, which Section 4.5 reports in full, includes the two procedures that treat the 21 sectors themselves as the sample, which do not reject.

the real and nominal series cannot separate. At monthly frequency the gap is absent during the lockdowns and opens as the economy reopens. A foreign-input channel of comparable size is identified alongside it in the same regression and arrives two years later, in 2022–24. The two channels are separable in time as well as across industries, and import-focused policies act on only one of them. Benchmarked on the same shock, reliance moves output about as much per standard deviation as import penetration, although its rank among the exposures depends on which are included (Section 4.6). Six other statistical systems collect a version of the same margin, on different units of observation and against different denominators, and the same design yields negative acute-2020 estimates in all of them.

Third, the gap persists, and where it persists is itself informative. Five years after the 2019 base year the gap is still open in employment, and it is open through 2023 in establishment counts. In real output at six-digit resolution it persists only under exact parallel trends, the nominal gap narrows after 2022, and in the value-weighted between-family real aggregate the gap closes. Since a production index is exactly that aggregate, this reconciles our persistence with a literature that reports acute-then-recovery; the two measure different objects. The fall in establishment counts runs, on the most likely reading, through depressed entry, so the lost capacity is not replaced. Reliant industries also moved away from the arrangement itself. On a balanced set of industries, reliance fell from 1.01 to 0.75 cents per dollar shipped between 2019 and 2024, about one-quarter, almost entirely within industry, and the most reliant third cut contract work by nearly a quarter while its shipments rose. The pull-back was not larger where the 2020 loss was larger (Section 5.4). That pattern is consistent with the account of Acemoglu and Tahbaz-Salehi (2025), in which a severed relationship-specific link is not re-formed because the private return to re-forming it falls short of its social return, and harder to square with re-optimization toward safer networks (Kopytov et al., 2024); it tests neither.

Fourth, the cost is not visible under the other disruptions the panel contains, although most of those tests are underpowered. In normal times, contract work is an ordinary, minor input that imposes no steady-state performance drag. It accounts for only about 1.6 percent of total manufacturing costs, and a control-function production estimate puts its output elasticity near 0.05. Moreover, the cross-sectional relationship between contract reliance and R&D intensity is null, which weighs against a capability-atrophy or “hollowing out” explanation. The output cost revealed by the pandemic therefore appears contingent on an acute, synchronized supply-chain collapse, or on its size, which the panel cannot separate. The design detects no comparable loss in calm times, under diffuse supply-chain pressure, or during the demand-driven 2008–09 recession. The minimum detectable effects under those disruptions are of the order of the COVID estimate or larger, however, so the contingency is a reading these nulls permit without establishing it. Two targeted shocks are more informative. After the 2011 Tōhoku earthquake the coefficient is +0.016, with an interval that excludes the COVID estimate, and the 2025 tariff wave adds no penalty to reliant industries’ employment (−0.002; minimum detectable difference 0.007).

Related literature. The paper’s measure is narrower than the firm boundary studied in the integration literature (Antràs and Helpman, 2004; Grossman and Helpman, 2002; Antràs and Chor, 2013; Alfaro et al., 2019; Atalay et al., 2014; Goldschmidt and Schmieder, 2017): it records tolling of the manufacturer’s own materials, not the make-or-buy decision, and a low value is consistent with either integration or arm’s-length purchase of finished inputs (Section 2). We read this margin against the industry-equilibrium sorting of Grossman and Helpman (2002) and the property-rights and firm-level integration decisions of Antràs and Chor (2013) and Alfaro et al. (2019). Atalay et al. (2014) caution that vertical integration is about control, not physical input flow. The contract-work

item is therefore a sourcing margin. It does not measure the vertical boundary.

Mechanically, the paper draws on production-network propagation, where disruptions travel through inputs that are hard to substitute (Barrot and Sauvagnat, 2016; Carvalho et al., 2021; Boehm et al., 2019), and on two accounts of why cost-minimizing sourcing can be fragile. The first, due to Elliott et al. (2022), is technological: a chain of essential inputs fails discontinuously once reliability falls below a threshold, and firms that choose their own reliability settle at exactly that threshold—an equilibrium that is fragile even though fragility is always inefficient. The second, due to Acemoglu and Tahbaz-Salehi (2025), is organizational. Supplier–customer links carry relationship-specific surplus, rents the two parties bargain over, and a link is severed once a shock pushes its private value below the cost of maintaining it. Because neither party internalizes the surplus the link creates for third parties, the equilibrium network is a strict subnetwork of the efficient one. In other words, equilibrium output can fall discontinuously where the planner’s would not. The model does not compare a step performed in house with the same step routed through an outside processor; what it supplies is the asymmetry between a link that can be severed and a step that has no such margin, together with the reason a severed link is not re-formed. That is the part this paper takes to data. Kopytov et al. (2024) offer a competing account of the same post-shock reorganization, in which firms facing supply-chain uncertainty choose less productive but safer networks, so that a smaller network after a disruption reflects re-optimization rather than an externality. Likewise, Grossman et al. (2024) treat supply-chain resilience as an under-provided investment, which is the reading of the policy question we return to in Section 8. The redundancy-versus-flexibility and correlated-supplier-risk trade-offs at its center are long established in operations management (Sheffi, 2005; Tomlin, 2006; Ivanov and Dolgui, 2020), and recent surveys of supply-chain risk frame the question this paper takes to data (Baldwin and Freeman, 2022; Carvalho and Tahbaz-Salehi, 2019).

A quantitative general-equilibrium literature now places aggregate risk inside models of sequential production. In Castro-Vincenzi et al. (2026), firms commit sourcing links before shocks land, so risk flattens revealed comparative advantage and a disruption costs more the further downstream it falls. That work is a calibrated multi-country model whose object is the allocation of sourcing under risk, and it orders disruptions by position in the chain; ours is a measured organizational margin in one country, identified off a dated shock, and it orders disruptions by synchronization.⁴

Closest to this paper’s policy question, Grossman et al. (2023) ask whether resilience policy should promote international diversification or reshoring. The results here are the empirical counterpart to that question: they place the domestic vulnerability on a margin the import lever does not reach.⁵ The framework in Section 7.2 adapts the relocation-option logic of Bergin et al. (2009) (henceforth BFH), in which outsourcing is valuable because assembly can be shifted across locations in response to shocks, and augments it with a synchronized shock that disables that option. In structure it is a portfolio problem in which diversification stops short of the aggregate: substitutable contractors spread plant-specific risk but not the component their shocks hold in common. That is the same reason idiosyncratic shocks fail to wash out in production networks (Acemoglu et al., 2012). Recent evidence on Mexican manufacturing finds that industries more intensive in cross-border processing bore higher employment volatility (Walke and Weiler, 2026), work that sets the domestic contracting margin aside.

The shock evidence joins a growing literature on supply chains and the pandemic. Lafrogne-Joussier et al. (2023) show that French firms exposed to China’s early lockdown lost sales, and

⁴The two exercises are complements. A model of where risk falls in a chain does not say which domestic sourcing arrangements broke in 2020, and an industry panel cannot price the general-equilibrium reallocation that follows.

⁵The broader reallocation debate is surveyed by Alfaro and Chor (2023).

Meier and Pinto (2024) document the same channel for U.S. industries, with production falling most where Chinese-input exposure was highest; that import-exposure channel is the foreign side we estimate jointly with the domestic one. What such disruptions imply in the aggregate is worked out in Guerrieri et al. (2022), where a sectoral supply contraction can induce a demand shortfall larger than the shock itself, and in Bonadio et al. (2021), whose many-country model quantifies transmission through global supply chains. Neither framework separates the within-border margin studied here. Closest in spirit, Khanna et al. (2022) find, using Indian firm-to-firm data, that firms buying complex, hard-to-substitute inputs broke fewer links and held up better; their resilience runs along input specificity and ours along contracting reliance, and the two are consistent once contract-reliant industries are read as the substitutable, easily-contracted margin that lost its in-house buffer.

Where the pandemic literature has measured the international margin of exposure to foreign suppliers and imported inputs (Lafrogne-Joussier et al., 2023; Meier and Pinto, 2024), this paper measures a domestic organizational margin of comparable magnitude that the import lever does not reach, identifies its output cost from a synchronized shock, and asks why that cost appears contingent on the kind of shock. Three papers are closest in object. Fort (2017) identifies ICT as a determinant of domestic versus foreign fragmentation, the sourcing margin measured here. Forbes and Lederman (2009) show that airlines keep in house the routes on which adaptation decisions are frequent and costly, and Macchiavello and Morjaria (2015) show that during a supply shock Kenyan rose exporters allocated scarce output across buyers according to the value of each relationship. Forbes and Lederman’s result is also an identification threat. If producers kept in house the steps that are hardest to manage at arm’s length when conditions change, pre-determined reliance is not exogenous. The selection would run against our finding, however: the industries that stayed reliant would be those where adaptation mattered least, which biases the estimate toward zero. Macchiavello and Morjaria’s evidence is the closest micro analog of the rationing mechanism of Section 7.2. One contribution relative to these papers is a shock-identified output consequence of contract reliance across an industrial economy. Another is a production-function account of why that consequence may be contingent on the kind of shock. A third is on measurement and does not depend on the shock: Section 3.1 establishes the margin’s domesticity through checks that other work can apply to the same Census item, namely a customs-based floor on offshore processing, a European balance-of-payments mirror, a Mexican mirror, and the receipts side of the Economic Census.

In Section 2 we build the panel and show that contract reliance is a structural attribute of an industry: about four-fifths of its variation is across industries rather than over time. Section 3 establishes that the margin is predominantly domestic and that foreign sourcing did not displace it. In Section 4 we estimate the output cost of the 2020 shock, test it against customer shocks, steel tariffs, and composition, and show that it fell on volume and opened at reopening. Section 5 follows the gap through 2024 and documents the pull-back from contract production, and Section 6 re-estimates the design in five European countries and Japan. Section 7 asks why the cost appears contingent on the kind of shock, and Section 8 turns to incidence and policy.

2 Data and Measurement

The estimation panel is a set of 358 six-digit NAICS manufacturing industries observed annually from 2005 to 2024 (188 of them before 2012), with a gap in 2017 (a Census of Manufactures year for which contract-work detail is not published).

Contract-production reliance. Contract-work and shipments data come from three successive Census instruments: the Annual Survey of Manufactures through 2021, the 2022 Economic Census, and the Annual Integrated Economic Survey, which absorbed the ASM beginning with reference year 2023 and whose 2024 wave is the panel’s final year.⁶ Contract-production reliance is

$$\text{CPR}_{it} = \frac{\text{cost of contract work}_{it}}{\text{value of shipments}_{it}}. \quad (1)$$

Read plainly, the ratio answers one question: of each dollar an industry ships, how many cents were paid to outside establishments for work performed on the industry’s own materials? Because the measure is a ratio, a level discontinuity in dollars across the survey redesign cancels, and the data bear that out, so the years pool under industry and year fixed effects.⁷ A second seam, at 2011-to-2012, changes coverage only. It cancels in the ratio as well, and none of the acute, persistence, or Tōhoku estimates spans it.⁸

Scope of the measure. The Census defines the cost of contract work as the cost of work done by others on materials furnished by the establishment. This is a *specific* sourcing margin—tolling and commission processing of the firm’s own materials—not the full make-or-buy boundary. A producer that outsources by buying finished components records that in the cost of materials, not contract work, and a vertically integrated producer and an arm’s-length buyer can both show low contract reliance. The measure should therefore be read as reliance on outside contractors for production steps. It is not the inverse of vertical integration.

A second limitation of the measure is a truncation at the extreme of outsourcing, and its direction can be signed. Producers that outsource *all* physical production exit manufacturing into wholesale as factoryless goods producers (Bernard and Fort, 2015; Kamal, 2018) and leave this frame entirely, so they cannot contaminate the panel, though the measure understates outsourcing at that extreme.⁹ The truncation compounds a bias Houseman et al. (2011) document for these data through a different channel. They show that price declines on offshored inputs go largely uncaptured, so the statistics understate real input growth and overstate the real value added and productivity of surviving manufacturing. Frame exit, our own inference here rather than a finding

⁶The ASM is drawn from two API endpoints covering 2005–2016 and 2018–2021. The 2023 and 2024 codes (NAICS-2017 basis) map to the panel spine as an identity; the 2022 codes (NAICS-2022 basis) are bridged with the Census 2022-to-2017 and 2017-to-2012 concordances, recovering 330 of the 358 spine industries. Industries with fully suppressed survey records drop from that year’s outcome sample—31 in 2023, among them automobile and petrochemical manufacturing, and 27 in 2024.

⁷The within-industry change in CPR across the 2021-to-2023 transition has a median absolute value of 0.0019, in line with the 2016-to-2018 endpoint seam (0.0019) and with ordinary adjacent-year movement; the 2023-to-2024 change, within the same survey, is 0.0012.

⁸We disclose this seam because the cross-sectional design rests on the pooled panel. The pre-2012 API endpoint publishes contract-work detail for 188 of the 358 spine industries and the post-2012 endpoint for all of them, so shipments summed over the panel jump by two-thirds between 2011 and 2012 while the published all-manufacturing total rises 4 percent; the seam is one of coverage alone, leaving units and activity unaffected. Because reliance is a ratio it cancels in the treatment: the median absolute within-industry change in reliance across 2011-to-2012 is 0.0016, in line with ordinary adjacent-year movement. Two exhibits span the seam: the 2008–09 placebo window (2005–2012) ends in the seam year, and the production function of Table 2 pools 2005–2016 in deflated quantities under industry fixed effects, which absorb a one-time level shift. The variance decomposition of Table 1 is estimated on the 2005–2021 base panel; re-estimated on the 2012–2021 years alone (on the raw six-digit file rather than the 358-industry spine), the between-sector share is 0.35 and the within-sector, between-six-digit share 0.55, so the seam does not generate the cross-sectional variation the design uses.

⁹Bounding how many producers sit near that boundary is beyond an industry panel, and it bears only on the most extreme cases.

of theirs, adds a second understatement of how much sourcing is occurring at all. Both work against the gradient we find, since the industries most likely to cross into the factoryless frame are precisely the contract-reliant ones.

Work performed by *other establishments of the same firm* also counts, so the item includes intra-firm tolling alongside arm’s-length work (Ramondo et al., 2016); a multi-plant firm can re-prioritize across its own network in a way the operational-control reading does not contemplate, and the two cannot be separated at the industry level. The concern is bounded on two sides. The output response appears in the U.S. and Japanese measures, which are establishment-based and *include* intra-firm tolling, and equally in the European measure, which is enterprise-based and *nets it out* (Appendix C), so an artifact of intra-firm tolling would not survive that switch of frame.

Other measures. Productivity and real quantities come from NBER-CES, real activity through 2024 from the Census instruments above, R&D intensity from the non-production worker share, and foreign-input exposure from the NAICS-6 import penetration of intermediate inputs, with a China-specific component used below.¹⁰ Appendix Table A13 reports summary statistics. One magnitude recurs throughout: contract work is a minor input, averaging about 1.6 percent of total manufacturing cost, and Section 7.1 develops what follows from that share.¹¹

Variance decomposition. A variance decomposition of contract-production reliance (Table 1) attributes 81 percent of its variation to differences across industries and only 19 percent to within-industry movement; the largest single component, 50 percent, is across six-digit industries within a three-digit sector, and the common year effect is essentially zero. (On the balanced sub-panel the cross-industry share is 0.73.) Contract reliance is thus a structural attribute of an industry rather than a cyclical margin. This shapes what can be identified. A design that leans on within-industry time-series variation runs on a thin, noisy slice; the credible variation is cross-sectional, or comes from a sharp shock. Accordingly, the central design below uses pre-shock reliance as a cross-sectional treatment.

3 Contract Reliance and Imports

The vulnerability this paper studies sits on a margin separate from imports on two counts: it is predominantly domestic, and foreign sourcing did not displace it. Both cut against the premise that reversing offshoring would rebuild the domestic capacity at risk.

3.1 Location of contract work

Where are the contractors? The Census definition does not restrict their location, so we establish the margin’s domesticity empirically, with five checks.

¹⁰NBER-CES supplies total factor productivity through 2016 and real output, employment, materials, and value added through 2018. The R&D proxy is validated against NSF Business Enterprise R&D expenditure (National Center for Science and Engineering Statistics, 2024); the trade measures are built from the HS10-to-NAICS6 concordance of Pierce and Schott (2012).

¹¹Scaled instead by the value of shipments, which exceeds total cost, the same spending is the 1.2 percent reliance ratio of Appendix Table A13. Each denominator is used where it is natural: shipments for the reliance measure, total cost for the input’s expenditure share.

First, across industries CPR is nearly orthogonal to import penetration: a regression of the industry mean of CPR on the HS10-derived foreign-input share (Pierce and Schott, 2012) yields a cross-industry R^2 of 0.01 (Pearson $r = 0.10$), and the same holds within industries over time—with industry and year fixed effects, import penetration explains essentially none of the variation in contract reliance (within R^2 below 0.001), so the industries whose imports rose were not the ones that leaned harder on contract work (Figure 1).¹²

Second, the only directly measured component of offshore contracting is small. The customs provision that records offshore processing of U.S.-owned components and their re-import (HTS 9802.00) ran \$2–3 billion of value added abroad per year over 2017–2021 (on gross 9802 value of \$4–5 billion). Against total manufacturing contract-work spending of \$53 billion in 2019, that is a directly measured offshore share of 5 percent on the value-added basis and 9 percent on the gross basis. This is a floor, not a ceiling: a 9802 claim is optional and worth filing only when it lowers duty, and USMCA (like NAFTA before it) admits qualifying Mexican-processed goods duty-free without one, so the provision records a shrinking, self-selected slice of offshore processing. The domesticity reading therefore does not rest on 9802; it rests on the orthogonality above and the mirrors below.

Third through fifth, three mirror checks point the same way. The balance-of-payments item that would measure the fee side directly—manufacturing services on physical inputs owned by others—is not published for the United States. The partner mirror, however, is informative. EU-27 credits vis-à-vis the United States on that item, the fees European firms earn processing U.S.-owned inputs, ran about 1.9 billion euros in 2019, roughly 4 percent of contract-work spending.

The receipts side of the same market corroborates this. Domestic establishments *selling* contract manufacturing services reported receipts of \$41 billion in 2017 and \$46 billion in 2022 against contract-work purchases of \$49 and \$51 billion in those benchmark years, positively rank-correlated across industries.¹³

The Mexican mirror addresses the largest offshore counterpart. Mexican contractors earn about 85 percent of their income from foreign principals, overwhelmingly U.S. firms (about \$33 billion in 2019). But the maquiladora model is factoryless goods production, so those principals are predominantly wholesale-classified and sit outside the manufacturing buying panel. The manufacturer-side offshore payments that would enter it are the ones the 9802 measure already records. The unmeasured piece is a scope boundary, offshore processing bought by U.S. wholesalers rather than manufacturers, not a hidden foreign share of the domestic measure.¹⁴

Our results here are therefore consistent with Fort (2017), who finds domestic fragmentation to be the larger and more ICT-sensitive margin, and with the evidence that outsourcing within U.S. manufacturing is substantial and imperfectly captured in the standard statistics (Houseman, 2007;

¹²A firm-based measure agrees: the Census experimental Trade by Industry and Product Statistics assign each import to the industry of the importing firm through the Business Register, and matched to contract reliance across manufacturing industries the cross-industry correlation is near zero (Pearson 0.02, Spearman 0.11), so the orthogonality does not turn on the concordance method.

¹³Because seller receipts also include work for wholesale-classified factoryless principals, the match corroborates a domestic seller population of the right order.

¹⁴Mexico is the one system that publishes both the buying and the selling item, so the concept can be checked from both ends. No payee-location split of the Mexican expense item exists, but three checks read it as domestic to first order: balance-of-payments manufacturing-services credits match the survey’s foreign-client earnings nearly one-for-one; the debit series falls 89 percent over 2019–2023 with Pemex’s crude-toll-processing wind-down while industries’ contract spending stays nearly flat (–7 percent); and domestic contractors’ income from domestic principals is roughly twice manufacturers’ total contract spending. The Mexican payer-side item averages 1.2 percent of gross output, close to the U.S. ratio.

Houseman et al., 2011). The foreign-input margin is real, and we estimate it directly in Section 4; it is simply recorded in a different variable.

3.2 Foreign sourcing and contract reliance

The margin is predominantly domestic (Section 3.1). The other leg of the reshoring premise is substitution: that foreign sourcing displaced domestic production, so reversing it would rebuild domestic capacity. A shift-share test in the tradition of Autor et al. (2013), using foreign suppliers’ export growth to other high-income countries as supply-side variation in U.S. foreign-input availability, tests it. If foreign sourcing displaced domestic contracting, industries that gained more access to Chinese inputs should have cut reliance. Instead, the point estimate is positive: more China-sourced input availability is associated with *higher* contract-production reliance. Admittedly, the first stage of the estimating instrument is weak—a NAICS-3-clustered F near 5.7—so we read the test as evidence against substitution, and take no magnitude from it; the Anderson and Rubin (1949) weak-instrument-robust interval excludes zero under the original China-instrument NAICS-3 construction ($[0.004, 0.254]$) and includes it under both alternatives: the instrument rebuilt at NAICS-6 ($[-0.011, 0.144]$) and the aggregate foreign-share instrument at NAICS-3 ($[-0.024, 0.160]$). The China-instrument 2SLS point estimate is $+0.040$ with a conventional p of 0.14.¹⁵ Either way, the test gives no sign of a substitution for an import restriction to reverse.

4 The 2020 Output Loss

Contracting a production step out diversifies idiosyncratic supplier risk but loads on a common supply factor, as Section 7 develops. It follows that a broad, synchronized shock should fall hardest on the most contract-reliant industries. The COVID-19 disruption of 2020—acute, dated, and economy-wide—is that common-shock case. This section estimates that cost, tests it against the alternatives that could produce the same pattern, and compares it with the other exposures competing to explain the same collapse.

4.1 Design

We use pre-COVID contract reliance as a pre-determined, cross-sectional treatment interacted with the shock, alongside the foreign-input-dependence channel:

$$Y_{it} = \beta (\text{CPR}_i^{\text{pre}} \times \text{post2020}_t) + \gamma (\text{Foreign}_i^{\text{pre}} \times \text{post2020}_t) + \mu_i + \delta_t + e_{it}, \quad (2)$$

where $\text{CPR}_i^{\text{pre}}$ is the 2015–2019 mean of contract reliance (standardized), $\text{Foreign}_i^{\text{pre}}$ is pre-COVID China-input exposure (standardized), μ_i and δ_t are NAICS-6 and year fixed effects, and the window is 2018–2023; adding the 2024 survey year leaves Panel A and the Panel B nulls in Table 3 essentially unchanged (its note reports the 2018–2024 window, where one Panel B row moves at baseline). One standard deviation of reliance—1.6 cents of contract work per dollar shipped—compares with

¹⁵A Rotemberg-weight decomposition (Goldsmith-Pinkham et al., 2020) confirms China carries the identification (weight 1.16; its own unclustered first-stage F , a component diagnostic rather than the strength of the instrument used above, is 48); the pooled multi-partner instrument carries negative weights and is not used. Pushing inference onto the single-country shock dimension (Borusyak et al., 2022) widens the interval to include zero, and a 2015-renminbi-devaluation natural experiment reproduces the point estimate ($\hat{\beta} = 0.039$). Full diagnostics are in the replication package.

a median industry at 0.7 cents and a ninetieth percentile at 2.8 (Appendix Table A13).¹⁶ The identifying variation is only the interactions: the cross-industry level of each exposure is absorbed by μ_i and the common COVID timing by δ_t , so β is the differential post-2020 movement of the outcome in high- versus low-reliance industries. Hence, a concrete example of the comparisons being made by the estimator is between post-2020 output in industries such as forgings and powder-metallurgy parts, where contract work runs well above the 2.8-cent ninetieth percentile, and in industries near the 0.7-cent median, in the same years.

Every number that follows depends on two features of the estimator. β is an equal-weighted-industry estimand, and least squares does not weight industries equally in identifying it. The regression’s own identifying weights concentrate sharply: the Kish effective sample is 31 industries of 358, with fabricated metal products, apparel, and leather carrying most of the weight, and mean contract reliance in that effective sample is 5.6 percent of shipments against 1.2 percent in the panel (Appendix A, Table A4). What β estimates is the response of a substantially reliant industry. It is neither the average manufacturing industry nor the average manufacturing dollar. The signed contributions behind that leverage are diffuse and two-sided, so the concentration does not make the estimate fragile; but it does define what the estimate is about, and the weighting path in Appendix A shows that the acute and pooled results do not carry over to the dollar-weighted estimand while the persistence contrast does.

Because the treatment is continuous, parallel trends must hold in the dose: absent the shock, outcomes would have evolved independently of the level of reliance, and the per-standard-deviation coefficient summarizes dose-specific responses under a linearity restriction (Callaway et al., 2024, henceforth CGS). The above-median contrast (−6.4 percent; Table 3) and the quartile dose-response (Appendix A) report the same object without that restriction.¹⁷ Identification rests on three conditions: (i) conditional parallel trends, tested below; (ii) contract reliance being pre-determined; and (iii) the absence of a confounding shock correlated with reliance beyond what the fixed effects and the controls absorb. The binding threats are that reliance proxies for sector type, or for industry type within sectors, which we address with composition controls and direct horse races in Section 4.3.

Clustering follows the assignment design (Abadie et al., 2023): the panel is the population of six-digit manufacturing industries rather than a sample of sectors, and reliance varies at NAICS-6, with most of its cross-sectional variance across six-digit industries within a sector (Table 1), so primary inference clusters at NAICS-6. Coarser levels bound any residual sectoral correlation from above, and Section 4.5 reports that full ladder through NAICS-4 to NAICS-3, where the 21-cluster wild bootstrap (Cameron et al., 2008; MacKinnon and Webb, 2017; Roodman et al., 2019) is the conservative benchmark.

4.2 Baseline estimates

Table 3 reports estimates of equation (2). Contract-reliant industries lost more output: about 2.4 percent per standard deviation of pre-COVID reliance (1.6 cents of contract work per dollar shipped, against a 0.7-cent median), or roughly 1.5 times the fee that one standard deviation represents.

¹⁶This 1.6-cent standard deviation is numerically close to, but distinct from, the 1.6 percent average cost share of Section 2: the first is the spread of the reliance ratio across industries, the second its mean as a share of total cost.

¹⁷Because the shock date is common to every industry, the design has no staggered adoption, so the two-way fixed-effects estimator makes no forbidden comparisons and places non-negative weight on the underlying dose-specific effects. Estimating the CGS dose-response objects directly leaves the sign intact and the loss concentrated at the top of the reliance distribution, so the linear slope is the conservative summary; the dose-specific effects and their inference are in Appendix A.

In other words, an industry paying 1.6 cents more per dollar shipped to contractors lost about 2.4 cents more of each dollar of output. An aggregate translation, and why the design does not support it, is in Appendix A. The foreign-input-dependence channel is comparable in size over the full window (about -1.8 percent for output, -2.4 percent for employment), and the two exposures are identified separately in the same regression. Both sourcing margins, contractors at home and suppliers abroad, shaped which industries held output. Their timing differs, however, and what that implies for reading γ is taken up in Section 4.7.

Nor did the contraction come through profitability: the operating-surplus share, the labor-cost share, the value-added share, and per-worker productivity show no contract-reliance effect. Employment falls in the baseline but is the more fragile of the two real outcomes, as Section 4.5 shows.

Figure 2 reports the event study with 2019 as the base year. The output interaction is flat before 2020, breaks sharply downward in 2020, deepens through 2022, and then narrows: the 2023 and 2024 coefficients are about -0.02 , roughly four-fifths of the acute-year loss, and no longer individually distinguishable from zero. Employment does not narrow. Its gap reaches -0.035 in 2024, five years after the 2019 base year, and is the widest in the series. Estimating both channels’ year interactions jointly places the acute break on contract reliance alone (Table 3, Panel C): the China-exposure coefficients for 2020–21 are -0.004 and -0.001 , against -0.023 and -0.031 for reliance, and the China-exposure loss arrives only in 2022–24. The two exposures are separable in time as well as in the cross-section.

At monthly frequency the acute loss dates itself more precisely still. If lockdown incidence drove the loss, the reliance gradient should be widest in April and May 2020, when the shutdowns bound. Instead, the reliance gradient is *absent* during the lockdown months of spring 2020 and opens as the aggregate reopens. In the Federal Reserve’s G.17 production indexes, aggregated to 63 industry groups covering 91 percent of the panel’s 2019 shipments, it is widest in July–August (-3.1 and -2.6 percent per standard deviation; June–December pooled -2.1 to -2.3 percent). By those months the China-input effect in the monthly import-shock literature had already dissipated (Meier and Pinto, 2024; Lafrogne-Joussier et al., 2023). Contract-reliant industries were not differentially shut down; they failed to differentially restart. That timing separates the contract-reliance margin from lockdown incidence and from the import channel in one exhibit, and it is the pattern the mechanism discussion of Section 7.2 builds on; the pooled June–December coefficients carry p between 0.08 and 0.12 at the 63-group aggregation. Construction detail is in Appendix A.

4.3 Alternative explanations

Customer shocks and steel tariffs. Is the estimate a reliance effect, or the shock that hit these industries’ customers? The industries that carry most of the estimate are metalworkers (forging, powder metallurgy, foundries, oilfield machinery; Table A6), and in 2020 their customers took shocks of their own: the oil price collapse, the 737 MAX production halt and the grounding of airline fleets, and the spring shutdown of vehicle assembly. Section 232 steel tariffs had raised their input costs from 2018. We control for the tariffs with pre-period steel and fabricated-metal input shares and measure exposure to the three customer groups in two ways, interacting each measure with post-2020 (Table 3, Panel D; Appendix Table A7). The first is the share of an industry’s commodity output bought in 2017 by oil and gas extraction, aerospace, and motor vehicles and parts, from the BEA detailed Use table, whose commodity detail is coarser than NAICS-6. The second measure varies at six digits: the density of each customer group’s employment in the counties

where an industry’s workers are, relative to the national density, a standard proxy for input-output links (Ellison et al., 2010). Both measures predict 2020 output losses with similar strength; their construction is in Appendix A.

With the tariff shares and the six-digit measure, the 2020 estimate is -1.6 percent ($p = 0.014$) and the 2018–2023 estimate -2.1 percent ($p = 0.010$), and the tariff shares do most of that work; co-location alone moves the 2020 estimate only from -2.2 to -2.0 percent. The BEA shares move it further, to -1.35 percent ($p = 0.023$; 0.018 clustering on sectors), and adding all three sets leaves -1.4 percent in 2020 and -1.7 percent over 2018–2023 ($p = 0.030$). We read -1.35 percent as a lower bound, not as a point estimate of the reliance effect. The six-digit measure predicts the losses about as well as the BEA shares yet leaves most of the reliance estimate in place, which is what one would expect if the commodity-level shares partly absorb the between-industry variation that reliance itself carries. Three further checks point the same way. Among auto suppliers alone, the top quartile by auto customer share, reliance still predicts a loss of 2.4 percent. If the vehicle-assembly shutdown carried the reliance estimate, the reliance gap should open with the auto exposure in March–May 2020. Instead, the auto exposure registers in those months while the reliance gap is zero; the reliance gap opens in the reopening months, and the exposures take about two-fifths off it, as in the annual data, although on 63 industry groups neither version is significant at 5 percent. With the BEA controls interacted with year, the leads stay flat ($p = 0.65$) and the gap still widens after 2020, to -2.5 percent in 2021 and -3.7 percent in 2022. The harshest variant drops every industry in the top decile of any BEA customer share, a quarter of the panel, and leaves -1.1 percent ($p = 0.12$).

Composition. A broader threat is sector type. Interacting post-2020 with pre-COVID upstreamness (Antràs et al., 2012), the consumer-demand share, an occupation-based teleworkability index, durable status, physical proximity, and labor intensity leaves the output effect at -0.024 to -0.028 —if anything, further from zero. Because the controls push the coefficient away from zero rather than toward it, its stability across specifications is the reassurance. Selection on unobservables would have to run opposite to selection on observables to overturn it.¹⁸

Three of those controls carry the demand- and lockdown-based alternatives, against the readings that these industries simply faced falling demand, could not send their work home, or worked at close quarters. A direct control for the 2020 shutdown designation points the same way, holding the effect at -0.022 (Table 5) with a null essential interaction. Essential industries are if anything *less* contract-reliant ($r = -0.31$), so the shutdown margin works against the estimate. Thin inventory buffers, the mediator that amplifies shipping delays in Alessandria et al. (2023), run the other way: reliant industries hold slightly *more* inventory before the shock. Purchased temporary staffing, a labor-supply channel of about the same size as contract work that could mimic the timing, is only weakly related to reliance and carries nothing when raced against it. Constructions of the occupation-based and inventory controls, the in-window inventory dynamics the pre-2018 series cannot reach, the staffing race, and a decomposition of the response into intensive and extensive margins are in Appendix A.

Industry type. The loss concentrates in low-R&D industries, on a cell-by-cell pattern we read as descriptive rather than as a tested contrast (Appendix A, Table A3): the commodity margin most easily contracted out. That raises a sharper version of the composition threat: that reliance

¹⁸The Oster (2019) δ that would drive the coefficient to zero is -4.8 at $R_{\max} = 1.3\bar{R}$; the bound shrinks toward zero as $R_{\max}$ rises, and because the controls barely move the coefficient or the within- R^2 , the extrapolation is highly leveraged, so the direction and stability of the coefficient carry more information than any single δ .

merely marks the standardized, low-R&D end of manufacturing. It does not. Contract reliance is essentially uncorrelated with R&D intensity ($r = -0.02$ with the non-production worker share), so the treatment does not sit on the R&D axis. Racing the two interactions leaves reliance unchanged (about -2.4 percent) while R&D-by-post carries nothing. With R&D, size, and the full composition controls entered together, reliance stays intact at -0.026 . The R&D interaction even turns positive ($+0.036$): once demand-side attributes are held fixed, high-R&D industries fare relatively better, mirroring the concentration just described. Pre-period industry scale is null and leaves the estimate unmoved (Table 5), and a placebo sharpens the point: another minor input’s cost share—energy, of comparable size—interacted with post-2020 yields nothing (Table 5). The contraction attaches to the contracted-production margin. Low-R&D status, size, and small input shares in general do not carry it.

Trade exposure and measurement. Four one-issue races, each a row of Table 5, close the remaining measurement readings. Two ask whether the contraction is really trade exposure at one remove, and neither says it is. A pre-determined upstream COVID-exposure control, the input-weighted 2019–2020 contraction of an industry’s supplier sectors, leaves the effect at -0.021 , and a finer NAICS-6 upstreamness measure is null. Embodied import content, the foreign content reaching an industry indirectly through its domestic suppliers, is the sharper version: reliance is essentially orthogonal to it (cross-industry correlations -0.16 with the through-supplier component, 0.02 with the total), and adding it leaves the effect at -0.024 while the embodied channel is itself present and correctly signed (about -0.023). Entering the direct and through-supplier components separately—collinear with each other and with the China channel—is the one variant that moves the estimate, to -0.019 (CRV1 $p = 0.025$, wild 0.081), roughly four-fifths of the baseline. The other two ask whether the result turns on how a variable was measured, and it does not: substituting total import penetration for the China measure, rebuilding the import measure from public BACI flows, or deflating by industry producer prices each leaves reliance at about -0.024 (Table 5).

4.4 Within-sector variation

Which comparison carries the 2020 loss? Mostly a comparison across sectors. Inside three-digit sectors the 2020 estimate is -0.8 percent and not significant; a within-sector gradient appears over 2018–2023, and it depends on how customer exposure is measured. Section 6 runs the same checks on the European panel.

Sector-by-year fixed effects. Half the panel variance in reliance, and 64 percent of the pre-COVID treatment cross-section, is across six-digit industries within a three-digit sector (Table 1), so the design can be pushed inside sectors, where the sector type common to a three-digit family is held fixed. Sector-by-year fixed effects, which absorb each sector’s whole pandemic path, leave the estimate at -1.8 percent per SD (-2.4 percent once composition controls are added). The within-sector estimate compares more and less reliant industries inside the same sector-year, not contract-heavy sectors against unlike ones, and the headline -2.4 percent is the between-inclusive estimate. It is not a floor: pushed one level finer the estimate falls further, so credibility and precision move in opposite directions along this ladder. Because sector-by-year effects push the identifying variation inside NAICS-3 clusters, where the 21-cluster bootstrap loses power, the specification’s design-based test is within-sector randomization inference: it gives $p = 0.055$, and the primary NAICS-6 clustering gives 0.044 , with the full ladder in Appendix Table A1. It survives a six-digit

measure of customer exposure but not the coarser commodity-level one (see the customer-shock paragraph above).

The within-sector estimate depends on which customer measure is used. Inside three-digit sectors over 2018–2023 it is -1.7 percent ($p = 0.054$) with the six-digit measure and the tariff shares, close to its unconditional -1.8 percent, and -1.2 percent ($p = 0.18$) with the BEA shares. We therefore treat the within-sector gradient as suggestive.

Pushed one level finer still, to NAICS-4-by-year fixed effects, the gradient genuinely flattens. The coefficient falls to -0.013 , roughly half the baseline, and out of significance at any clustering level. A split-sample measurement-error correction attributes only about a fifth of that decline to attenuation, so the steeper slope sits in the between-family variation (Appendix A, which also reports the same correction at the three-digit level).

A resolution decomposition states what the ladder identifies. Correlating the 2019-to-2020 change in log shipments with pre-COVID reliance measured at successively finer levels, the correlation is -0.25 at the three-digit sector mean, -0.21 at the four-digit family mean, -0.21 at six-digit, and -0.06 on the within-four-digit residual, which carries 38 percent of the variance in reliance. The coarser the measure, the better it predicts the collapse. In other words, knowing an industry’s three-digit sector predicts its 2020 loss at least as well as knowing its own reliance. The design therefore establishes that reliance marks the industries that lost output and that the marking survives sector type; it does not establish that a producer’s contracting choice, holding its product fixed, carries the loss, because the variation that could identify that choice, within four-digit families, carries almost none of the response. The organizational reading requires the establishment-level design named in the introduction.

4.5 Inference

The clustering ladder. Appendix Table A1 reports the full ladder of clustering choices for the two key specifications, none suppressed; Section 4.1 gives the rationale for taking NAICS-6 as primary. The baseline rejects at every rung: $p = 0.002$ at NAICS-6 ($G = 358$), 0.010 at NAICS-4 ($G = 86$, wild cluster bootstrap 0.029), and 0.022 under the 21-cluster NAICS-3 wild bootstrap (Webb weights, null imposed), the standard small- G calibration (Cameron et al., 2008; MacKinnon and Webb, 2017; Roodman et al., 2019) and the most conservative rung. A collapsed pre/post cross-section, which removes the serial-correlation rationale for clustering altogether (Bertrand et al., 2004), gives $p = 0.001$ with heteroskedasticity-robust errors and 0.006 clustered at NAICS-3 (Table 5); within-sector randomization inference, which holds sector composition fixed, gives 0.030 . The two procedures that treat the 21 sectors themselves as the sampling unit are the ladder’s outer bound: a sector block bootstrap gives $p = 0.068$ and Ibragimov–Müller does not reject ($p = 0.23$). The last fails for lack of power over a heterogeneous set of sector coefficients, twelve of them negative and nine positive, among which electrical equipment alone is -0.40 on a range of reliance that lies wholly below the panel mean (Appendix A). An exact decomposition of the pooled coefficient into industry contributions shows who carries it. Ten industries holding 0.9 percent of shipments, forgings, powder-metallurgy parts, foundries, industrial molds, oilfield machinery, and apparel contractors among them, carry 60 percent of it, and fabricated metals, machinery, primary metals, and food carry 85 percent of the sector-level sum (Table A6). The 21 leave-one-sector-out refits all lie in $[-0.028, -0.020]$, so no sector is necessary to the pooled estimate, which says nothing about uniformity across sectors. What holds at every rung is the above-median contrast (-0.064 , $p < 0.001$; wild NAICS-3 $p = 0.008$). The employment effect is weaker under the conservative

bound (wild $p = 0.029$ at baseline, 0.076 under controls). Table 5 collects the remaining checks and Appendix A reports them in full.

4.6 Magnitude

How large is a loss of 2.4 percent per standard deviation? We benchmark it in two ways, each on the same panel and shock with the standardized exposures entered one at a time. Figure 3 illustrates the comparison. On a common footing, contract reliance moves output about as much as the foreign-sourcing exposure on which the pandemic literature has concentrated (Lafrogne-Joussier et al., 2023; Meier and Pinto, 2024): -2.7 percent per standard deviation, against -3.0 percent for total import penetration and -2.2 percent for Chinese-input exposure, through a margin that literature does not measure. The exposures the tariff instruments actually reach—Section 301 and Section 232 coverage—carry no significant response of their own. A Shapley decomposition of the cross-industry variance in the 2019–2020 shipments change yields the same ranking (panel b). Contract reliance accounts for 3.9 percent of the total, nearly two-thirds of what the three exposures jointly explain, against 1.6 percent for import penetration and 0.6 percent for teleworkability, which stands in for lockdown exposure. Contract reliance ranks first whichever foreign measure it is benchmarked against. That ranking is conditional on the exposure set, and one variant reverses it: adding Section-232 steel-input coverage and Section-301 tariff exposure raises the total explained variance from 6.1 percent to 9.3 percent and puts steel first (3.9 percent of the variance, 42 percent of the explained portion, against 3.0 percent and 32 percent for contract reliance). We do not lead with that variant because steel-input coverage reads as a proxy for durable-metals composition, which had its own 2020 collapse, and neither tariff measure carries an individually significant output response. Entered jointly with foreign exposure, teleworkability, physical proximity, and upstreamness on one fixed sample, the reliance coefficient holds at -0.025 .

Two limits discipline the comparison. The three exposures together explain only about 6 percent of the cross-industry dispersion in 2020 output losses, so the shares above are shares of that explained portion. The ranking is also within a single shock; how contract reliance would rank under a disruption of another kind is the question Section 7 bounds.

4.7 Interpretation

An input of this size might be expected to move output by at most its own share. It moves output by more. One standard deviation of reliance is about 1.6 cents per dollar shipped, and the coefficient runs about 1.5 times that fee. Some gap of this kind is expected. Contract work is a charge for an operation performed on materials the buyer already owns, and those materials sit in the buyer’s cost of materials rather than in the fee; the fee therefore understates the throughput at risk. It does not run this far. Calibrating gated throughput from the toll rate, the contractors’ own 2020 throughput loss, and a downstream multiplier reproduces at most about a fifth of the estimated response, on routes and criteria fixed before measurement (Appendix B, Table A8).

That failure is informative about what the treatment measures. If the fee’s own throughput can account for only a fifth of the response, then reliance is not usefully read as a small input share whose outage propagates in proportion to its size. It marks a production arrangement. Two facts already point the same way. Contractors did not stop producing in 2020, so the loss is not the arithmetic of a missing input (Section 7.2), and the response survives controls for the input’s closest observable analogs, including a placebo on another minor input’s cost share (Section 4.3). The residual is the price of that reading. What separates a production arrangement from the character

of the production routed through it is the firm-level bundle of the introduction; the calibration bounds the mechanical pass-through channel but does not isolate what fills the remainder.

Complementarity between contracted and in-house tasks is the natural reading of that residual. If the contracted step cannot be substituted for within the buyer’s own production, a disruption to it idles the in-house tasks it feeds, and the output loss is bounded by the step’s role in production rather than by its cost share. Baqaee and Farhi (2019) give the general statement: away from the Cobb–Douglas case a sector’s contribution to the aggregate loss departs from its Domar weight, and under complementarity the departure is upward for small inputs that the rest of production cannot do without. The near one-for-one transmission of disrupted imported inputs to output that Boehm et al. (2019) document is that elasticity measured at the firm level, and the fee-to-response ratio here is the same object at the industry level. What the reading requires is an elasticity of substitution between the contracted step and the rest of production well below one; the industry data do not identify it. One unit conversion also belongs here. The framework of Section 7.2 differentiates with respect to a share of tasks, while the empirical treatment is a fee over revenue, and the two coincide only up to the toll rate that Appendix B measures at 0.51, so a 1.6-cent fee corresponds to roughly a 3 percent task share. Table 7 collects what the paper’s timing facts do to the candidate mechanisms of Section 7.2: contractor output held up in 2020, the buyer loss opened at reopening, and reliance stayed low afterward, which together rule out an outage and leave rationing, a missing in-house fallback, and a dissolved link standing, none of them tested directly.

Volume carried the loss, leaving margins and productivity untouched (Table 3, Panel B). The framework delivers that split by construction: a disruption removes a fraction of contracted output at unchanged unit economics, so it surfaces as lost quantity, and the business these industries retained ran normally. This operational-control reading is an interpretation, not a finding: the contract-work item records spending rather than control, so at the industry level it cannot be separated from the alternative that the standardized production routinely routed to contractors was itself more exposed to the shock.

The timing of Table 3, Panel C completes the reading of γ : the foreign channel’s annual footprint arrives only in 2022–24, pooling disruption with the goods-cycle unwind, so γ is a full-window average rather than an acute-disruption response.¹⁹

Relative prices complete the picture. Had contract-reliant industries been rationed while their customers’ demand held firm, their prices should have risen. Instead, relative prices did not move. Estimating equation (2) with the log industry producer price index as the outcome gives a precise zero through 2023, with or without the full controls, and the price event-study coefficients are individually insignificant in every post year through 2023. In 2024 the price coefficient on the PPI-balanced membership is +0.012 ($p = 0.08$), which is by construction the gap between that year’s real and nominal output coefficients (Appendix A); the two are not statistically distinguishable, so the final-year real persistence rests on a relative-price movement the data do not establish. Through 2023, a quantity loss with no relative-price response says the market did not treat these industries’ output as scarce: demand for it is elastic, and the reliant industry simply lost the sale. The aggregate supply-chain-pressure inflation of the period ran through import prices and economy-wide costs (di Giovanni et al., 2022), not through the *relative* price of the contract-reliant margin. Our null here is therefore consistent with that literature. That is what one expects at the low-R&D,

¹⁹An annual panel cannot see the V-shaped early-2020 import interruption documented at monthly frequency (Lafrogne-Joussier et al., 2023; Meier and Pinto, 2024), so the near-zero 2020–21 China coefficients are consistent with that literature; what the annual data add is that the differential loss still visible at annual frequency in 2020–21 belongs to the contract-reliance margin. The 2022–24 attribution is the one the demand-rotation checks of Section 5.2 complicate.

standardized end of manufacturing, where the contraction concentrates (Appendix A), and it makes the establishment losses of Section 5.1 easier to understand: capacity that loses volume without pricing power is not replaced.

Who, if anyone, filled the gap? The trade data give a partial answer: not imports. Estimating equation (2) with trade outcomes in place of shipments, import penetration of these industries' product markets shows no differential rise at any horizon through 2024 (the penetration series behind the foreign channel, available through 2021, agrees), and competing imports *fell* in the acute year, by about 3.7 percent per standard deviation; restricting to industries with import data in every year gives much the same. The synchronized shock hit foreign producers of the same goods, which is what Section 6 documents directly.²⁰ Apparent consumption, shipments plus competing imports, falls roughly one-for-one with the output loss (about -2.4 percent against a -2.8 percent shipments reference on the same panel). The differential loss was therefore not offset from abroad: total supply of these products to the U.S. market fell and stayed lower. Whatever demand substitution occurred ran across product lines, a margin an industry panel cannot see; it did not run from domestic to imported varieties of the same products.

5 Persistence, 2021–2024

This section follows the 2020 gap through 2024, one outcome at a time, and then turns to what reliant industries did with the arrangement itself.

5.1 Persistence across data sources

A gap that opens in 2020 and is still open four years later—in employment through 2024, in the establishment stock through 2023—is the paper's third result, and because it is the claim most exposed to a measurement artifact, we establish it in sources that share no collection machinery with the survey that supplies the treatment. Table 4 assembles them: a mandatory establishment survey, a voluntary payroll survey, an administrative universe of unemployment-insurance records, a business register, and a value-weighted production index. Seven of the eight rows show the gap. The eighth does not, and its identity turns out to be informative.

County Business Patterns establishment counts supply an extensive-margin counterpart to the persistence. Because the counts reference March of each year, the 2020 count predates the lockdowns and shows no differential movement. The gap then opens and widens through 2023. Dating the post period at 2021 and dropping the transition year, the loss is about 2.2 percent per standard deviation. It survives removing each industry's own pre-2020 trend, which leaves a flat detrended pre-period and a loss of about 1.4 percent—roughly three-fifths of the undetrended figure.²¹ The persistent output gap therefore extends past a flow shortfall. Contract-reliant industries lost establishments, a durable loss of capacity. The two margins can be separated exactly, and they operate in sequence: in 2020 the loss is production at plants already standing, with the establishment count contributing about a seventh of it, while by 2023 the count carries about four-fifths of the remaining gap (Appendix A).

Business Dynamics Statistics—which separates establishment entry from exit at the four-digit level—decomposes that decline. Its net-count specification reproduces the CBP estimate closely

²⁰The import series carries positive event-study coefficients in 2015–16 ($+0.067$, $+0.033$): import growth into contract-reliant industries was already slowing before COVID. The pre-period at the 2018 base of the estimation window is flat (-0.002).

²¹ $p < 0.001$ at NAICS-6, with a wild NAICS-3 bound of 0.009; the detrended estimate has $p = 0.026$.

(about -2.6 percent per standard deviation), and the split then attributes the loss to entry: post-2021 entry rates fall by 0.46 percentage points per standard deviation of reliance, about 6 percent of the 7.2 percent post-period entry rate for manufacturing as a whole.²² That benchmark is itself elevated: business formation surged nationally after 2020 (Dinlersoz et al., 2021; Decker and Haltiwanger, 2023), so the contract-reliant shortfall opens against a rising tide rather than a falling one. The exit margin does not move: the exit rate is statistically flat and exits themselves *fall*, the mechanical footprint of a shrinking stock closing at an unchanged hazard. Contract-reliant industries did not shed plants faster; they stopped adding them. The entry shortfall is open in the first COVID entry cohort (the BDS flow year covering March 2020 through March 2021) and remains open through 2023: the capacity lost is capacity not being replaced. Appendix A reports the details. The industry-specific detrending removes a linear pre-2020 trend but not a nonlinear acceleration, so the panel cannot fully separate acute scarring from a pulling-forward of the secular decline these industries were already in (Pierce and Schott, 2016). Whether the tail instead reflects the post-pandemic rotation of goods demand is taken up in Section 5.2: the gap survives that control essentially unchanged.

Outside the Census collection system the labor margin agrees. In the Quarterly Census of Employment and Wages, the administrative universe of establishments covered by unemployment insurance, observed at the same six-digit level, employment in contract-reliant industries falls in 2020 and stays below its 2019 level in every post year through 2024 (Table 4).²³ The administrative precision also surfaces a mild pre-period drift the survey series is too noisy to detect: contract-reliant industries were already declining relative to the rest by roughly half a percentage point a year before 2019. The 2020 break is four to five times that slope, and the detrended estimates absorb it—applying the same industry-specific detrending as the establishment series removes the pre-period trend and leaves a gap of about 1.7 to 2.3 percent in every post year through 2023. In 2024 the detrended gap thins to 1.6 percent ($p = 0.065$), at the edge of that range rather than outside it, while the raw gap does not thin at all.²⁴ BLS payroll employment at the four-digit level moves the same way, drift included.

One row of Table 4 shows no persistence, and it is exactly the one a reader of this literature would reach for first. In the Federal Reserve’s G.17 industrial production index the reliance gap opens in 2020 and then closes during 2021, which is the acute-then-recovery profile the pandemic literature reports. That divergence is a property of the object each source measures. The production index is a value-weighted aggregate across industry groups. The persistent real gap in our data lives in the unweighted six-digit cross-section, where it is also least robust to a pre-trend violation (Section 5.2), and, more weakly, in unweighted family aggregates. Between value-weighted family aggregates, what persists is nominal, carried by relative prices. Deflated shipments aggregated to the index’s own 63 groups reproduce the index almost exactly, while the same shipments at six-digit resolution retain the gap. A value-weighted, between-family real index is the single cut of these data in which no source shows persistence, ours included. The production index is exactly that cut. Reading the persistence therefore requires saying at which resolution and under which weights it holds, which is what Appendix A sets out.

²²BDS inference clusters at NAICS-4, the level at which its aggregated treatment varies; both estimates reject at conventional levels ($p < 0.001$ and $p = 0.005$).

²³The 2025 QCEW year is also available and shows much the same raw gap (-0.033). It overlaps the tariff wave that began in April 2025, a separate treatment, and Section 7.3 reports it as one.

²⁴Detrending does not flatten the pre-period entirely: the largest surviving pre-coefficient, -0.008 in 2017 ($p = 0.003$), is a residual of the 2015–2019 trend fit and runs a third the size of the smallest raw post-2020 coefficient, and half the smallest detrended one.

5.2 Pre-trends and demand rotation

Pre-trends sensitivity. Because the design rests on parallel trends, we bound the violation it would take to overturn the result (Rambachan and Roth, 2023): under their relative-magnitudes restriction the 2020 output effect stays significant for post-period violations as large as the single largest pre-period deviation ($\bar{M} = 1.0$) and the averaged effect survives half that, with breakdown values identical under either clustering, while the employment effect, whose pre-period drifts up, does not survive the same relaxation in its first post-year. We run the same sensitivity on the two series that carry the persistence claim, and they separate. QCEW administrative employment is the most robust series in the paper on this criterion: its first-post effect survives violations half again as large as the worst pre-period deviation ($\bar{M} = 1.5$), raw and detrended alike, and the averaged effect survives $\bar{M} = 0.5$. That the raw series does as well as the detrended one is not an accident of the correction, since its pre-period drift runs *upward* while the estimated effect is negative, so extrapolating the drift works against the finding rather than generating it. The real six-digit output series, with pre-years 2015, 2016, and 2018 on a chain-linked deflator, sits between the two. Its 2020 effect survives $\bar{M} = 1.0$; its 2024 effect and the 2020–24 average exclude zero only at $\bar{M} = 0$, so real-output persistence holds under exact parallel trends and not under any relaxation of them. The BDS entry margin does not clear the same bar: it excludes zero only at $\bar{M} = 0$, so any violation proportional to its pre-period deviations overturns it. The entry decomposition should accordingly be read as the most likely account of the establishment decline rather than as an independently robust result. Removing each industry’s own linear trend, fit on 2015–2019 and extrapolated forward, gives a flat detrended pre-period and a break at the shock: -2.2 percent in the acute window and -2.5 percent over the full window (Table 5; Appendix A). Fit on 2012–2019 instead, the full-window estimate is -0.019 and not significant ($p = 0.116$; Appendix A). The contraction is a break rather than the continuation of a differential trend.

Demand rotation. The persistence into 2023–24 invites a demand-side reading: contract-reliant industries sit at the standardized end of manufacturing, and the 2021–22 goods boom and its unwinding could have held exactly them down for reasons unrelated to 2020. We address the post-period path directly with a realized demand-rotation index and with the final-use shares interacted with year fixed effects. Neither moves the estimate. The full-window coefficient moves from -0.024 to -0.023 , and the drop-2020 persistence contrast from -0.025 to -0.024 (Table 5); the construction is in Appendix A. A variant that adds $\log(\text{shipments plus imports})$ as a control moves the coefficient to -0.009 ($p = 0.25$); that control contains the outcome, so we report it as the bad-control bound. If the goods boom and its unwinding drove the gap, it should open in 2021–22, with the rotation. Instead, the reliance gap opens in 2020–21, before the rotation years, and the channel whose decline coincides with the rotation window is the import one (Section 4.2).

5.3 Changes in contract reliance after 2020

The panel also records what reliant industries did afterward, and the answer is that they moved away from the arrangement. On the 297 industries observed in every year from 2016 to 2024, contract-production reliance fell from 1.01 to 0.75 cents per dollar shipped, 21 percent below its 2016–18 average. Contract work fell 11 percent in dollars while shipments rose 20 percent and the shipment-weighted producer price index rose 27 percent, so the contracted volume fell by roughly a third in real terms. A shift-share split attributes 94 percent of the decline in the aggregate ratio to within-industry change and almost none to composition. The fall begins in 2021 inside the Annual

Survey of Manufactures and continues through 2023–24 inside its successor survey, so no collection seam carries it.

Figure 5 shows where it happened. Sorting industries into terciles of pre-COVID reliance and indexing each tercile’s sums to 2019, the most reliant third cut contract work 23 percent by 2024 while its shipments rose 14 percent and its employment held at 99; the middle third held contract work flat as shipments rose 20 percent; the least reliant third raised contract work with shipments. Contract work in the reliant third fell five to eight times more than the output it served.

If plant exit drove the decline, the fall in reliance should track the fall in establishment counts. Instead, across industries the 2019–2024 change in reliance is unrelated to the 2019–2023 change in establishment counts (-0.14 per log point, $p = 0.64$, with pre-period reliance and sector effects), and the top tercile falls by the same amount on the 138 industries with no net plant loss. Even an exit concentrated at the most reliant plants of each sector could move the ratio by about zero against the observed median fall. What fell is contract work at surviving plants.

Whether those plants replaced the work in house the panel can say only through 2022, and imprecisely. Table A5 reports the post-2020 change in each cost item as a share of shipments, pooled across post years, on the same balanced set. Contract work falls (-0.10 per standard deviation of reliance, $p < 0.001$). Materials, production hours, and capital expenditure move in the direction an in-house replacement would predict, though none is individually distinguishable from zero, and employment per dollar shipped rises (0.009 , $p = 0.07$). The matched cost items end in 2022, before most of the reliance decline. The contractor side is visible for longer. In the administrative employment series, machine shops, heat treating, and metal coating were 4 to 15 percent below their 2019 employment in 2025, and apparel contractors and printing support 30 to 40 percent below, against 1.5 percent for manufacturing as a whole. Reliant industries walked away from contract production at surviving plants; whether they rebuilt the capability in house is a firm-level question.

5.4 Interpreting the persistence

Why should a severed arrangement stay severed? Acemoglu and Tahbaz-Salehi (2025) provide one answer. In their model a supplier–customer relationship dissolves when a shock pushes its private value below its fixed cost, and it is not re-formed once the shock passes, because the parties do not internalize the surplus the link creates for the rest of the network; the post-shock equilibrium network is therefore a strict subnetwork of the efficient one. An entry shortfall that outlasts the disruption is consistent with that externality, and if the externality is the mechanism the persistence is a welfare loss and not only a reallocation, since the links that fail to re-form would have been worth re-forming; that reading is conditional on the externality, which we do not test. It is equally consistent with the endogenous reorganization of Kopytov et al. (2024), in which firms facing higher perceived disruption risk choose safer, less productive supply networks, with no externality required. The industry data do not separate the two.

One number from Section 5.3 bears on the choice. Across all 358 industries, the correlation between an industry’s contribution to the 2020 coefficient and its percentage change in contract work is 0.03. The pull-back is not where the loss was. Under re-optimization toward safer networks (Kopytov et al., 2024), the industries that learned most about their exposure should have moved most; under a severed link that is not re-formed (Acemoglu and Tahbaz-Salehi, 2025), the pull-back follows the failure of the relationship rather than the size of the loss and need not line up with it. The pattern is consistent with the second reading and harder to square with the first; it tests

neither. The industry data observe establishments, not the links themselves, and Section 5.2 sets out how much weight the entry margin can bear. Inside three-digit sectors the 2018–2023 gradient holds with one measure of customer exposure and not with the other, so the within-sector evidence on persistence is suggestive (Section 4.4).

6 Evidence from Europe and Japan

Contract-production reliance is not an American accounting peculiarity. Six other statistical systems collect a version of the same concept, and this section shows that the measure is portable and that its output gradient is not an artifact of any one country’s conventions. The first claim does not depend on the second, and the second is evidence no single country’s microdata could supply.

Portability is the harder of the two, since the systems disagree about almost everything a measurement convention can disagree about. The Census item is establishment-level and denominated in shipments. The European measure is enterprise-based and therefore *nets out* tolling between two establishments of the same firm, exactly the component the U.S. and Japanese establishment-based measures retain, and we divide it by production value, the closest European analog of shipments. Its legal definition is broader still, pooling furnished-materials tolling with design-specification subcontracting. Japan’s item sits closest to the U.S. concept but is published only for larger establishments. Levels consequently differ by an order of magnitude across systems and are not comparable. Rankings are another matter, and where two conventions can be checked against each other on the same industries they correspond: across German four-digit industries the harmonized European variable and the strict furnished-materials national item, both measured before 2020, rank industries at Spearman 0.96, although in Germany the two appear to share a survey source (Appendix C.4). What the concept picks out survives the change of instrument even though its magnitude does not.

That portability is what makes the second claim testable. A gradient estimated from the Census item alone could reflect how U.S. manufacturers book tolling arrangements rather than what happened to their production, and finer U.S. data would not settle it, since the booking convention is common to the whole frame. Re-estimation where the accounting differs does settle it. An artifact of one collection convention should not survive a switch to the other.

Such re-estimation is possible because the design requires only a pre-2020 cross-section of subcontracting intensity and a real production series at fine industry detail, and both exist abroad. For Europe, the treatment is payments to subcontractors over production value from the harmonized Structural Business Statistics, excluding repair and installation (NACE 33), which lies outside U.S. manufacturing, and the outcome is the industrial production index, a volume measure, for the five countries that publish it at four-digit detail (Germany, Italy, France, Spain, Greece); the pooled specification carries country-by-industry and country-by-year fixed effects, so each country’s aggregate pandemic path is differenced out and β is identified from within-country, cross-industry variation. For Japan, the final Census of Manufactures separately reports the cost of consigned production on furnished materials, the closest foreign analog of the Census contract-work item. The outcome is the METI production index over 75 hand-concorded industry cells.²⁵ In every case the treatment is standardized to one standard deviation within country, so only the within-country ranking of industries is used; Appendix C details the sources, their conceptual and collection

²⁵The European treatment is the 2017 SBS wave at NACE four-digit; the Japanese treatment is the 2019 Census of Manufactures at JSIC four-digit. Appendix C.4 explains the production-value denominator and the exclusion of repair, and reports every estimate under the earlier value-added definition as well.

differences, and what the fixed-effects structure does and does not absorb.

Figure 4 and Table 6 report the acute-window estimate (the 2018–2020 sample, post equal to 2020), which isolates the synchronized phase of the shock, for each economy, alongside the U.S. estimate from the same window. All eight point estimates (the United States, Japan, the pooled European panel, and the five European countries within it) are negative. The pooled European panel gives about -3.6 percent per standard deviation and Japan -2.8 percent, against the U.S. -2.2 percent; Italy (-7.5 percent) and Greece (-4.4 percent) are individually significant, while Germany, Spain, and France are negative but imprecise. The event studies share the U.S. shape at the shock: a flat pre-period, then a sharp 2020 break of about -3.1 percent in pooled Europe and -2.4 percent in Japan, against the U.S. -2.2 percent on the same window (Appendix Figures A1–A2). The European pre-period is longer than the 2018 base of the pooled figure shows. Extending the production series back to 2012, reliant industries were already declining relative to others in most countries before 2019: joint tests of the pre-2019 leads reject in fourteen of sixteen countries at the division level. Removing each industry’s own 2012–2019 trend leaves the 2020 coefficient negative in fifteen of sixteen countries (median -0.034) and significant at the 10 percent level in nine, and at four-digit detail in Italy (-0.064) and Greece (-0.018), while Germany, France, and Spain are small and imprecise at that resolution with or without the trend; what detrending removes is part of the later gap (Appendix Figure A1).

Germany allows one further check. Its cost-structure survey collects the U.S. concept itself, contract work on the firm’s own materials as a share of gross production value, so the German leg can be estimated on the U.S. definition as well as on the harmonized one. On the same 168 industries, the national tolling treatment gives -2.7 percent in 2020 ($p = 0.085$) and -2.3 percent through 2025 ($p = 0.10$), against -2.3 and -1.6 percent with the harmonized treatment. The German estimate is larger and closer to significance on the U.S. definition, which is the direction that measurement error in the broader concept predicts (Appendix C.4).

Where the economies differ is afterward, and much of that difference is a property of the instruments rather than of the economies. Both foreign outcomes are volume indexes built on value weights, the European series on base-year weights and the Japanese on value-added weights. That is the construction of the Federal Reserve’s production index, and Section 5.1 established what it does to U.S. data of known content: aggregated on value weights, the U.S. series reproduces the acute 2020 response and loses the persistence, while the same shipments re-aggregated over the same groups without weights retain it (Appendix A). A U.S.-sized persistent gap abroad would therefore be largely invisible to the series available abroad, which makes the foreign non-persistence weak evidence that no such gap opened. A second part of the difference is trend. The European series carry pre-2019 relative declines that the U.S. six-digit panel shows only faintly, and removing them shrinks the post-2021 tails further. The foreign full-window estimates therefore combine a value-weighted outcome with a differential trend, and neither favors persistence. Full-window estimates abroad are negative but attenuated (pooled Europe about -2.1 percent through 2025, $p = 0.075$; Japan -0.6 percent, not significant), and once industry trends are removed, smaller still. The acute window is therefore the comparable object, and the country-level full-window estimates, one of which turns mildly positive as recovery overshoots the dip, are reported in Appendix C.²⁶

Employment can be compared as well, on a wider panel. Eurostat’s business statistics report persons employed at four-digit detail on both sides of the 2021 change in their framework, which

²⁶As in the U.S. panel, the published clusters are the conservative sector-level bound: clustered at the country-industry unit at which each treatment varies, the pooled acute estimate has $p < 0.001$, Japan’s $p = 0.007$, and the pooled 2018–2024 estimate moves from $p = 0.080$ to $p = 0.036$, while Germany, Spain, and France remain individually imprecise at every clustering level (Appendix C).

extends the comparison to 23 countries with the same treatment (Appendix C.5). After a pre-period screen that removes industries whose employment moved by more than half between 2018 and 2019, employment in more reliant industries fell 1.2 percent more per standard deviation in 2020 and 2.3 percent more over 2018–2024, and both estimates survive removing each country’s pre-2020 reliance trend. On the industries where both outcomes exist, though, employment registers about a third of the output loss in 2020 (-1.2 against -3.6 percent). One reading is that short-time work schemes held matched employment relationships in place over the months in which U.S. establishments let them dissolve. The other is that European headcounts respond to output with a longer lag, which the 2018–2024 estimate, at -2.3 percent, would partly show. The business statistics cannot test the first reading, because their hours series moves one for one with headcount. Scheme take-up by sector is published, so the test is available; we do not run it here. The two readings carry different implications for whether the persistence this paper documents is technological or institutional.

The European result holds across industries but not within them, and textiles and apparel carry much of it. Two checks show this, and we run both on the U.S. panel as well (Appendix Table A12). Adding country-by-division-by-year fixed effects, so that only four-digit industries inside the same two-digit division of the same country are compared, removes the European result: the production estimate becomes $+1.3$ percent (standard error 1.8) and the employment estimate -0.4 percent (0.4), with standard errors little changed. Textiles, apparel, and leather account for about three-fifths of the European employment estimate. Without them the employment estimate falls to -0.6 percent and is insignificant (Appendix Table A12), and the 2020 production estimate falls to -2.8 percent ($p = 0.007$) while the full-window production estimate disappears. The U.S. panel differs on the second check and agrees on the first. Textiles, apparel, and leather contribute about a fifth of the U.S. 2020 estimate, and dropping them raises it to -2.5 percent. Within three-digit sectors the U.S. estimate for 2020 alone is -0.8 percent and insignificant; over 2018–2023 it is -1.7 percent, and -2.6 percent without textiles, and it survives one of the two customer-exposure controls of Section 4.3. In both regions, then, the acute loss is a pattern across broad sectors. The within-sector gradient appears only in the U.S. panel after 2020, and even there it is suggestive.

The exercise delivers sign and shape, not magnitudes, and no single country carries it: the pooled European full-window estimate leans on Greece and Italy, although the acute estimate is -3.5 percent without Greece, and Japan’s acute estimate rests on the pooled fine-cell variation (Appendix C reports these sensitivities). What survives them is the sign pattern across seven statistical systems that collect subcontracting on different units of observation and against different denominators. The estimate is still identified from one shock; what it is no longer identified from is one country’s way of recording the margin.

7 Normal Times and Other Shocks

Why should a margin that is only about 1.6 percent of cost move output by more than its own share, and only under some shocks? The share bounds the input’s marginal product, not the output a disrupted step gates, and what the step costs when it fails depends on how synchronized the disruption is. This section argues both halves: contract work is an ordinary input in calm times, and a simple framework of correlated supplier risk says when reliance raises or lowers an industry’s output response to a shock.

7.1 Production-function estimates

Is contract work simply an unproductive input, so that contract-reliant industries pay a steady-state penalty for using it? The elasticity says no, and the cost share says the input is a minor one. Its cost share averages about 1.6 percent of total cost (Section 2), and its output elasticity is small in absolute terms: a gross-output production function that treats contract work as an endogenous input puts the contract-work elasticity at 0.05 under both OLS and the control-function GMM, collapsing to a noisy 0.007 under within-industry fixed effects, as the thin within variation would predict (Table 2).²⁷

That 0.05, while small in absolute terms, runs about three times the 1.6 percent cost share.²⁸ A direct test points the same way: regressing R&D on contract reliance is null across NSF R&D expenditure, the non-production share, and non-production employment (on NSF expenditure, a coefficient of 0.29 with a standard error of 5.7), so the data show none of the capability atrophy a hollowing-out account would predict (Pisano and Shih, 2009), though the intervals are wide enough that they do not exclude it. Contract work is an ordinary input when conditions are normal. The cost identified in Section 4 is therefore contingent on disruption, not a property of the input’s marginal product.

7.2 A framework of correlated supplier risk

Whether the domestic-contracting margin is fragile is not obvious a priori: it depends on how correlated supplier disruptions are. This section sets out a stylized framework for when contract reliance raises or lowers an industry’s output response to a shock. It is an interpretive device, not a structural model: its key parameter is not measured, and no estimate elsewhere in the paper is derived from it. It adapts the relocation-option logic of BFH, in which outsourcing is valuable because assembly can be shifted across locations, and augments it with a common supply factor that disables that option. Their empirical counterpart is an incidence result that bears directly on this paper: exercising that margin loads volatility onto the contractor side.²⁹ The economic content is an asymmetry in exposure to that common factor between in-house and contract production (Assumption 1); the resilience consequences turn on how synchronized the disruption is.

Mechanisms behind the asymmetry. The inequality $\rho > \rho_h$ is a reduced-form stand-in for several micro-mechanisms, each of which raises the contracted loss relative to the in-house loss under a common shock: (i) an outright contractor shutdown; (ii) scarce capacity rationed toward the contractor’s priority customers; (iii) no in-house fallback for a step tolled precisely because the capability sits outside; (iv) a visibility lag through an unobserved supplier; (v) the freight and requalification costs of moving work in process; and (vi) dissolution of the relationship itself.

The data cut against shutdown. Aggregate receipts of the establishments that sell contract

²⁷Estimated by the control-function approach of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Akerberg et al. (2015), as a two-step estimator with a Levinsohn–Petrin proxy; the one-step GMM form of Wooldridge (2009) is degenerate on this panel (Table 2 note).

²⁸Reading a gap between an output elasticity and a cost share as a wedge is the standard cost-minimization diagnostic and inherits its assumption. The comparison is also asymmetric in measurement: the cost share is observed directly, while β_c is a lower bound because contract work has no dedicated deflator and is deflated by the materials index (Table 2). Both features leave the wedge understated rather than manufactured, but neither makes it an estimated parameter.

²⁹The Mexico/U.S. standard-deviation ratios are 2.03 on the production-worker wage bill and 2.21 on employment, against 1.53 and 0.77 for aggregate manufacturing.

manufacturing rose from \$41 billion in 2017 to \$46 billion in 2022, and in 2020 those industries lost about 5 percent of output against 9.7 percent for manufacturing as a whole (Appendix B). The buyer-side loss also opens at reopening, later than an outage would put it (Section 4.2). That evidence favors rationing and a missing in-house fallback over shutdown, mechanisms that the theory of relational contracts formalizes (Baker et al., 2002). The buyer retains title to the materials, so the residual control right of the property-rights theory is not at issue. What is at issue is a non-contractible priority over the contractor’s scarce capacity, which a self-enforcing agreement sustains in normal times and a large aggregate shock can break. The last mechanism needs no asymmetry in exposure: in Acemoglu and Tahbaz-Salehi (2025) a relationship-specific link with a maintenance cost dissolves once a shock pushes its value below that cost, and a step performed in house has no such margin to sever. Assumption 1 is consistent with either reading, and the framework uses only its sign predictions.

Assumption 1 (Control–flexibility asymmetry). In-house production is under the producer’s direct control: facing a common, industry-wide disruption it can re-sequence and prioritize, losing a fraction ρ_h of in-house output. Contract production is spread over $N \geq 2$ substitutable contractors, so an idiosyncratic outage at one is re-sourced to the others. A common disruption removes that option at all of them at once, and a contractor allocates whatever capacity it retains across its own book of customers, where this buyer’s schedule is one claim among many, costing a fraction $\rho > \rho_h$ that can be neither re-sourced nor backfilled in house. The assumption does not require the contractor to stop producing. Its content is that contractors are *more* exposed to the common factor than in-house production, not that in-house is immune to it.

Let a producer contract a share c of its tasks and perform $1 - c$ in house, with c pre-determined when a disruption strikes. A common factor strikes with probability q —the synchronization of supplier shocks—while independently each contractor suffers an idiosyncratic outage with probability $\pi < 1$, and in-house production loses an expected fraction θ to its own idiosyncratic failure. A contracted task and all $N - 1$ alternates are down together only with probability π^N , so expected contracted loss per unit is $q\rho + (1 - q)\pi^N$: substitutable contractors diversify the idiosyncratic outages, but the synchronized component leaves a floor $q\rho$ that market thickness cannot lower. That floor is the resilience counterpart of the failure of idiosyncratic shocks to wash out in production networks (Acemoglu et al., 2012), and Section 7.3 tests the null it implies. Combining the two bundles, the marginal value of reliance is

$$\frac{d\mathbb{E}[Y]}{dc} = \theta - (1 - q)\pi^N - q(\rho - \rho_h) \xrightarrow{N \rightarrow \infty} \theta - q(\rho - \rho_h). \quad (3)$$

The first term on the right, θ , is the in-house idiosyncratic risk a contracted task escapes. The second term is the diversification benefit of a thick contractor market, and it vanishes as N grows. The third term, $q(\rho - \rho_h)$, is the contractor’s *excess* exposure to the common factor, and market thickness cannot lower it. In a thick market the expression falls in the synchronization q and changes sign at $q^* = \theta/(\rho - \rho_h)$. That threshold lies in $(0, 1)$ whenever the in-house idiosyncratic loss is smaller than the contractor’s excess common-factor loss, which is the case of interest. Routing a task to a contractor trades plant-specific risk for a share of the economy-wide risk, and the trade is favorable when disruptions are weakly synchronized.

The re-sourcing option is the relocation option of BFH, here rendered inert by the common factor, and the fragility at $q \rightarrow 1$ is the efficient-yet-fragile network of Elliott et al. (2022). What the framework adds to either is Assumption 1. The mapping to the estimated shocks follows: a broad, synchronized collapse (q near one) predicts a negative contract-reliance coefficient, the COVID

case. Diffuse pressure that realizes little synchronized loss predicts roughly zero, the GSCPI case. A demand-driven contraction carries no common *supply* factor at all, so it maps to $q = 0$, where the expression is non-negative in a thick market. The 2008–09 case is therefore a placebo for the proposition that not every downturn loads on reliance, and not a third point on the synchronization ordering.

Section 4 estimates the COVID case; the other two shocks are read through the same lens, as we turn to next. The data line up with the ordering, and they permit one out-of-sample check of the sign, a low-synchronization shock that should turn the coefficient non-negative. That is the 2011 Tōhoku episode, reported with its limited power in Section 7.3. Complementarity between in-house and contracted tasks would sharpen the prediction, the loss rising with reliance, and the quartile point estimates do rise with reliance, though their curvature is not separately identified (Appendix A).

7.3 Market thickness and other shocks

Supplier-market thickness. The common floor in equation (3) predicts that a thick contractor market should not buffer the loss under a broad, synchronized shock. It does not, on a test of limited power. Interacting reliance with the pre-determined 2015–2019 mean count of establishments per NAICS-6 (County Business Patterns), the triple interaction on output is null (-0.006 , SE 0.005), and the effect is -0.025 in thin against -0.027 in thick markets. The null is the framework’s prediction, since thickness insures against idiosyncratic supplier risk, which a synchronized shock does not engage; but the minimum detectable moderating effect is about 0.014 per standard deviation of log establishments, so the test separates the prediction from a moderating effect of modest size only weakly. For employment the pattern is different and the framework does not predict it: the triple interaction is -0.012 ($p = 0.058$; wild $p = 0.041$), and the split gives -0.013 in thin markets (insignificant) against -0.037 in thick ones ($p = 0.004$). Thick markets did worse on employment, which no term in equation (3) delivers.

Diffuse pressure and the 2008–09 recession. Is the cost a general property of contract reliance or specific to COVID? Two tests on the same panel point to the latter. Replacing the post-2020 dummy with the annual Global Supply Chain Pressure Index (Federal Reserve Bank of New York, 2025), interacted with pre-shock reliance, yields no negative relationship: -0.002 (SE 0.005) on the full window and 0.006 (SE 0.003) on the pre-COVID window that isolates the 2008–09 crisis, the 2011 Tōhoku earthquake, and the 2017–18 pressures, the non-negative sign the framework gives diffuse pressure; the 2008–09 recession, a demand shock, if anything moved reliant industries the other way (0.036 , SE 0.023; its pre-trends are not clean). Neither test is powerful. The minimum detectable effect is about 0.014 per standard deviation of the index for the pressure interaction and about 0.063 for the 2008–09 contrast, larger than the COVID estimate, and an acute, economy-wide collapse and a smooth pressure index are different objects in any case. The design therefore detects no comparable loss under these disruptions, but it is underpowered to rule one out; the nulls are consistent with the ordering without testing it. The null also disciplines the persistence: measured pressure had normalized by 2023 while the output gap and establishment losses had not, so the persistent component tracks the establishment scarring of Section 5.1, not ongoing disruption.

Two orderings. The sign pattern is what Section 7.2 implies: negative under a broad synchronized collapse, muted under diffuse pressure, non-negative under a demand shock. A second, independent

ordering yields much the same ranking. In Acemoglu and Tahbaz-Salehi (2025) what governs the network’s response is the size of the shock: a small one dissolves only the least valuable links and is barely amplified, while a large one reaches progressively more productive relationships and is amplified sharply. The 2020 collapse is both the most synchronized and the largest supply shock in the panel, and the GSCPI episodes are neither. Accordingly, the pattern does not depend on which ordering is correct. Whether a large but unsynchronized disruption, of which the panel contains none, would separate the two orderings remains open. For a targeted, weakly synchronized shock the framework predicts a non-negative coefficient. The panel contains two such cases, Tōhoku in 2011 and the 2025 tariff wave, and in neither do we detect a penalty.

Tōhoku 2011. The earthquake is the one dated check of the sharper sign-change prediction: a brief, Japan-specific supply shock whose propagation through Japanese input-output links Carvalho et al. (2021) document. On the 188 industries observable over 2008–2011 the contract-reliance coefficient is 0.016 ($p = 0.19$ – 0.26 across the clustering ladder), against -0.024 for the same industries under COVID. The test is underpowered for the sign itself: the minimum detectable effect is about 0.038, half again the COVID magnitude. What the interval supports is a difference between the two shocks. The 95% confidence interval, $[-0.011, 0.042]$, excludes the COVID point, and treating the two estimates as independent the difference is 0.040 with a standard error near 0.014. The two are estimated on the same industries over non-overlapping periods, so that standard error is approximate. Appendix A reports the construction.

The 2025 tariff wave supplies a fourth case, a targeted shock in the panel’s most recent year. In the administrative employment series the reliance gradient in 2025 differs from its 2024 value by -0.002 ($p = 0.49$; minimum detectable difference 0.007), with or without controls for the change in each industry’s own-product and input tariffs, and detrending does not change it. Reliance correlates 0.30 with the 2025 own-product tariff increase, so the tariffs landed disproportionately on reliant industries without adding a penalty to them. This is the most precise of the four comparison shocks, and it sits on the sign the framework gives a targeted disruption.

8 Incidence and Policy

Section 4.6 sized the margin against the other exposures that competed to explain the 2020 collapse. The natural next question is what the gradient comes to in output, and this design does not support that translation: the coefficient is an equal-weighted-industry estimand, and Appendix A shows the acute and pooled responses do not carry over to the dollar-weighted one. The arithmetic is set out there for a reader who wants the order of magnitude; nothing in this section rests on it.

What the design does support is the relative gap. It nets out anything common to all industries and cannot see demand that reallocated across product lines, so it bounds the economy-wide loss from above; to that extent it measures which industries, establishments, and places absorbed the shock. The one offset the data do rule out is the most direct. The gap was not filled from abroad: apparent consumption of these products fell roughly one-for-one with output (Section 4.7). Within these product markets, therefore, the shortfall is real. Which workers and places bear the transition cost is an incidence question the industry data leave open.

Where the vulnerability sits and where industrial policy is aimed do not obviously line up (the cell mapping is descriptive; Appendix A). The low-R&D, low-export-control set holds 31 percent of manufacturing shipments and 32 percent of employment, clusters in Southern and Plains

states—Alabama, Arkansas, Mississippi, and Georgia hold 40–44 percent of their manufacturing employment in it, against 24–31 percent in Massachusetts, New Hampshire, and California—and overlaps the regions most exposed to the 2001–2011 China trade shock (Autor et al., 2013). It is largely disjoint from the recipients of CHIPS Act awards and Section 232 protection: about 4 percent of the set’s employment lies in a covered industry, and none in the semiconductor industries CHIPS targets. Across states its employment share is negatively correlated with exposure to the two programs together (Pearson $r \approx -0.5$), strongly negative for CHIPS ($r \approx -0.65$) and mildly *positive* for Section 232 alone ($r \approx 0.12$); part of the CHIPS gap is mechanical, since the set is defined by low R&D while CHIPS exposure is proxied by the R&D-intensive semiconductor industry. The one industry in both maps is steel.³⁰

Two policy implications follow. First, the import lever is poorly aimed at this margin. That conclusion rests on the domesticity evidence—the near-orthogonality to import penetration and the direct-measurement and mirror bounds of Section 3.1—rather than on the weakly identified substitution test of Section 3.2, which only corroborates it. The one import restriction that reaches part of the vulnerable set is Section 232, through steel (the mildly positive $r \approx 0.12$ above); but it does so by protecting a domestic commodity industry, not by reversing an offshoring that the orthogonality evidence says did not occur, so it is a narrow exception. Recent evidence sharpens the point: the 2018–2019 tariffs, the most direct use of the import lever, raised input costs and lowered output, employment, and exports among the globally connected U.S. manufacturers they reached (Flaaen and Pierce, 2024; Handley et al., 2025), so the instrument is poorly aimed at the contract-production margin and costly where it lands. And the lever that carries most reshoring spending is not an import restriction at all: whether subsidy- or procurement-based industrial policy could rebuild domestic contracting capacity is a separate question this design does not address.

Second, the industries that bore the output cost are low-R&D, geographically concentrated, and largely outside current industrial-policy support, so the vulnerability map and the policy map do not align. The evidence identifies where the vulnerability sits, and it narrows the instrument menu. Inventory is not the fix, since reliant industries held more of it before the shock and lost output anyway (Section 4.3). Dual sourcing is not either: the thickness test of Section 7.3 finds no buffer from a thicker contractor market, which is what the framework predicts when the disruption is common to all suppliers. What remains is capacity reservation, a claim on a contractor’s output under stress that a relational contract cannot enforce on its own; its cost and feasibility are not measured here.

The incidence within these industries is also unresolved here. The output loss could fall on the firms that contract production out, on the contractors performing the work (whose own exposure an industry aggregate does not reveal) or on workers through the fissured-employment margin that domestic outsourcing runs along (Weil, 2014; Goldschmidt and Schmieder, 2017). Separating them is a firm- and worker-level question, and it bears on whether resilience policy should target buyers, suppliers, or places.

A final qualification bounds the policy reading. The result is identified from a broadly synchronized shock, the $q \rightarrow 1$ case of Section 7.2, while the disruptions that motivate reshoring are mostly the opposite kind: a decoupling from one supplier country, a single-strait contingency, an embargo on one input. Those are low- q , targeted shocks, where the diversification logic runs the other way,

³⁰Cell shares are the NAICS-6 cell assignment of Appendix A apportioned to states through the County Business Patterns NAICS-3 employment distribution. Section 232 coverage is the steel and aluminum NAICS-6 industries; CHIPS exposure is proxied by semiconductor manufacturing (NAICS 3344), since award dollars are not yet realized as output. The largest announced fabrication awards (Arizona, Ohio, New York, Texas) sit in low-vulnerability states, so an award-weighted map would, if anything, widen the gap.

and about them the design is silent in either direction. A policy that treats all supply disruption alike may misread one case or the other; which class of disruption an instrument is meant to address is the prior question.

9 Conclusion

We began by asking whether the import lever that dominates resilience policy reaches the margin through which disruption cut domestic output in 2020. It largely does not. Reliance on outside contractors for production steps is predominantly domestic and nearly orthogonal to import penetration, and in normal times it is an ordinary, minor input. When COVID-19 hit, industries one standard deviation more reliant lost about 2.4 percent more output over 2020–2023, and 1.35 percent in 2020 once their customers’ own shocks and steel tariffs are controlled for. The loss was in physical volume. It persisted in employment through 2024, the real output gap persists only under exact parallel trends, and reliant industries then cut contract work at surviving plants. Contracting a step out diversifies idiosyncratic supplier risk while loading on a common supply factor, so the kind of shock matters. Most of the disruptions that motivate reshoring are targeted, and about those the design says little: the one targeted shock in the panel, the 2025 tariff wave, adds no penalty to reliant industries, on a single year. Exposure measures used in resilience assessments could carry the Census contract-work item alongside import penetration. It is already collected, and on the 2020 shock it moved output about as much.

Three boundaries frame these results. The estimates come from a single shock, though no longer from a single economy. They cover one organizational margin, tolling and commission work on owned materials, which is narrower than the make-or-buy boundary. And they sit at one level of aggregation: the industry design identifies the output response of the contract-reliant margin but cannot separate it from the fragility of the standardized production that tends to be routed to contractors, and the acute loss runs mostly across sectors in both the United States and Europe. It remains to be seen whether the organizational friction of the contracting relationship or the character of the production routed through it carries the loss.

We conclude by mentioning two directions in which the results could be extended. The first is the establishment-level design named in the introduction, which would compare plants with different contract-work shares inside the same industry and year. The second is the European employment side. European headcounts registered about a third of the 2020 output loss in reliant industries, short-time work schemes are the natural explanation, and their take-up by sector is published, so whether they absorbed the scarring is a test the data now allow.

References

- Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138(1):1–35, 2023.
- Daron Acemoglu and Alireza Tahbaz-Salehi. The macroeconomics of supply chain disruptions. *Review of Economic Studies*, 92(2):656–695, 2025. doi: 10.1093/restud/rdae038.
- Daron Acemoglu, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.
- Daniel A. Ackerberg, Kevin Caves, and Garth Frazer. Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451, 2015.
- George Alessandria, Shafaat Yar Khan, Armen Khederlarian, Carter Mix, and Kim J. Ruhl. The aggregate effects of global and local supply chain disruptions: 2020–2022. *Journal of International Economics*, 146:103788, 2023. doi: 10.1016/j.jinteco.2023.103788.
- Laura Alfaro and Davin Chor. Global supply chains: The looming “great reallocation”. Working Paper 31661, NBER, 2023. Prepared for the Jackson Hole Economic Policy Symposium.
- Laura Alfaro, Pol Antràs, Davin Chor, and Paola Conconi. Internalizing global value chains: A firm-level analysis. *Journal of Political Economy*, 127(2):508–559, 2019.
- Theodore W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20(1):46–63, 1949.
- Pol Antràs and Davin Chor. Organizing the global value chain. *Econometrica*, 81(6):2127–2204, 2013.
- Pol Antràs and Elhanan Helpman. Global sourcing. *Journal of Political Economy*, 112(3):552–580, 2004.
- Pol Antràs, Davin Chor, Thibault Fally, and Russell Hillberry. Measuring the upstreamness of production and trade flows. *American Economic Review: Papers and Proceedings*, 102(3):412–416, 2012.
- Peter M. Aronow and Cyrus Samii. Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1):250–267, 2016.
- Enghin Atalay, Ali Hortaçsu, and Chad Syverson. Vertical integration and input flows. *American Economic Review*, 104(4):1120–1148, 2014.
- David H. Autor, David Dorn, and Gordon H. Hanson. The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6):2121–2168, 2013.
- George Baker, Robert Gibbons, and Kevin J. Murphy. Relational contracts and the theory of the firm. *Quarterly Journal of Economics*, 117(1):39–84, 2002. doi: 10.1162/003355302753399445.
- Richard Baldwin and Rebecca Freeman. Risks and global supply chains: What we know and what we need to know. *Annual Review of Economics*, 14:153–180, 2022.

- David Rezza Baqaee and Emmanuel Farhi. The macroeconomic impact of microeconomic shocks: Beyond Hulten’s theorem. *Econometrica*, 87(4):1155–1203, 2019.
- Jean-Noël Barrot and Julien Sauvagnat. Input specificity and the propagation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, 131(3):1543–1592, 2016.
- Paul R. Bergin, Robert C. Feenstra, and Gordon H. Hanson. Offshoring and volatility: Evidence from Mexico’s maquiladora industry. *American Economic Review*, 99(4):1664–1671, 2009.
- Andrew B. Bernard and Teresa C. Fort. Factoryless goods producing firms. *American Economic Review: Papers and Proceedings*, 105(5):518–523, 2015.
- Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.
- Christoph E. Boehm, Aaron Flaaen, and Nitya Pandalai-Nayar. Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. *Review of Economics and Statistics*, 101(1):60–75, 2019.
- Barthélémy Bonadio, Zhen Huo, Andrei A. Levchenko, and Nitya Pandalai-Nayar. Global supply chains in the pandemic. *Journal of International Economics*, 133:103534, 2021.
- Kirill Borusyak, Peter Hull, and Xavier Jaravel. Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213, 2022.
- Brantly Callaway, Andrew Goodman-Bacon, and Pedro H. C. Sant’Anna. Difference-in-differences with a continuous treatment. Working Paper 32117, NBER, 2024.
- A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427, 2008.
- Vasco M. Carvalho and Alireza Tahbaz-Salehi. Production networks: A primer. *Annual Review of Economics*, 11:635–663, 2019.
- Vasco M. Carvalho, Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi. Supply chain disruptions: Evidence from the Great East Japan Earthquake. *Quarterly Journal of Economics*, 136(2):1255–1321, 2021.
- Juanma Castro-Vincenzi, Adry Gracio, Gaurav Khanna, and Nitya Pandalai-Nayar. Supply chain risk, trade and economic fragility. NBER Working Paper 35496, National Bureau of Economic Research, July 2026.
- Ryan A. Decker and John Haltiwanger. Surging business formation in the pandemic: Causes and consequences? *Brookings Papers on Economic Activity*, pages 249–302, 2023. Fall 2023 issue.
- Julian di Giovanni, Şebnem Kalemli-Özcan, Alvaro Silva, and Muhammed A. Yıldırım. Global supply chain pressures, international trade, and inflation. Working Paper 30240, NBER, 2022.
- Jonathan I. Dingel and Brent Neiman. How many jobs can be done at home? *Journal of Public Economics*, 189:104235, 2020.
- Emin Dinlersoz, Timothy Dunne, John Haltiwanger, and Veronika Penciakova. Business formation: A tale of two recessions. *AEA Papers and Proceedings*, 111:253–257, 2021.

- Matthew Elliott, Benjamin Golub, and Matthew V. Leduc. Supply network formation and fragility. *American Economic Review*, 112(8):2701–2747, 2022.
- Glenn Ellison, Edward L. Glaeser, and William R. Kerr. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3):1195–1213, 2010. doi: 10.1257/aer.100.3.1195.
- Federal Reserve Bank of New York. Global supply chain pressure index (GSCPI). Technical report, Federal Reserve Bank of New York, 2025. Monthly series, 1997–2026; <https://www.newyorkfed.org/research/policy/gscpi>.
- Aaron Flaaen and Justin R. Pierce. Disentangling the effects of the 2018–2019 tariffs on a globally connected u.s. manufacturing sector. *Review of Economics and Statistics*, 2024. forthcoming.
- Silke Januszewski Forbes and Mara Lederman. Adaptation and vertical integration in the airline industry. *American Economic Review*, 99(5):1831–1849, 2009. doi: 10.1257/aer.99.5.1831.
- Teresa C. Fort. Technology and production fragmentation: Domestic versus foreign sourcing. *Review of Economic Studies*, 84(2):650–687, 2017.
- Deborah Goldschmidt and Johannes F. Schmieder. The rise of domestic outsourcing and the evolution of the German wage structure. *Quarterly Journal of Economics*, 132(3):1165–1217, 2017.
- Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020.
- Gene M. Grossman and Elhanan Helpman. Integration versus outsourcing in industry equilibrium. *Quarterly Journal of Economics*, 117(1):85–120, 2002.
- Gene M. Grossman, Elhanan Helpman, and Hugo Lhuillier. Supply chain resilience: Should policy promote international diversification or reshoring? *Journal of Political Economy*, 131(12):3462–3496, 2023.
- Gene M. Grossman, Elhanan Helpman, and Alejandro Sabal. Optimal resilience in multitier supply chains. *Quarterly Journal of Economics*, 139(4):2377–2425, 2024. Earlier version: NBER Working Paper 31739, “Resilience in Vertical Supply Chains”.
- Veronica Guerrieri, Guido Lorenzoni, Ludwig Straub, and Iván Werning. Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? *American Economic Review*, 112(5):1437–1474, 2022.
- Kyle Handley, Fariha Kamal, and Ryan Monarch. Rising import tariffs, falling exports: When modern supply chains meet old-style protectionism. *American Economic Journal: Applied Economics*, 17(1):208–238, 2025.
- Susan Houseman, Christopher Kurz, Paul Lengermann, and Benjamin Mandel. Offshoring bias in U.S. manufacturing. *Journal of Economic Perspectives*, 25(2):111–132, 2011. doi: 10.1257/jep.25.2.111.
- Susan N. Houseman. Outsourcing, offshoring and productivity measurement in United States manufacturing. *International Labour Review*, 146(1-2):61–80, 2007.

- Dmitry Ivanov and Alexandre Dolgui. Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. *International Journal of Production Research*, 58(10):2904–2915, 2020.
- Fariha Kamal. A portrait of u.s. factoryless goods producers. Working Paper CES-18-37, U.S. Census Bureau Center for Economic Studies, 2018.
- Gaurav Khanna, Nicolas Morales, and Nitya Pandalai-Nayar. Supply chain resilience: Evidence from Indian firms. Working Paper 30689, NBER, 2022.
- Alexandr Kopytov, Bineet Mishra, Kristoffer Nimark, and Mathieu Taschereau-Dumouchel. Endogenous production networks under supply chain uncertainty. *Econometrica*, 92(5):1621–1659, 2024. doi: 10.3982/ECTA20629.
- Raphaël Lafrogne-Joussier, Julien Martin, and Isabelle Méjean. Supply shocks in supply chains: Evidence from the early lockdown in china. *IMF Economic Review*, 71(1):170–215, 2023.
- James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341, 2003.
- Rocco Macchiavello and Ameet Morjaria. The value of relationships: Evidence from a supply shock to Kenyan rose exports. *American Economic Review*, 105(9):2911–2945, 2015. doi: 10.1257/aer.20120141.
- James G. MacKinnon and Matthew D. Webb. Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254, 2017.
- Matthias Meier and Eugenio Pinto. COVID-19 supply chain disruptions. *European Economic Review*, 162:104674, 2024.
- Simon Mongey, Laura Pilossoph, and Alexander Weinberg. Which workers bear the burden of social distancing? *Journal of Economic Inequality*, 19(3):509–526, 2021.
- National Center for Science and Engineering Statistics. Business enterprise research and development: 2022. Nsf 24-335, U.S. National Science Foundation, 2024. September 30, 2024.
- G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996.
- Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204, 2019.
- Justin R. Pierce and Peter K. Schott. A concordance between ten-digit u.s. harmonized system codes and sic/naics product classes and industries. *Journal of Economic and Social Measurement*, 37(1-2):61–96, 2012.
- Justin R. Pierce and Peter K. Schott. The surprisingly swift decline of us manufacturing employment. *American Economic Review*, 106(7):1632–1662, 2016.
- Gary P. Pisano and Willy C. Shih. Restoring american competitiveness. *Harvard Business Review*, 87(7/8):114–125, 2009.
- Ashesh Rambachan and Jonathan Roth. A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591, 2023.

- Nátália Ramondo, Veronica Rappoport, and Kim J. Ruhl. Intrafirm trade and vertical fragmentation in U.S. multinational corporations. *Journal of International Economics*, 98:51–59, 2016.
- David Roodman, James G. MacKinnon, Morten Ørregaard Nielsen, and Matthew D. Webb. Fast and wild: Bootstrap inference in Stata using boottest. *Stata Journal*, 19(1):4–60, 2019.
- Yossi Sheffi. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage*. MIT Press, 2005.
- Gary Solon, Steven J. Haider, and Jeffrey M. Wooldridge. What are we weighting for? *Journal of Human Resources*, 50(2):301–316, 2015.
- Brian Tomlin. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, 52(5):639–657, 2006.
- Adam G. Walke and Stephan Weiler. Trade, offshoring, and manufacturing employment volatility in Mexico. *Economic Inquiry*, 64(1):350–370, 2026. doi: 10.1111/ecin.70023. Published online October 2025.
- David Weil. *The Fissured Workplace: Why Work Became So Bad for So Many and What Can Be Done to Improve It*. Harvard University Press, 2014.
- Jeffrey M. Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114, 2009.

Table 1. Variance decomposition of contract-production reliance

Component	Share of variance
Between three-digit sectors	0.31
Between six-digit, within sector	0.50
<i>Total cross-industry</i>	0.81
Within industry, over time	0.19
of which common year effect	0.002

Notes: Nested ANOVA of contract-production reliance on the 2005–2021 base panel (unbalanced). Shares sum to one up to rounding; the balanced sub-panel gives a 0.73 cross-industry share.

Table 2. Production-function elasticities: contract work is an ordinary input

Input elasticity	OLS	Within FE	Control-function GMM
Contract work (β_c)	0.050***	0.007	0.054*
Materials	0.70***	0.54***	0.72***
Capital	0.18***	−0.06	0.21**
Labor	0.07***	0.27***	0.04
Returns to scale	0.99	0.82	1.00

Notes: Gross-output production function on a deflated real-quantity NAICS-6 panel, 2005–2016, applied at the industry level as a prototype, not a firm-level estimator. All three columns estimate the full input vector; an energy elasticity is included in every specification but not shown, so the rows above do not sum to the returns to scale reported beneath them. It is small and insignificant outside the fixed-effects column, and negative under OLS (−0.012) and the control function (−0.024), which is why those two columns sum to slightly more than their returns to scale; under fixed effects it is +0.066 ($p = 0.012$). Returns to scale is the sum of all five input elasticities. The control-function column is the Akerberg et al. (2015) two-step estimator with a Levinsohn–Petrin proxy; standard errors are industry-block bootstrapped. The one-step Wooldridge (2009) GMM version on the same panel is degenerate ($\beta_c = 0.52$, energy −0.87, returns to scale 1.15) and is not used. Materials are net of contract work in every column; with gross materials instead, β_c is 0.029 (OLS) and −0.002 (within FE), so the ordinary-input reading does not depend on the netting. Contract work is deflated by the materials deflator, so β_c is a lower bound on magnitude.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. The COVID shock: two exposure channels, real quantities and margins

Outcome	CPR ^{pre} × post		Foreign ^{pre} × post	
	$\hat{\beta}$ (SE)	p	$\hat{\gamma}$ (SE)	p
<i>Panel A. Real quantities</i>				
log output (shipments)	−0.024 (0.008)	0.002	−0.018 (0.008)	0.024
above-median contrast	−0.064 (0.016)	< 0.001	—	—
log employment	−0.022 (0.007)	0.002	−0.024 (0.006)	< 0.001
log labor productivity	−0.002 (0.004)	0.69	0.005 (0.005)	0.24
<i>Panel B. Margins and factor shares</i>				
operating-surplus share	−0.001 (0.002)	0.44	0.000 (0.002)	0.89
labor-cost share	0.000 (0.001)	0.73	−0.001 (0.001)	0.32
value-added share	−0.002 (0.002)	0.32	−0.001 (0.002)	0.81
<i>Panel C. Timing: joint year-by-year coefficients, log output</i>				
2020	−0.023 (0.006)	—	−0.004 (0.006)	—
2021	−0.031 (0.007)	—	−0.001 (0.008)	—
2022	−0.035 (0.009)	—	−0.035 (0.013)	—
2023	−0.018 (0.011)	—	−0.028 (0.012)	—
2024	−0.020 (0.012)	—	−0.023 (0.016)	—
<i>Panel D. Log output, adding customer and steel-tariff controls in turn</i>				
+ Section 232 input shares	−0.022 (0.008)	0.007	−0.019 (0.008)	0.020
+ co-location customer exposure	−0.021 (0.008)	0.010	−0.021 (0.008)	0.009
+ BEA customer shares	−0.017 (0.008)	0.030	−0.022 (0.008)	0.008

Notes: Each row is a separate estimate of equation (2) on 2018–2023 with NAICS-6 and year fixed effects ($N = 2112$). The continuous treatment is one SD of pre-COVID contract reliance; the foreign channel is one SD of pre-COVID China-input exposure; the above-median contrast replaces the continuous treatment with an indicator for reliance above the industry-level median (its employment analog is -0.057 , $p < 0.001$). Standard errors and p -values are CRV1 clustered at NAICS-6, the primary level (Section 4.5); the conservative NAICS-3 wild cluster bootstrap bounds (Webb weights, null imposed, $B = 9999$) are 0.022 for output, 0.029 for employment, and 0.008 for the above-median contrast, with the full clustering ladder in Appendix Table A1. Panel A shows the cost falls on real volume; Panel B shows it does not run through profitability or factor shares for either channel. Panel C reports the year-by-year coefficients from estimating both channels’ year interactions jointly (base year 2019, CRV1 standard errors at NAICS-6): the reliance loss arrives in 2020–21, when the China-exposure coefficients are zero, and the China-exposure loss arrives only in 2022–24. Panel D adds, cumulatively, the pre-period steel and fabricated-metal input shares, a six-digit measure of co-location with oil and gas, aerospace, and auto employment, and the BEA commodity-level shares of output sold to the same three customer groups, each interacted with post (Section 4.3; the full set is in Table A7). Extending the window to 2018–2024 leaves Panel A unchanged: output -0.023 (NAICS-3 wild bootstrap $p = 0.032$), employment -0.025 ($p = 0.023$), above-median contrast -0.063 ; the productivity, operating-surplus, and labor-share rows stay null, while the value-added share is -0.004 ($p = 0.043$) at baseline and null with controls ($p = 0.92$).

Table 4. The persistent gap in five collection systems

Source	Frame	Outcome	Acute	Latest year
<i>Panel A. Output</i>				
ASM/EC/AIES	survey, NAICS-6	log shipments, nominal	−0.022	−0.025
ASM/EC/AIES	survey, NAICS-6	log shipments, real	−0.035	−0.030
G.17 index	value-weighted, 63 groups	real production	−0.021	0.00
<i>Panel B. Labor</i>				
QCEW	UI universe, NAICS-6	log employment	−0.023	−0.031
CES	payroll survey, NAICS-4	log employment	—	−0.020
ASM/EC/AIES	survey, NAICS-6	log employment	—	−0.035
<i>Panel C. Establishments</i>				
CBP	business register, NAICS-6	log establishments	−0.006	−0.025
BDS	business register, NAICS-4	log entries	—	−0.072

Notes: Coefficients on one standard deviation of pre-COVID contract reliance interacted with the indicated year, from equation (2) with industry and year fixed effects. The six sources span five collection systems (CBP and BDS both derive from the Census Business Register); the systems share no collection machinery with each other, and only the first two rows share it with the survey that supplies the treatment. The acute column is the event-study coefficient for 2020 except in the G.17 row, where it is the pooled June–December estimate, the months in which the monthly gradient opens (Appendix A); the CBP acute cell is insignificant by construction, since the count references March 12 and predates the lockdowns. The final column is the event-study coefficient for the latest year each source reports: 2024 for the survey and QCEW rows, 2023 for the CES, G.17, CBP, and BDS rows. The nominal-shipments row reports the 2021–2023 pooled persistence contrast with 2020 dropped (its 2024 event-study coefficient is -0.019 , $p = 0.059$; the 2019-shipment-weighted variant of the same series gives -0.021 , $p = 0.11$), and the CES row the midpoint of its -0.018 to -0.022 range over 2021–2023. QCEW coefficients run -0.023 to -0.031 raw over 2020–2024 and -0.016 to -0.023 after the industry-specific detrending whose numbers the text quotes. Real shipments are PPI-deflated on the 289 industries with a continuous six-digit series through 2024. BDS is published at NAICS-4, so its standardized magnitude is per four-digit standard deviation and is not comparable in size to the six-digit rows. The G.17 row is the one value-weighted, between-family real object in the table, and the text and Appendix A explain why it is the cut in which no source, ours included, shows persistence. Inference, clustering, and pre-period treatment for each row are in Section 5.1 and Appendix A.

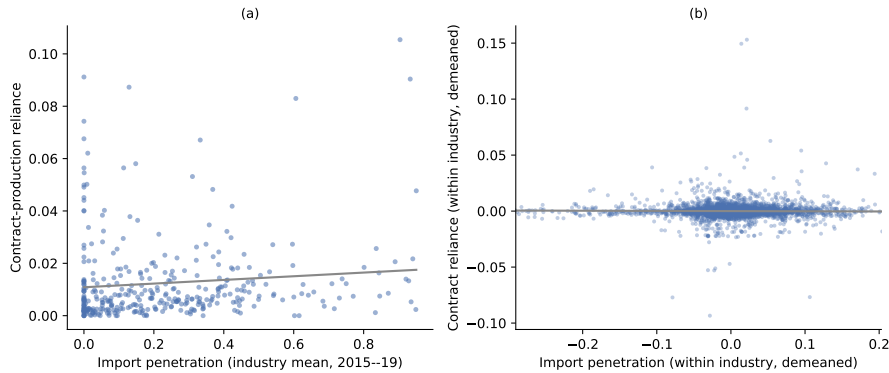


Figure 1. Contract reliance is statistically distinct from imports

Notes: Contract-production reliance against import penetration (imports over shipments plus imports, built from the HS10-to-NAICS6 concordance of Pierce and Schott, 2012). Panel (a) plots industry means for 2015–19; the cross-industry fit has $R^2 = 0.01$ ($r = 0.10$). Panel (b) residualizes both variables on NAICS-6 and year fixed effects, so its slope is the within-industry, over-time relationship, which is flat (within $R^2 < 0.001$, coefficient insignificant). If contract work were relabeled imports, both panels would slope up steeply. A horse race against the foreign-input channel, in which the two enter as separate co-equal coefficients, is in Table 3.

Table 5. Robustness of the output contraction

Specification	$\hat{\beta}$ (CPR ^{pre} × post)	p
Baseline (nominal shipments, full sample)	−0.024	0.022 ^b
<i>Panel A. Composition and confound controls</i>		
+ full composition controls	−0.028	0.035 ^b
+ pre-COVID inventory intensity × post	−0.022	0.019 ^b
+ upstream COVID exposure × post (ADB-MRIO)	−0.021	0.018 ^b
+ embodied import content × post (BEA I–O)	−0.024	0.025 ^b
+ physical proximity × post	−0.024	0.023 ^b
+ R&D intensity and size × post	−0.024	0.020 ^b
+ energy cost share × post (placebo input)	−0.024	0.025 ^b
+ essential-industry (CISA) × post	−0.022	0.033 ^b
+ realized demand-rotation index	−0.023	0.023 ^b
+ final-use composition × year fixed effects	−0.023	0.019 ^b
<i>Panel B. Measurement</i>		
Foreign channel = total import penetration	−0.024	0.035 ^b
Real output (PPI-deflated, common sample) + controls	−0.034	0.022 ^b
<i>Panel C. Fixed-effects structure and trends</i>		
NAICS-3 × year fixed effects	−0.018	0.055 ^r
NAICS-4 × year fixed effects	−0.013	0.225 ^c
Industry pre-trends removed (fit 2015–19)	−0.025	0.051 ^b
<i>Panel D. Inference variants</i>		
Collapsed pre/post cross-section (two periods)	−0.024	0.001 ^h
same, CRV1 at NAICS-3 / NAICS-4	−0.024	0.006 / 0.009
Randomization inference (baseline)	−0.024	0.003
Leave-one-sector-out (each of 21 NAICS-3)	[−0.028, −0.020]	all < 0.05

Notes: Each row is the CPR^{pre} × post coefficient on log output from equation (2), 2018–2023, NAICS-6 and year fixed effects. ^b is a wild cluster bootstrap p (Webb weights, null imposed, $B = 9999$, 21 NAICS-3 clusters), the conservative rung of the clustering ladder; the baseline’s primary NAICS-6-clustered p is 0.002 (Appendix Table A1). The baseline randomization-inference row permutes the cross-sectional treatment across industries. ^r is randomization inference permuting the treatment across industries *within* NAICS-3 sectors ($B = 1999$), the design-based test for the sector-by-year specification, whose NAICS-6, NAICS-4, and NAICS-3-wild p -values are 0.044, 0.062, and 0.21. ^c is CRV1 at NAICS-4 (86 clusters); this specification identifies only from within-four-digit variation, and a split-sample measurement check (Section 4.4) finds its smaller coefficient is mostly a genuinely flatter within-family gradient rather than attenuation. ^h is heteroskedasticity-robust on the 358-industry collapsed cross-section (mean post-2020 minus mean pre-2020 log output), which removes the serial-correlation rationale for clustering (Bertrand et al., 2004) but not the design rationale, since treatment may be correlated within sectors; the clustered versions of the same regression are reported beneath it. The occupation-based and inventory controls are constructed in Appendix A; the others are described in Section 4.3. The real-output row deflates by industry PPI on the 300 covered industries and reports the full-control specification; on that common sample the deflated and nominal coefficients are within a tenth of a log point of each other without controls (−3.5% against −3.6%) and half a point apart with them (−3.4% against −3.9%), so the gap to the −0.024 baseline is the narrower sample and the added controls, not deflation (Appendix A). Dating the post period at 2021 with 2020 dropped, the persistence contrast is −0.025 (wild $p = 0.026$) and stays −0.024 under either demand-rotation control. Appendix A reports the leave-one-industry-out influence diagnostics.

Table 6. The acute COVID output response of contract-production reliance across economies

Economy	Treatment source	Outcome	Cells	Acute 2018–20	
				$\hat{\beta}$ (SE)	p
United States	ASM contract work, 2015–19	shipments	358	−0.022 (0.007)	0.001
EU-5 pooled	SBS subcontracting, 2017	prod. index	690	−0.036 (0.010)	0.0003
Germany	(same)	prod. index	168	−0.022 (0.015)	0.17
Greece	(same)	prod. index	89	−0.044 (0.013)	0.002
Spain	(same)	prod. index	145	−0.037 (0.026)	0.17
France	(same)	prod. index	144	−0.012 (0.012)	0.33
Italy	(same)	prod. index	144	−0.075 (0.019)	0.001
Japan	Census of Mfg. consigned prod., 2019	prod. index	75	−0.028 (0.011)	0.020

Notes: Each row estimates the same design: log real activity on pre-COVID contract-production intensity (standardized to one SD within the economy) interacted with post-2020, industry and year fixed effects, on the acute window 2018–2020 (post = 2020); the full-window estimates are in Appendix Table A9. U.S.: NAICS-6, log nominal shipments, two-channel specification of equation (2), p under the primary NAICS-6 clustering used throughout the paper; CRV1 on 21 NAICS-3 clusters, for comparability with the sector-level clustering of the foreign rows, gives 0.007, and the NAICS-3 wild cluster bootstrap 0.038 (dropping the import channel gives −0.023). EU: treatment is Eurostat SBS payments to subcontractors over production value (NACE four-digit excluding repair and installation, NACE 33; 2017 wave); outcome is the STS industrial production index; the pooled row uses country-by-industry and country-by-year fixed effects, clustered on country-by-NACE-2; country rows use industry and year fixed effects, clustered on NACE-2. Japan: treatment is consigned-production cost over shipments from the final Census of Manufactures (2019, establishments with 30 or more employees); outcome is the METI production index over 75 concorded cells, clustered on JSIC-2. As in the U.S. panel, these sector-level clusters are the conservative bound; the unit-level (NACE-4 and concordance-cell) ladder is reported in Appendix C. Sources, conceptual differences, and construction detail are in Appendix C.

Table 7. Candidate mechanisms against three timing facts

Mechanism	Contractor output, 2020	Buyer loss opens	After 2020	Fits
Contractor shutdown	collapses	at lockdown	recovers	no
Scarce capacity rationed by priority	holds up	at reopening	buyers pull back	yes
No in-house fallback for the step	holds up	at reopening	rebuilt in house	yes
Link dissolved, not re-formed	holds up	at reopening	reliance stays low	yes
Mechanical input share				at most a fifth
Observed	−5% against −9.7% for all manufacturing	absent April–May, widest July–August	reliance −26%; contractor employment −4 to −40%	

Notes: Candidate mechanisms from Section 7.2. Contractor output is the 2020 output change of the establishments that sell contract manufacturing (Appendix B); timing is the monthly production-index event study of Section 4.2; the after-2020 column is Section 5.3. The input-share row is the calibration of Appendix B. Mechanisms are ranked by elimination; none is tested directly.

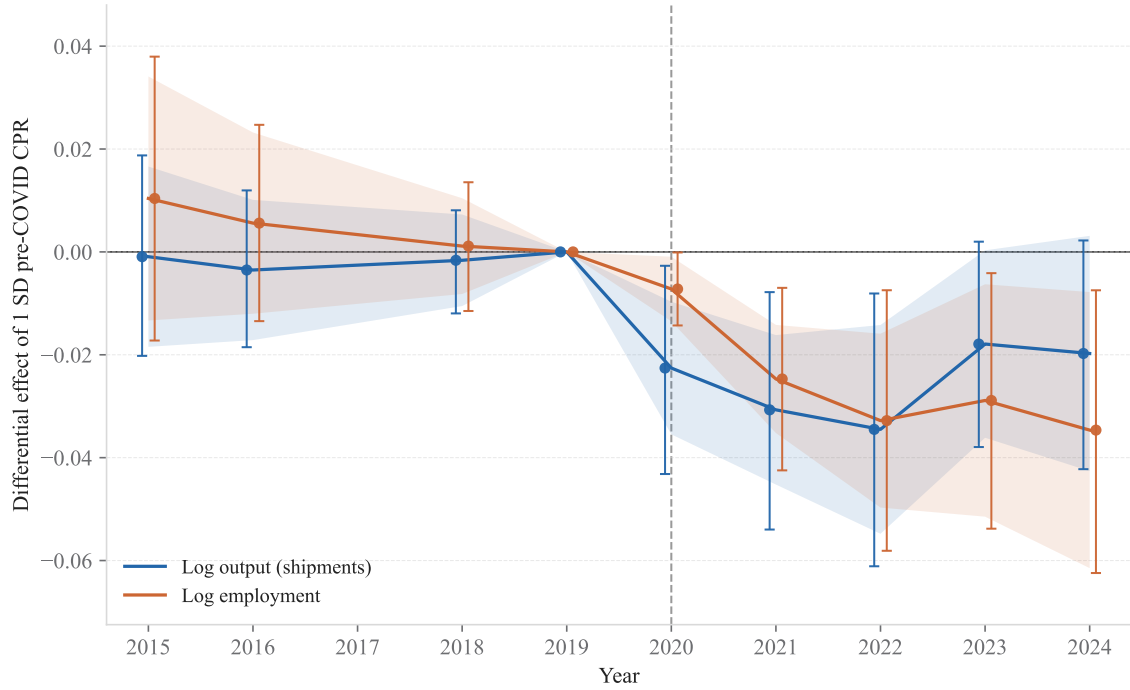


Figure 2. Event study: persistent contraction in contract-reliant industries

Notes: Coefficients on pre-COVID contract reliance (one SD, standardized) interacted with year dummies, base year 2019, window 2015–2024, NAICS-6 and year fixed effects; the event study extends the pre-period of the 2018–2023 estimation window back to 2015 to display longer pre-trends. There is no 2017 observation (a Census of Manufactures year for which contract-work detail is not published), and the plotted series interpolates across it. Shaded bands are CRV1 95% intervals; whiskers are wild cluster bootstrap 95% intervals (Webb weights, null imposed, $B = 9999$, 21 NAICS-3 clusters, the conservative rung of Appendix Table A1). Pre-2020 coefficients are small and, for output, insignificant; the employment series shows a mild pre-2020 drift discussed in Section 5.2.

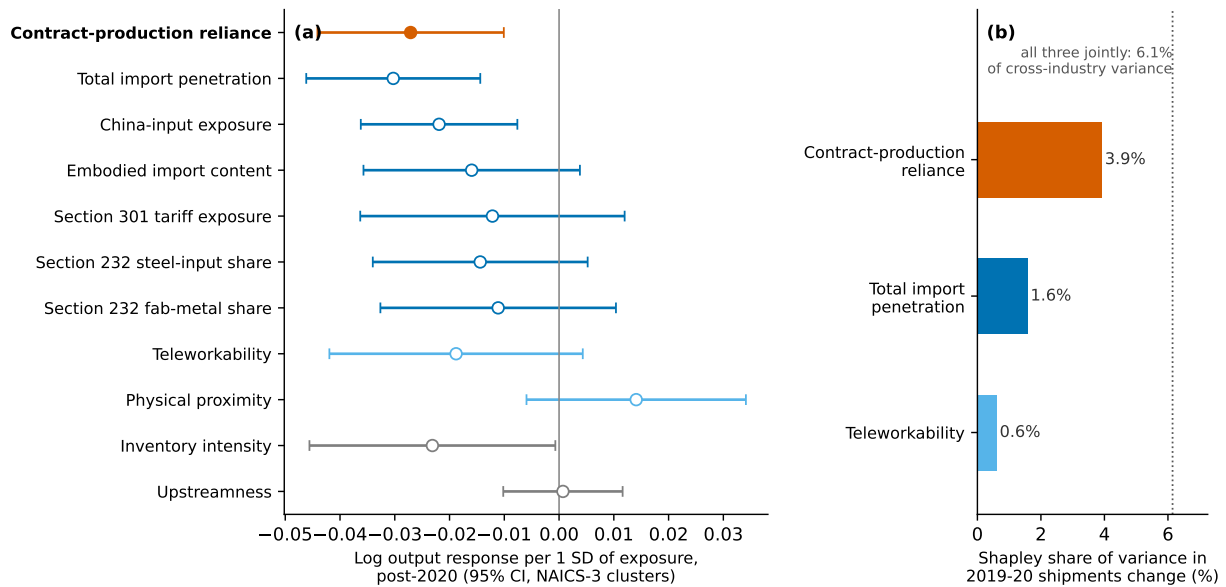


Figure 3. Contract reliance against the measured exposures, on a common footing

Notes: Panel (a): the post-2020 log-output response per one SD of each exposure, entered one at a time in equation (2) on the same panel and window (Section 4.6); whiskers are CRV1 95% intervals on 21 NAICS-3 clusters, the conservative level. Rows are grouped by the kind of exposure—the contracting margin, foreign sourcing, tariff coverage, pandemic operability, and supply-chain structure—not sorted by magnitude. Of the eleven, only contract reliance and the two import measures carry intervals that exclude zero. Panel (b): Shapley shares of the cross-industry variance in the 2019–2020 shipments change for the three exposures entered jointly; the dotted line is their joint explained share. Most of the cross-industry dispersion in 2020 losses is explained by none of the measured exposures, so the shares are shares of that explained portion.

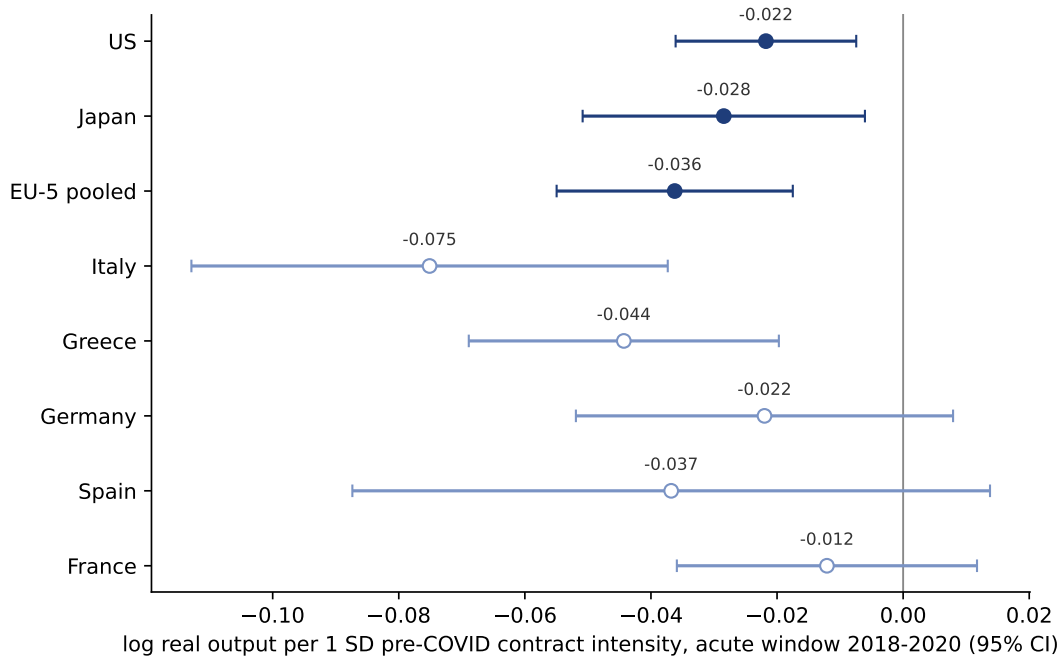


Figure 4. The acute-2020 output response of contract-production reliance, by economy
Notes: Acute-window (2018–2020) coefficients on pre-COVID contract-production intensity (one within-economy SD) interacted with post-2020, with 95% confidence intervals; specifications and sources as in Table 6. Filled markers are the three headline panels (U.S., Japan, pooled EU-5); open markers are the five European countries estimated individually inside the pooled panel. Magnitudes are not comparable across economies because the underlying subcontracting concepts differ in breadth (Appendix C); the exhibit’s content is the sign pattern.

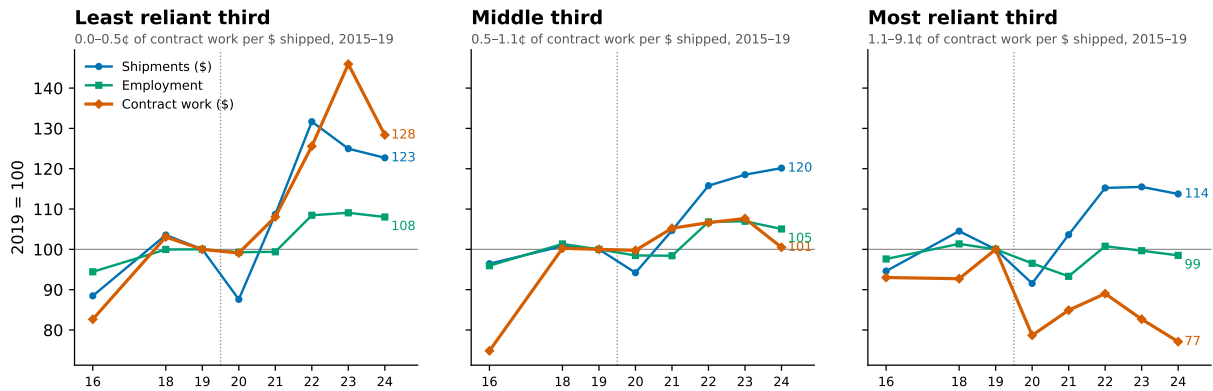


Figure 5. Contract work, shipments, and employment by pre-COVID reliance tercile
Notes: Sums over the 297 industries observed in every year 2016–2024, split into terciles of 2015–19 mean reliance, each series indexed to 2019 = 100. Contract work and shipments are nominal dollars from the Annual Survey of Manufactures (2016–2021), the 2022 Economic Census, and the 2023–24 Annual Integrated Economic Survey; employment from the same sources. The most reliant third ends 2024 with contract work at 77, shipments at 114, and employment at 99.

A Extended Robustness and Inference

This appendix details the checks summarized in Table 5 and Sections 4.3–4.5 and 5.2. The production-function elasticities of Section 7.1 (Table 2) are reported with the other tables, which follow the references and precede this appendix.

The dose-response and the heterogeneity-robust estimand. Splitting pre-COVID reliance into quartiles, the post-2020 output gap relative to the least-reliant quartile is -0.4% in the second quartile, -4.7% in the third, and -8.3% in the most reliant (Figure A3): the contraction rises with the dose and is concentrated in the most contract-reliant industries. The individual quartile magnitudes are imprecise under a cluster bootstrap over the 21 sectors, so we read the dose-response as corroborating the *direction* of the headline effect—more reliance, more loss—rather than as identifying its curvature.

Taking the CGS objects directly, the dose-specific effect relative to the least-reliant quartile and its treated-weighted average, the average output loss of an above-bottom-quartile industry is -0.045 over the full window and -0.027 in the acute window, where it is not individually significant ($p = 0.31$). Both are larger in magnitude than the -0.024 per-standard-deviation slope, because the loss concentrates at the top of the reliance distribution, which is why the linear coefficient is the conservative summary. Resampling whole NAICS-3 sectors, the overall effect is borderline (95% interval $[-0.079, 0.003]$, $p = 0.068$), and the implied average causal response, the slope of the dose-response curve averaged over the treated doses, is sensitive to the dose resolution—the same conservatism under which the curvature is not separately identified.

The clustering ladder and the Ibragimov–Müller rung. What the coarser rungs of Table A1 bound is the correlation of treatment and residuals across industries, and 64 percent of the variance of the pre-COVID reliance *cross-section*—nearly two-thirds—sits inside three-digit sectors.³¹ Ibragimov–Müller estimates the coefficient separately within each sector, so each of its 21 inputs is identified from the same within-sector variation as the pooled estimate; what the procedure gives up is the between-sector variation and, at $G = 21$, precision. The sector coefficients are heterogeneous (Table A2). Their mean is -0.028 , close to the pooled -0.024 , but electrical equipment (NAICS 335, 17 industries) alone is -0.40 , and without it the sector mean is -0.009 ; twelve of the 21 are negative and nine positive, two of the latter from sectors with only four and seven six-digit industries. An equally weighted t -test over that distribution has little power, and that, rather than any feature of the design, is why the rung does not reject. The electrical-equipment slope is also a different object from the pooled one. Reliance in that sector runs from 0.1 to 1.3 cents, all below the panel mean, and the steep slope comes from its least reliant industries (transformers, lighting, wiring devices) growing 20–40 percent after 2020, not from reliant industries falling; it contributes 5 percent of the pooled coefficient, which is -0.023 ($p = 0.002$) without it.

Split-half measurement-error correction. NAICS-4-by-year fixed effects attenuate the coefficient to -0.013 (Table 5), and a split-sample check shows the decline is mostly real rather than statistical. The pre-period reliance measure averages four survey years, so its 2015–16 and 2018–19 halves supply two readings of the same exposure; instrumenting one with the other corrects classical measurement error, with first-stage F between 190 and 474 across specifications. The halves

³¹This is the between/within split of the treatment as it enters the design, and so a different object from the panel decomposition of Table 1, which pools industry-years over 2005–2021.

Table A1. The clustering ladder: inference matrix for the output contraction

Procedure	Baseline NAICS-6 + year FE	Sector $\times$ year + NAICS-3 $\times$ year FE
$\hat{\beta}$ (CPR ^{pre} $\times$ post)	−0.024	−0.018
<i>Panel A. Industry-level variation as the effective sample</i>		
CRV1, NAICS-6 clusters ($G = 358$)	0.002	0.044
CRV1, NAICS-4 clusters ($G = 86$)	0.010	0.062
Wild cluster bootstrap, NAICS-4	0.029	0.090
Wild cluster bootstrap, NAICS-3 ($G = 21$)	0.022	0.21
Randomization inference, within-sector	0.030	0.055
<i>Panel B. Sector as the sampling unit</i>		
Sector block bootstrap (heterogeneity-robust estimand)	0.068	—
Ibragimov–Müller ($G = 21$ sectors)	0.23	0.23

Notes: p -values on the CPR^{pre} $\times$ post coefficient for the two key specifications (log output, 2018–2023, two-channel), all computed on the same sample; none suppressed. Primary cells in bold: the panel is the population of six-digit industries and the treatment varies at NAICS-6, with 64% of its variance inside three-digit sectors, so NAICS-3 clustering is the conservative bound on assignment correlation rather than the mandated level (Abadie et al., 2023). Panel A runs from the assignment level to that bound, so p rises as the effective sample coarsens; Panel B treats the 21 sectors themselves as the sample. Wild cluster bootstraps use Webb weights with the null imposed ($B = 9999$); within-sector randomization inference permutes pre-COVID reliance within NAICS-3 ($B = 1999$) and is the design-based test for the sector $\times$ year specification, whose identifying variation sector-by-year effects push inside clusters, where the NAICS-3 bootstrap loses power. CRV1 at NAICS-3 gives 0.007 (baseline) and 0.085 (sector $\times$ year) but over-rejects at $G = 21$; unrestricted randomization inference gives 0.003 and 0.013 but ignores the sector clustering of treatment and shock; neither is leaned on. The sector block bootstrap resamples whole NAICS-3 sectors for the heterogeneity-robust average effect of Section 4.5. Ibragimov–Müller t -tests the 21 sector-level coefficients and does not reject; its non-rejection reflects the dispersion of those coefficients (Table A2), not a loss of within-sector variation. An independent wild cluster bootstrap that refits the full model on each draw reproduces the NAICS-3 p -value (0.021), so that rung is not an artifact of the bootstrap implementation.

correlate at 0.73 within four-digit families, implying a within-family reliability of 0.84 against 0.92 in the baseline, so a homogeneous-coefficient attenuation prediction for the within-family estimate is $-0.024 \times 0.84/0.92 = -0.022$. The error-corrected estimates are -0.013 to -0.019 (the larger in magnitude with $p = 0.082$), so attenuation accounts for roughly a fifth of the observed decline and the within-family gradient is genuinely flatter. The same correction runs the other way at the three-digit level, moving the within-sector estimate from -0.018 to an instrumented -0.023 ($p = 0.024$).

The cell grid. Table A3 reports equation (2) estimated within each cell of the R&D-by-export-control-exposure grid. Point estimates sit at -0.030 and -0.049 in the two low-R&D cells and near zero in the two high-R&D ones, so the heterogeneity runs on the R&D dimension and not on the export-control axis. The cells are not statistically distinguishable, since the pooled triple interaction has $p = 0.63$, which is why the concentration is read as descriptive throughout.

Real output, common sample. Deflating by industry PPI (300 of the 358 industries) gives -3.5% per SD real against -3.6% nominal in the baseline, and -3.4% real against -3.9% nominal with the full controls (real CRV1 $p = 0.001$, wild $p = 0.022$ with controls). Deflation does not move the estimate; the PPI-covered industries simply contracted somewhat more than the full panel (-3.6% versus the -2.4% headline), so the gap reflects sample composition. The industries

Table A2. Sector-level acute coefficients behind the Ibragimov–Müller rung

NAICS-3 sector	$\hat{\beta}$	Six-digit industries
335	−0.400	17
314	−0.126	5
323	−0.073	4
331	−0.069	18
333	−0.064	38
313	−0.056	7
326	−0.037	16
312	−0.035	7
336	−0.031	27
332	−0.027	36
324	−0.027	5
339	−0.001	16
337	0.015	12
334	0.016	24
316	0.018	4
311	0.025	43
315	0.030	7
322	0.048	11
321	0.049	14
327	0.073	18
325	0.095	29
Mean (21 sectors)	−0.028	
Mean excluding 335	−0.009	

Notes: Equation (2) estimated separately within each of the 21 three-digit sectors, 2018–2023, NAICS-6 and year fixed effects; each coefficient is identified from within-sector variation across six-digit industries. Twelve of 21 are negative. The Ibragimov–Müller rung of Table A1 is an equally weighted t -test over this column.

Table A3. The output contraction concentrates in low-R&D industries

Cell	log output $\text{CPR}^{\text{pre}} \times \text{post}$	log employment
Low R&D, low export-control	−0.030	−0.023
Low R&D, high export-control	−0.049	−0.033
High R&D, low export-control	−0.011	−0.014
High R&D, high export-control	−0.004	−0.016

Notes: Equation (2) estimated within each cell of the R&D $\times$ export-control-exposure grid, 2018–2023. The pooled triple interaction testing whether the low-R&D/low-export-control cell is *differentially* more negative is insignificant ($p = 0.63$ for output), so the concentration is on the low-R&D margin. Significance stars are omitted: cell-level CRV1 rests on the few NAICS-3 clusters within each cell, where it over-rejects (Table A1), and the wild cluster bootstrap p for the low-R&D/low-export-control output coefficient is 0.23. The cell pattern is read as descriptive in the text, not as a tested contrast.

without a continuing six-digit PPI series are concentrated in machinery, electrical equipment, and transportation equipment, where BLS publishes at coarser product detail. Estimated year by year on the PPI-balanced membership (289 industries, 2018–2024), the real series reproduces the nominal path and then outlasts it: -3.5% , -3.5% , -2.3% , -2.7% , and -3.0% per SD for 2020–2024, each $p \leq 0.01$, while the nominal coefficient on the same membership thins to -1.9% ($p = 0.059$) in 2024. The log-PPI event study on the same membership is the third panel, and by construction it equals the nominal minus the real coefficient: -0.5 , 0.2 , -0.6 , -1.0 , 0.2 , and 1.2 log points for 2018 and 2020–2024. No post-2020 coefficient is significant at 5 percent; the 2024 value has $p = 0.083$ at NAICS-6 and 0.10 at NAICS-3, and the 2018 pre-coefficient is significant ($p = 0.004$) but runs against a post-2020 price rise. Deflation therefore preserves the year-by-year persistence through 2023 without changing it. In 2024 the real series alone carries it, and it does so through a 1.2-point relative-price rise that is not distinguishable from zero; the 2024 real and nominal coefficients (-3.0% and -1.9% , standard errors 1.1 and 1.0) are not statistically distinguishable from each other. A 2019-shipment-weighted variant shows no significant real gap through 2023 and a gap of -2.7% in 2024 ($p = 0.044$ at NAICS-6, 0.037 at NAICS-3), but the weights concentrate heavily (Kish effective $N \approx 61$ of 358; petroleum refining alone carries a tenth of the weight), so we place little weight on it in either direction; the measurement decomposition below locates where the real persistence sits.

Persistence across data sources (detail). The monthly reopening-phase timing reported in Section 4.2 is estimated on the Federal Reserve’s G.17 industrial production indexes at the industry-group level: 63 groups cover 91% of the panel’s 2019 shipments, with both exposures aggregated on shipment weights, estimated at monthly frequency over 2015–2023 with the twelve months of 2019 pooled as base. The June–December 2020 pooled coefficients carry p between 0.08 and 0.12 at this aggregation, and the 2020 average matches the annual acute estimate. (The China-exposure series in the production index is dominated by the quality-adjusted computer and electronics group; excluding that group, its profile matches the annual timing of Section 4.2, with no acute footprint and losses arriving in 2022–2023.)

In the same data the reliance gap closes during 2021 while the shipment gap does not. At the identical 63-group aggregation, nominal shipments retain the persistent path (-2.0% , -2.8% , -2.2% for 2021–2023) while PPI-deflated shipments match the production index, both within a point of zero after 2020; Section 5.1 states what that reconciliation implies, and the next paragraph separates its weighting and resolution components.

Weighting and resolution can be separated, and weighting carries more of the difference. Re-aggregating the same deflated shipments to the same 63 groups as an unweighted mean of six-digit logs, on an identical PPI-balanced membership, puts the 2024 gap at -2.9% ($p = 0.042$) against -2.2% ($p = 0.21$) for the value-weighted sum, close to the -3.0% ($p = 0.005$) estimated at six-digit. How the treatment is aggregated does not matter: shipment weights and equal weights give coefficients within 0.002 of each other, the two exposures correlating at 0.98 . On the shorter 2018–2023 window the unweighted re-aggregation closes about half the distance to the six-digit estimate rather than most of it, so resolution contributes as well. Coarse aggregates are also the less stable object here: adding one year to the window moves the unweighted 2023 coefficient from -1.2% to -2.5% , a range the six-digit estimate does not traverse (-2.5% to -2.7%), because with the 61 of the 63 groups that enter this window a single additional year re-estimates the unit effects appreciably. Both aggregations carry a wrong-signed 2018 pre-coefficient, larger unweighted ($+1.1\%$, $p = 0.018$) than at six-digit ($+0.7\%$, $p = 0.068$), so these comparisons bear on where the persistence is visible and not on whether the pre-period licenses it.

The labor margin needs no such qualification: administrative QCEW employment (Section 5.1) and BLS payroll employment at the four-digit level (-1.8% to -2.2% per SD, 2021–2023) agree with the Census series. QCEW worksite counts also fall monotonically (-3.9% by 2024) but are not significant net of industry trends on their own; the register-based CBP and BDS series remain the basis for the establishment margin.

The effective sample. Ordinary least squares identifies a variance-weighted average, so the regression’s own weights answer the question of which industries identify the reliance coefficient (Aronow and Samii, 2016). Squaring the residualized treatment and aggregating to industries, the identifying weight concentrates: its Kish effective sample is 31 industries of 358, the ten largest weights carry 49% of the total, and by sector the weight sits in fabricated metal products (34% of identifying weight against 7% of shipments), apparel (16% against 0.2%), and leather (8% against 0.1%) (Table A4)—the formal statement of the commodity tilt that the leave-one-industry-out exercise flags. The effective sample is also far more reliant than the panel it is drawn from: its mean contract reliance is 5.6% of shipments against 1.2% in the panel as a whole, which is to say the coefficient is identified in the region of the dose distribution where the nonparametric dose-response bites. It is not, however, drawn from systematically smaller industries: mean 2019 shipments are \$9.5 billion against \$15.9 billion in the panel, with the same median. The concentration is one of leverage alone. Signed contributions to the coefficient are diffuse and two-sided: no industry moves it by more than 0.0025 of its -0.024 , negative contributors are spread across fabricated metals, machinery, primary metals, and food, and the single largest-weight industry (a leather-goods industry) contributes *against* the effect, which is why every leave-one-out refit stays negative. Dropping each of the 358 industries one at a time leaves every estimate negative and within -0.028 to -0.022 (a different exercise from the sector-level range in Table 5), with only three moving the coefficient by more than a tenth of its value; the highest-leverage lines are metal-processing and apparel industries, and dropping every food, textile, and metal-processing industry at once leaves the coefficient at -0.030 , if anything larger.

The weighting path. The unweighted estimate is the average industry’s response; weighting by size moves the estimand toward the dollar-weighted response, a different parameter and also of interest (Solon et al., 2015). Figure A4 traces the coefficient as the weights move continuously from equal to 2019-shipment ($w_i = s_i^\alpha$, α from 0 to 1). The persistence contrast (post period 2021–2023, 2020 dropped) is flat along the entire path, -0.025 unweighted and -0.026 fully dollar-weighted ($p = 0.067$); employment is likewise stable, and the administrative QCEW employment persistence is unchanged under worker weighting ($p < 0.001$ in every post year).

Two cells that carry the headline do not survive the path, and we state them in full. The pooled post-2020 coefficient runs -0.024 , -0.025 , -0.024 , -0.019 ($p = 0.071$) and -0.013 ($p = 0.352$) as α moves from 0 to 1; the acute coefficient runs -0.022 , -0.020 , -0.014 ($p = 0.204$), -0.001 ($p = 0.975$) and $+0.023$. Part of the decay is mechanical: fully dollar-weighted estimates lose precision because the weights themselves concentrate (Kish effective $N \approx 61$), and capping any industry at one percent of total weight restores precision without restoring significance for either cell (pooled -0.020 , $p = 0.083$; acute -0.012 , $p = 0.332$). The $\alpha = 1$ acute sign flip is separately contaminated by the petroleum price collapse, which makes the least reliant giant industries crash in dollars. But petroleum does not account for the decay that precedes it: the acute estimate is already indistinguishable from zero at $\alpha = 0.75$. The acute and pooled results are therefore properties of the average *industry* rather than of the average *dollar*, while the persistence contrast holds under both. Consistent with a common response across the size distribution, size-tercile estimates

show no gradient (output -0.018 , -0.040 , -0.023 from smallest to largest tercile, differences insignificant), the employment response is, if anything, larger in bigger industries (reliance by post by log size -0.010 , $p = 0.031$). What weighting changes is not the presence of the response but which measurement cut retains it (the preceding decomposition): the one value-weighted object without persistence remains between-family real output.

Aggregate translation. A reader may want the gradient expressed as a share of manufacturing output. Reading the size off the reliance-quartile gaps and quartile shipment shares, a counterfactual in which every industry entered 2020 at the bottom-quartile mean of reliance puts the post-2020 differential output gap at about 2.6% of manufacturing shipments a year, and the linear per-standard-deviation version at 1.2%. Both weight an equal-weighted-industry coefficient by dollars, which is the substitution the weighting path above rules out, so we treat them as an upper-end order of magnitude and rest no claim on them.

Table A4. The effective sample of the baseline regression

<i>Panel A. Ten largest identifying weights (industries)</i>					
NAICS-6	Sector	Weight (%)	Cumulative (%)	Contribution	Reliance (%)
316992	Leather	8.2	8.2	+0.0008	10.5
332112	Fabricated metals	7.3	15.6	−0.0021	9.1
332993	Fabricated metals	6.6	22.2	+0.0015	8.7
315240	Apparel	5.7	27.9	−0.0010	9.0
332117	Fabricated metals	4.5	32.5	−0.0025	7.4
315190	Apparel	4.4	36.9	+0.0020	8.3
315210	Apparel	3.6	40.5	−0.0010	6.8
332722	Fabricated metals	3.4	44.0	−0.0006	6.7
336611	Transportation equipment	3.0	46.9	+0.0005	6.2
333514	Machinery	2.5	49.4	−0.0007	5.8
<i>Panel B. Identifying weight by sector (six largest)</i>					
Sector	Weight (%)	Shipment share (%)	Contribution	Industries	
Fabricated metals	33.7	6.6	−0.0063	36	
Apparel	15.8	0.2	−0.0012	7	
Leather	8.5	0.1	+0.0015	4	
Machinery	8.4	6.6	−0.0053	38	
Primary metals	6.5	4.2	−0.0046	18	
Transportation equipment	4.6	16.5	−0.0002	27	

Notes: Identifying weights are the squared residualized treatment of equation (2) (the two-channel baseline, 2018–2023), aggregated to industries and normalized (Aronow and Samii, 2016); the Kish effective sample of the weight distribution is 31 industries of 358, and the fifty largest weights carry 81% of the total. Contribution is the industry’s signed term in the coefficient; the terms sum to -0.024 . Reliance is the raw pre-COVID mean of contract work over shipments. Sector labels are three-digit NAICS names; shipment shares use 2019 values.

Establishment counts. The scarring evidence of Section 4 uses County Business Patterns establishment counts by NAICS-6, 2015–2023, merged to the panel treatment (358 industries); inference follows the primary NAICS-6 clustering of Section 4.5, with the NAICS-3 procedures as the conservative bound. Because CBP references March 12, the 2020 count predates the lockdowns and the post period is dated at 2021 with 2020 dropped as a transition year: the estimate is -0.022 per SD ($p < 0.001$; NAICS-3 CRV1 $p = 0.012$, wild $p = 0.009$); keeping 2020 and dating post at it gives -0.018 ($p < 0.001$; wild $p = 0.010$), attenuated as the timing predicts. CBP March employment falls 3.4% per SD on the same design ($p < 0.001$; wild $p = 0.018$). The event study runs

+0.004 (2018), -0.006 (2020, insignificant), -0.013 , -0.021 , -0.025 (2021–2023); its pre-period carries a mild positive drift, so the same industry-specific detrending as the output series (trends fit on 2015–2019, extrapolated) is applied and leaves a flat pre-period and a monotone post gap of -0.009 , -0.015 , -0.016 , with the pooled detrended estimate -0.014 ($p = 0.026$; wild $p = 0.023$). A conservative in-sample unit-trend specification absorbs the break, as it mechanically must when the gap deepens monotonically through the sample’s end. The output-side demand-rotation controls of Section 5.2 leave the post-2021 gap intact, so the establishment losses read as scarring rather than the thinning of industries whose demand rotated away.

Entry and exit margins. Business Dynamics Statistics separates the net change into establishment entry and exit at NAICS-4.³² The net-count specification reproduces the CBP result (-0.026 per SD, $p < 0.001$; wild $p = 0.009$).

The split attributes it to entry: log entries fall 7.8% per SD ($p = 0.003$; wild $p = 0.036$) and the entry rate falls 0.46 percentage points ($p = 0.005$; wild $p = 0.048$) against a manufacturing-wide post-period entry rate of 7.2, a rate raised by the post-2020 national surge in business formation (Dinlersoz et al., 2021; Decker and Haltiwanger, 2023). The exit rate is statistically flat (-0.17 percentage points, $p = 0.11$; wild $p = 0.22$), and log exits *fall* (-4.3% per SD, $p = 0.017$; wild $p = 0.050$), the mechanical footprint of a shrinking stock at an unchanged hazard; even the 2020–21 cohort shows no differential closure.

Event studies show no significant pre-period coefficient on either margin, but the entry margin is not powered to establish a flat pre-trend: the 2017 entry coefficient is -0.039 with a standard error of 0.026, so the interval comfortably admits a pre-existing shortfall as large as the post-period effect itself, and the non-rejection should be read as silence. BDS flows dated t cover March $t - 1$ through March t —the same timing that dates the post period at 2021—so the 2020 entry coefficient (-0.065 , $p = 0.036$) largely predates the shock and reads as an incipient shortfall rather than the break itself; the gap is open in the first COVID cohort (-0.065 , $p = 0.026$ in 2021), widens in 2022 (-0.081 , $p = 0.031$), and remains open in 2023 (-0.072 , $p = 0.018$). BDS is published at the four-digit level, one level coarser than the paper’s six-digit panel, so these standardized magnitudes are per NAICS-4 standard deviation and are not directly comparable to the six-digit estimates.

Intensive and extensive margins. Establishment counts and output both fall, which raises the question of whether the output response is the arithmetic shadow of a thinner stock of plants. Log shipments decompose additively into log establishments and log shipments per establishment, so estimating all three on one sample splits the response into an extensive and an intensive margin that sum to it exactly. Detrending all three as the establishment series above is detrended, the acute-window response divides $-0.022 = -0.004$ extensive ($p = 0.20$) + -0.018 intensive ($p = 0.003$; NAICS-3 CRV1 $p = 0.077$, wild $p = 0.16$). The extensive share of the gap then climbs year by year: 15% in 2020, 28%, 34%, and 79% by 2023, when output per surviving establishment is back at zero (-0.004 , $p = 0.79$). Timing explains the acute year. CBP references March 12, so the 2020 count predates the lockdowns and the denominator cannot yet respond; what fell was production at plants already standing. The margins operate in sequence, the shock striking output at incumbent establishments and the entry shortfall converting that into a durable capacity gap as it accumulated.

³²86 manufacturing industries, the six-digit treatment aggregated to NAICS-4 with pre-period shipment weights, 2018–2023 with 2020 dropped, post dated at 2021, NAICS-4 and year fixed effects, the China-exposure control, $N = 430$. Because the aggregated treatment varies at NAICS-4, primary inference clusters there (86 clusters), with the 21-cluster NAICS-3 wild bootstrap as the conservative bound.

The split also speaks to the factoryless alternative, since reclassification out of manufacturing would move plants and shipments together and load on the extensive margin, which in 2020 does not move.

Two limits. Pooled across the whole post period the extensive margin carries 83% of the response undetrended and 48% detrended, on an insignificant pooled intensive term. What the split establishes is therefore the timing of the two margins. The pooled coefficient is a separate question. And the intensive estimates, significant under the primary clustering, do not clear the conservative NAICS-3 bound, the same power loss Table A1 documents throughout. Extensive shares are ratios of two estimated coefficients and carry no standard error; inference rests on the intensive coefficient.

Post-2020 changes and industry contributions. Table A5 reports the cost-item event studies behind Section 5.3: each item as a share of shipments, regressed on standardized pre-COVID reliance interacted with a post-2020 indicator, with industry and year fixed effects, on the balanced industry set. The matched items come from the Annual Survey of Manufactures through 2021 and the 2022 Economic Census; the 2023–24 integrated survey publishes contract work and energy only, which is why the windows differ. Table A6 reports the exact decomposition of the pooled acute coefficient into industry contributions used in Section 4.5: each industry’s contribution is its Frisch–Waugh–Lovell weight times its residualized outcome, and the contributions sum to the coefficient.

Table A5. Post-2020 change in cost items as shares of shipments, by pre-COVID reliance

Cost item (share of shipments)	Window	N	Post $\times$ reliance	SE	p	Predicted
Contract work	2018–2024	277	−0.104	0.019	< 0.001	down
Materials	2018–2022	323	0.005	0.004	0.15	up (weak)
Production hours	2018–2022	330	0.006	0.005	0.21	up
Capital expenditure	2018–2022	322	0.017	0.015	0.26	up
Fuel and electricity	2018–2024	292	0.007	0.010	0.46	flat
Employment	2018–2022	330	0.009	0.005	0.07	up

Notes: Each row is one regression of the log of the item over shipments on pre-COVID reliance (one SD) interacted with post-2020, with the China-exposure interaction of equation (2) as a control, industry and year fixed effects, and clustering at NAICS-6; coefficients are in log points. “Predicted” is the sign an in-house replacement of the contracted work would give. N is industries observed in every year of the window.

Relative prices. Estimating equation (2) with the log industry PPI as the outcome (annual means of the same series used for deflation, 300 industries) gives −0.001 per SD (wild $p = 0.93$) and −0.005 ($p = 0.34$) with the full controls; the event-study coefficients are small and insignificant in every post year through 2023, and the 2024 coefficient on the PPI-balanced membership is +0.012 ($p = 0.083$), reported beside the real and nominal series above. The pre-period carries one small negative coefficient, −0.006 in 2018, which is statistically significant ($p = 0.002$) and works against finding a post-2020 price rise. The foreign-input channel shows a modest relative price *decline* over the window (−0.017, CRV1 $p = 0.035$), which annual averaging over 2020–2023 can reconcile with the early-2020 import-price spikes documented at monthly frequency by Meier and Pinto (2024); the contract-reliance price response is a precise zero through 2023 and a marginal positive in 2024.

Composition and inventory controls. The teleworkability index is each industry’s May-2019 occupation mix from the BLS Occupational Employment and Wage Statistics, weighted by the work-from-home feasibility scores of Dingel and Neiman (2020) and mapped to NAICS; the

Table A6. Ten largest industry contributions to the pooled acute coefficient

Industry (NAICS-6)	Reliance (¢/\$)	Share of $\hat{\beta}$	Δ output (%)
Powder metallurgy parts (332117)	7.4	10	−7
Iron and steel forging (332111)	5.5	10	−14
Nonferrous forging (332112)	9.1	9	−25
Oil and gas field machinery (333132)	4.0	7	−25
Steel investment foundries (331512)	5.6	4	−4
Cut-and-sew apparel contractors (315210)	6.8	4	−3
Women’s cut-and-sew apparel (315240)	9.0	4	−26
Blank recording media (334613)	0.2	4	+36
Industrial molds (333511)	5.3	4	−4
Aluminum foundries (331524)	4.5	4	−14
Ten industries together		60	

Notes: Exact Frisch–Waugh–Lovell decomposition of the pooled acute coefficient (−0.024, 2018–2023, NAICS-6 and year fixed effects); contributions sum to the coefficient. Reliance is the 2015–19 mean in cents per dollar shipped. Δ output is the 2020–23 mean of log shipments relative to 2015–19, in percent. The ten industries hold 0.9 percent of 2019 shipments. Blank recording media enters through a large positive residual at near-zero reliance.

physical-proximity index applies the same occupation weighting to the O*NET physical-proximity score (Mongey et al., 2021) and is essentially uncorrelated with reliance ($r = -0.09$). Under these composition controls the employment effect attenuates by about a quarter, so part of it is composition. Pre-shock inventory intensity (NBER-CES inventories over shipments, 2015–2018) is, if anything, slightly *higher* in contract-reliant industries (cross-industry correlation +0.18), carries no output effect of its own (−0.016, $p = 0.13$), and leaves the contract-reliance effect essentially unchanged when interacted with post-2020 (−0.024 to −0.022, wild $p = 0.019$; the row in Table 5). In-window inventory dynamics can be observed directly: the AIES historical timeseries reports end-of-year inventories at NAICS-6 on one harmonized basis for 2012–2024. Estimated as outcomes in equation (2) on 2018–2024, total inventories fall in contract-reliant industries by about as much as shipments do (−0.020, $p = 0.049$), the inventory-to-shipments ratio does not move (0.004, $p = 0.67$), and the work-in-process share of inventories, the component that sits at or awaits a contractor, is likewise null. The ratio rises briefly in 2020 (0.013, $p = 0.027$) as shipments fall faster than stocks, and is flat thereafter. Reliant industries, that is, neither drew down nor failed to rebuild stocks relative to their output. Inventories scaled with the lost volume.

Temporary staffing. Purchased temporary staffing is collected in the same survey and is of the same size as contract work (the median industry spends 0.87 of its contract-work outlay on staffing agencies), so it is a labor-supply channel that could mimic the timing. It is only weakly related to reliance ($r = 0.20$; a third of top-quartile reliance industries are also top-quartile staffing users), and racing the two intensities against the 2019–2020 shipments change leaves reliance at −0.025 while staffing carries nothing (+0.000, $p = 0.97$). The staffing item is published for 2023–24 only, so that race is descriptive.

Customer shocks and steel tariffs. Table A7 reports the customer-exposure and Section 232 checks of Section 4.3 in full, with the construction of the exposures in its notes. It is generated by `code/seminar/sm_08_confounder_table.py` from the estimates of `sm_07_us_confounders.py` (BEA shares) and `sm_10_colocation_exposure.py` (co-location). The BEA commodity detail is

coarser than NAICS-6: the 358 industries have 150 distinct customer vectors, the three foundry industries share one, and so do forging and powder metallurgy. The co-location measure is built from County Business Patterns for 2016. It behaves as a customer-link proxy should: oilfield machinery sits at eleven times the national density of oil and gas employment, and aircraft parts at nearly five times that of aerospace. The Census 2017 materials-consumed table, a third candidate source, is no finer than the BEA detail for the industries at issue: it reports one code for all forgings and one for all ferrous castings, and does not itemize powder-metal parts or oilfield machinery.

Demand-rotation and detrending controls. The realized demand-rotation index is each industry’s 2017 final-use composition—consumption (split by durability), equipment, other investment, exports, and government purchases, from the BEA benchmark use table—weighted by the corresponding national real final-demand paths from the national accounts, expressed as log gaps from 2019; the alternative interacts the final-use shares and durable status with year fixed effects, so any industry with a given demand mix may follow any path. Reliance barely loads on the final-use composition (no share correlation exceeds 0.11 in magnitude), and the realized rotation, if anything, ran against contract-reliant industries ($r = -0.15$ with the 2022 index), which is why neither control moves the estimate (Section 5.2). The industry-specific detrending fits each industry’s linear trend on 2015–2019 and extrapolates it forward. The choice of fitting window matters and we report both with inference. On the 2015–2019 fit the acute estimate is -0.022 (CRV1 $p = 0.013$, wild $p = 0.041$) and the full-window estimate -0.025 ($p = 0.015$, wild 0.051); fitting on 2012–2019 instead gives an acute estimate of -0.019 that is marginal at the assignment level and fails the conservative bound ($p = 0.044$, wild 0.109), and a full-window estimate of -0.019 that is insignificant at both ($p = 0.116$, wild 0.186). The 2020 break is present on either window; the persistence beyond it is not established on the longer one. The cross-system agreement of Table 4, not the detrending, carries the persistence claim.

A sign-flip test (Tōhoku 2011). The framework predicts not just a smaller effect under a less synchronized shock but a *change of sign*: contracting out diversifies idiosyncratic supplier risk, so under a narrow, asynchronous disruption the contract-reliance coefficient should turn non-negative (Section 7.2). The 2011 Tōhoku earthquake—a brief, Japan-specific supply shock—is such a case. On the 188 industries observable over 2008–2011, the same design with the shock dated to 2011 gives a contract-reliance coefficient of 0.016 (insignificant), against -0.024 for those *same* industries under COVID: the point estimate flips from significantly negative to positive, the direction the framework predicts for a low-synchronization shock.

Section 7.3 reports the clustering ladder, the minimum detectable effect, and the comparison with the COVID interval on the same industries; the construction of the Japanese exposure measures follows.³³

A further heterogeneity prediction—that the loss should be larger where a contracted step is hardest to bring back in house—is not identified at this resolution. Three pre-2020 proxies for weak fallback (capacity utilization, capital intensity, average establishment size) give triple interactions

³³The contract-reliance estimate is positive but imprecise—one narrow event on 188 industries—so we read it as the mechanism’s predicted sign and take no magnitude from it. The flip is robust to how Japanese exposure enters: an import-penetration measure (built as the China one, from Tōhoku-window TDM imports through Pierce-Schott) and a BEA input–output measure of dependence on Japanese-penetrated suppliers both leave the contract-reliance coefficient near 0.017. The Japanese coefficients themselves are uninformative at this resolution. Penetration loads positively (0.038, Japanese import competition receding), input dependence is a noisy zero (-0.004 , only NAICS-3 granular), so the test rests on the contract-reliance sign alone.

Table A7. Customer shocks and steel tariffs

	2020 (acute window)		2018–2023	
	Estimate	NAICS-3 <i>p</i>	Estimate	NAICS-3 <i>p</i>
Baseline	−2.18*** (0.65)	[0.007]	−2.40*** (0.76)	[0.008]
<i>Dropping the exposed industries</i>				
Drop oilfield machinery, 333132 (357)	−2.06*** (0.64)	[0.010]	−2.26*** (0.74)	[0.007]
Drop aerospace-linked industries (349)	−1.76*** (0.65)	[0.036]	−2.13*** (0.81)	[0.020]
Drop powder metallurgy, 332117 (357)	−2.14*** (0.67)	[0.010]	−2.25*** (0.77)	[0.018]
Drop all three (347)	−1.55** (0.66)	[0.063]	−1.77** (0.79)	[0.043]
Drop top decile of any BEA customer share (268)	−1.07 (0.70)	[0.167]	−1.35 (0.82)	[0.092]
<i>Controlling for the exposures</i>				
+ Section 232 input shares	−1.59** (0.63)	[0.010]	−2.16*** (0.79)	[0.008]
+ BEA customer shares	−1.61*** (0.59)	[0.004]	−1.96*** (0.72)	[0.003]
+ BEA customer shares and Section 232	−1.35** (0.59)	[0.018]	−1.91** (0.76)	[0.004]
+ Co-location customer exposure	−2.04*** (0.66)	[0.008]	−2.26*** (0.77)	[0.007]
+ Co-location and Section 232	−1.61** (0.65)	[0.006]	−2.10** (0.81)	[0.008]
+ Both customer measures and Section 232	−1.38** (0.63)	[0.014]	−1.72** (0.79)	[0.016]
<i>Within three-digit sectors</i>				
NAICS-3 × year	−0.84 (0.85)	[0.321]	−1.79** (0.90)	[0.090]
+ BEA customer shares and Section 232	−0.59 (0.81)	[0.462]	−1.19 (0.89)	[0.212]
+ Co-location and Section 232	−0.84 (0.82)	[0.224]	−1.73* (0.89)	[0.078]
+ Both customer measures and Section 232	−0.52 (0.82)	[0.428]	−1.06 (0.90)	[0.248]
<i>Among auto suppliers</i>				
Top quartile of BEA auto share only (90)	−2.38*** (0.84)	[0.014]	−2.57*** (0.87)	[0.009]

Notes: Percent change in shipments per one SD of pre-COVID reliance interacted with post-2020, in the paper's two-channel specification (China import exposure × post included), NAICS-6 and year fixed effects, 358 industries unless the row gives the number of industries used. Standard errors CRV1 on NAICS-6 in parentheses, with the CRV1 NAICS-3 *p*-value alongside. Customer groups are oil and gas extraction and support activities (NAICS 211, 213), aerospace (3364), and motor vehicles, bodies and parts (3361–3363), each entered separately and interacted with post. BEA customer shares are the shares of each commodity's 2017 total use bought by each group, from the BEA 2017 detailed Use table, mapped to NAICS-6 through the table's own concordance; the commodity detail is coarser than NAICS-6, giving the 358 industries 150 distinct customer vectors, so these controls partly act as sector dummies. Co-location customer exposure is, for each six-digit industry, the density of each customer group's employment in the counties where the industry's workers are, relative to the national density, from County Business Patterns 2016, with suppressed county employment imputed from establishment size classes; every industry has its own value. Section 232 shares are the pre-period steel and fabricated-metal input shares. With the controls interacted with year, the 2015–2023 event study (base 2019) gives 2020 −1.51***, 2021 −2.50***, 2022 −3.75***, 2023 −1.57; the joint test of the 2015–2016 and 2018 leads has *p* = 0.65 (0.84 without the controls). * *p* < 0.10, ** *p* < 0.05, *** *p* < 0.01.

that are correctly signed on two of three but nowhere near significant, and the directional pattern does not survive holding the R&D moderator fixed. The build and estimates are in the replication package.

B The Gated Value Share

Contract work is a charge for an operation performed on materials the buyer owns, so the fee understates the output that operation gates. A natural conjecture is that the magnitude discussed in Section 4.7 is an accounting artifact of that gap. This appendix calibrates the conjecture and rejects it, on routes and criteria fixed in the replication log before any parameter was measured.

Three parameters connect the fee to the output at risk: a toll rate τ , the contractor’s charge as a share of the value of the item returned to the buyer; a throughput loss δ , the proportional reduction in contractors’ ability to process during 2020; and a downstream multiplier $m \geq 1$, the ratio of buyer shipment value that cannot be completed to the value of the gated item. Gated throughput per dollar shipped is CPR/τ , output at risk is m times that, and the loss is δ times the output at risk. Writing A for the ratio of the estimated response to one standard deviation of reliance, $A = 0.024,17/0.015,87 = 1.523$, the account requires

$$\delta m/\tau = A = 1.523. \quad (4)$$

The multiplier matters. A one-parameter version of the conjecture, in which a toll rate near 15% converts a 1.6% fee into a roughly 10% gated share, omits it: an operation sitting mid-chain blocks more of the buyer’s shipments than passes through it. With three unknowns and one equation this is a calibration rather than a test. It asks whether ordinary parameter values reproduce A , it identifies no mechanism, and it bears on none of the interpretive limits recorded elsewhere in the paper.

Measurement. $\tau = 0.511$, the receipts-weighted own-account value-added share of shipments among industries that sell contract manufacturing, with toll receipts netted from both sides. A contractor performing an operation on another firm’s materials must recover the costs and margin it earns performing that operation on its own account, which makes the own-account ratio the structural analog of the toll charge. It is stable at 0.510 on the 2017 Economic Census and 0.512 on 2022 (medians 0.505 and 0.519, interquartile range 0.43–0.58), and moves only from 0.497 to 0.522 as the value-added content of toll receipts is swept from 1.00 down to 0.80. Contractors are conversion-cost businesses, so a toll near half the returned item’s value is what their technology implies.

$\delta = 0.050$, the receipts-weighted 2019–2020 log decline in shipments among the 30 industries at toll intensity of 5% or more, which hold \$23.6 billion of the \$41 billion in contract-manufacturing receipts reported by manufacturing industries. Across all 178 industries with positive receipts it is 0.054. Contractor industries lost *less* output in 2020 than manufacturing as a whole, which fell 9.7%, and this excess is negative at every threshold.

$m \in [1.00, 2.67]$, bounded above by the ratio of shipments to materials (median 2.06, shipment-weighted 1.99, ninetieth percentile 2.67). An item returning from a contractor enters as a material, and no single material gates more output than the whole material bundle supports.

Table A8. The gated value share reproduces at most a fifth of the estimated response

δ	m	τ	implied A	share of $A = 1.523$
0.054	1.00	0.511	0.106	7%
0.054	2.06	0.511	0.217	14%
0.054	2.67	0.511	0.283	19%

Notes: Rows set the downstream multiplier at its lower bound, its median, and its ninetieth percentile. δ is the all-contractor value; the toll-intensive cut gives 0.050 and lowers every row. Implied A is $\delta m/\tau$ from equation (4), to be compared with $A = 1.523$.

Results. The most generous admissible combination reproduces 19% of the estimated response. Closing the gap would need $m = 14.4$ against a measured ceiling of 2.67, or $\delta = 0.778$ at $m = 1$, or a toll rate of 3.5% against 51% measured. A coherence condition fails independently: at the effective sample’s mean reliance of 5.6%, a gated output share of at most one requires $m/\tau \leq 17.9$, and reproducing A needs 28.2. The failure criterion fixed in advance was $\delta > 0.50$ or $m > 4$, and both are breached by wide margins.

The pre-specified reporting rule calls for two disclosures. A narrower cut at toll intensity of 10% or more returns a far more favourable $\delta = 0.007$ and is not used: it is six industries dominated by ammunition, government-owned and contractor-operated defense capacity on a procurement path, whose shipments grew 5.7% in 2020. An upper-end benchmark from HTS heading 9802 was flagged in advance as weak and is likewise unused; it would not have helped, since dutiable value added abroad runs 42 to 86% of entered value and corroborates a high toll rate.

Measurement leaves two soft spots. δ is built on contractor industries’ annual shipments rather than on the toll flow itself, since receipts are collected only in Economic Census years, so a toll segment that collapsed while own-account business held up would leave δ understated. And the calibration is annual on both sides, so a sub-annual outage averaged into an annual δ is invisible here. Neither is small, and neither is close to the order of magnitude the verdict would have to move. We report this as a bounded negative result: it closes the deflationary reading of the magnitude without replacing it with an identified alternative.

C Cross-Country Data and Comparability

This appendix documents the data behind Section 6: what each source collects and how, what its subcontracting variable measures, how the underlying data-generating processes differ, and which of those differences the empirical model absorbs. The design principle throughout is that the estimation never compares countries to each other. Each economy contributes only the within-country ranking of its industries and the within-industry time path of its own real output; everything cross-country is differenced out or standardized away.

C.1 Sources and collection

United States. The treatment and outcome come from the Census Bureau’s establishment surveys described in Section 2: a mandatory annual sample survey (ASM) benchmarked by quinquennial censuses, reporting establishment-level dollar values that Census aggregates to NAICS-6. The

contract-work item is defined as the cost of work done by others on materials furnished by the establishment; the outcome is the dollar value of shipments, with a PPI-deflated variant in Table 5.

Europe. Payments to sub-contractors come from the Structural Business Statistics, compiled by national statistical offices under a common EU regulation, at the enterprise (not establishment) level and disseminated at NACE four-digit.³⁴ The legal definition is broader than the U.S. item: a subcontracting relationship requires that the customer participate in the conception of the product or provide the materials to be processed, so the European variable pools furnished-materials tolling (the U.S. concept) with design-specification subcontracting in which the contractor buys its own inputs.

The outcome is the short-term-statistics industrial production index, a volume index built from deflated turnover or physical quantities with base-year weights; four-digit publication is voluntary, which is what restricts the fine panel to Germany, Italy, France, Spain, and Greece (verified against both the annual and the monthly series); the eleven other countries with a 2017 subcontracting cross-section enter the division-level comparison of Section 6 at two-digit only. A German validation anchors the broad treatment to the narrow concept: the Destatis cost-structure survey collects the strict furnished-materials item (Kosten für Lohnarbeiten), and across German four-digit industries the two intensities, both measured before 2020, correlate at Spearman 0.96 (Pearson 0.96); Appendix C.4 discusses what that comparison does and does not establish.

Japan. Japan’s treatment comes from the final Census of Manufactures (reference year 2019), an establishment census whose cost breakdown—published for establishments with 30 or more employees at JSIC four-digit—separates the cost of consigned production on furnished materials (the closest foreign analog of the U.S. item) from production-support outsourcing such as maintenance, inspection, and packaging; the intensity is the former over shipments, with the broader sum as a robustness variant. The outcome is the METI production index, built from roughly six hundred item-level quantity relatives aggregated with value-added weights. Because the index’s industry categories are goods-based rather than JSIC-based, we construct a concordance of 75 cells, each flagged for whether its mapping is exact or approximate, and estimate a conservative 22-cell sector panel alongside it.³⁵

Mexico. Mexico is the one economy whose annual manufacturing survey publishes both sides of the market—the payer-side item (gastos por maquila, spending on industrial processing performed by another firm on the establishment’s materials, the closest analog of the U.S. item) and a seller-side item split by client location—and it is used in Section 3.1 only to check that the reliance concept is domestic to first order. No estimate from it enters the comparison here. Its payer-side treatment carries a pre-existing relative decline and its seller-side variation is close to collinear with the sectoral bloc that carries Mexico’s lockdown incidence, so neither side supports a clean acute estimate. On the payer side, the 2018–2024 estimate of -0.047 is unmoved by controls for U.S. imports of the same product from Mexico, U.S. employment in the same product, exposure to the 2021 outsourcing reform, and the foreign-principal share, but a treatment-specific linear trend absorbs it (-0.009 ,

³⁴Variable V23110, drawn largely from a mix of business surveys and administrative or tax records. It was collected triennially; the 2017 wave is the pre-COVID cross-section, divided by production value (V12120) from the same framework, with repair and installation (NACE 33) excluded; Appendix C.4 explains both choices.

³⁵Where one JSIC three-digit group spans several index industries, the index rows are combined with their published weights; where one cell spans several JSIC groups, the treatment is aggregated as the ratio of summed numerators to summed denominators. At the 22-cell sector level the mapping is one-to-one by construction.

$p = 0.32$), which is why it is not read as evidence on persistence; the detail is in the replication package.

C.2 Variable definitions and collection differences

The U.S., European, and Japanese treatments are payer-side: they record what an industry’s producers spend on having outside firms perform production. They differ in breadth. The U.S., German, and Japanese items require that the customer own the materials; the harmonized European item does not. The unit differs too: U.S. and Japanese data are establishment-based, European data enterprise-based, so tolling between two plants of one European firm nets out. So does the frame, since the Japanese breakdown covers larger establishments only and European small-cell values are suppressed.

Those differences leave the distributions at very different levels—contract work averages about 1.2% of shipments in the United States, the strict German item about 3% of production value, the Japanese item about 5% of shipments, and the broad European measure, taken over production value, about 6% in Italy and 7% in France, with a long right tail in every country. Those level differences mix institutions (Mittelstand tolling networks, Japanese consignment production) with definitions, and no part of the analysis interprets them. The outcome processes differ too: the U.S. outcome is a nominal dollar aggregate from a survey, while the European and Japanese outcomes are constructed volume indexes with their own base years, weights, and revision practices. The choice matters. On the same five European countries and sample, the volume index shows the acute effect while the nominal SBS value added does not (-0.017 against -0.007 , insignificant, on the earlier value-added treatment), so nominal outcomes wash out a real-quantity effect; the U.S. panel, where deflation leaves the estimate unchanged (Table 5), does not share this sensitivity.

C.3 Fixed effects and identifying variation

Each layer of the specification is aimed at one class of the differences above.

Standardization within country. The treatment is z-scored within each economy, so only the ranking of industries inside a country identifies β ; the level differences across concepts never enter. Where a broader concept (V23110) proxies a narrower one (tolling), classical measurement error in the ranking attenuates the estimate, so the harmonized European coefficients are, if anything, conservative. In Germany, where the broad and the strict concepts rank industries almost identically, the national treatment gives a larger estimate than the harmonized one, which is the direction attenuation predicts.

Industry (country-by-industry) fixed effects. Any time-invariant feature of an industry’s measurement—concept breadth, size cutoffs, enterprise versus establishment units, suppression patterns, the index’s base-year normalization—is a level and is absorbed. This also handles the index rebasing (2015=100, 2020=100, 2021=100) mechanically, since a base change is a multiplicative constant in logs.

Country-by-year fixed effects. In the pooled European panel, each country’s entire aggregate path—the severity and timing of its lockdowns, its inflation, its fiscal response, any revision or redesign of its statistical system in a given year—is differenced out, so identification is within country and year, across industries. It is also why employment does not travel as an outcome: employment responses ran through country-specific short-time work schemes that a country-year effect absorbs only in the aggregate, so real output is the outcome carried across countries. This bears on the divergence discussed in Section 6. Kurzarbeit, cassa integrazione, chômage partiel and Japan’s

employment adjustment subsidy all operated at industry-varying intensity, so a country-year effect leaves the industry gradient of scheme take-up in the residual. Where a scheme was taken up most heavily in the industries this paper’s treatment ranks highest, the measured employment response abroad would be damped by the policy rather than by the absence of a shock. Take-up by country and sector is published by Eurostat and the OECD, which makes this checkable; we flag it as the natural next test of whether the persistence documented here is a feature of the production arrangement or of the labor-market institutions surrounding it.

A pre-determined treatment. The treatment is fixed at its pre-2020 cross-section (2017 in Europe, 2019 in Japan, the 2015–2019 mean in the U.S.), so post-period statistical redesigns—the 2021 European business-statistics reform, which changed the SBS population, and Japan’s 2020–2021 change of survey vehicle—cannot contaminate it. The outcome indexes are maintained as continuous series across those reforms.

Inference and stress tests. With few clusters, cluster-robust p -values are reported alongside randomization inference that permutes the treatment within country (2018–2024 pooled $p = 0.026$). The clustering ladder of Section 4.5 travels: moving from country-by-NACE-2 clusters to country-by-NACE-3 to the country-by-NACE-4 unit, the pooled acute p falls from 0.0003 to 0.0001 to below 0.0001 and the 2018–2024 p from 0.080 to 0.059 to 0.036, in line with the within-country randomization p of 0.026; Japan’s acute estimate is $p = 0.007$ clustered on the 75 concordance cells against 0.020 on JSIC-2 sectors. Assignment-level clustering does not rescue the individually imprecise countries: Germany ($p = 0.14$), Spain (0.14), and France (0.34) stay insignificant, so the sign pattern continues to carry the cross-country evidence, with no single country behind it. The pooled European point estimate is stable across fixed-effects structures (-0.019 under common-year effects, separable country effects, and country trends, and -0.021 weighted by value added), but its full-window significance leans on Greece and Italy: dropping Italy lowers the estimate by about two-fifths and dropping Greece by about a quarter, and neither leave-out is significant. The pooled event study also has a small positive 2018 coefficient (0.010). The first of these is stated in the text. Japan’s acute estimate is robust to the broad treatment definition (-0.031 , $p = 0.012$) but attenuates to -0.017 ($p = 0.18$) on exact-concordance cells only, and the between-sector panel is flat. A broader European panel at two-digit detail (seventeen countries) gives -0.049 ($p = 0.017$), directionally consistent but too coarse to weight. A quartile dose-response in the pooled European panel is not monotone, so the European evidence rests on the linear-in-standard-deviation gradient rather than on curvature, as in the United States (Appendix A).

What is not absorbed. Two things remain outside the design. Differences in effect size across economies cannot be separated from differences in measurement breadth, so the paper draws no cross-country magnitude comparisons; the exhibit content of Figure 4 is the sign pattern and the event-study shape. And countries whose production indexes exist only at two-digit detail are excluded from the fine panel rather than pooled with it, trading coverage for a common industry resolution.

C.4 Comparability of the European treatment

The European subcontracting variable differs from the Census item in its definition, in how it treats contracting across borders, in its denominator, and in its industry coverage. We specified four changes from those differences before estimating any of them, applied them cumulatively, and report every step in Table A10.

Table A9. Full-window estimates by economy

Economy	Acute 2018–20	Full window
United States	−0.022	−0.023
EU-5 pooled	−0.036	−0.021
Germany	−0.022	−0.015
Greece	−0.044	−0.069
Spain	−0.037	0.000
France	−0.012	0.009
Italy	−0.075	−0.046
Japan	−0.028	−0.006

Notes: Specifications and sources as in Table 6; the acute column is repeated from it for comparison. The full window is 2018–2025 for Europe and 2018–2024 for the United States and Japan. The U.S. loss persists while the European and Japanese gaps close by 2022–2023, so the full-window estimates abroad are attenuated; in France recovery overshoots the acute dip far enough to turn the point estimate mildly positive, and in Spain it returns to zero. The acute window is therefore the object compared across economies (Section 6).

Definition. Commission Regulation (EC) No 250/2009 defines payments to sub-contractors (V23110) as payments for industrial goods and services supplied within a subcontracting relationship. Such a relationship exists when the customer participates in the conception of the product, by supplying technical specifications, the materials to be processed, or both, and then sells the product and carries the after-sales liability. A supplier that buys its own inputs to the principal’s specification therefore counts. The Census item counts only work done by others on materials the establishment owns. The gap is widest where principals design and sell while their suppliers source the inputs, which describes much of European fashion. In Italy about four-fifths of textile, apparel, and leather enterprises have fewer than ten employees, and Istat describes them as working largely as subcontractors or to order for larger firms.³⁶ The regulation also notes that the variable is rarely identified separately in company accounts, so national offices may estimate it.

Contracting across borders. U.S. apparel production runs through yarn-forward rules of origin under CAFTA-DR and USMCA: U.S. mills export yarn and fabric to Central America and Mexico, and finished garments return as imports. None of that flow is recorded as contract work in the ASM. U.S. textiles, apparel, and leather (NAICS 313–316) paid \$1.4 billion for contract work in 2019, 2.6 percent of all manufacturing contract work on 1.1 percent of manufacturing shipments. The European regulation sets no residency condition on the subcontractor, so work sent to a sewing plant outside the EU may enter V23110; we have found no methodological note that settles how national offices record it. Both differences push in one direction. Relative to the U.S. item, the European variable ranks textiles, apparel, and leather higher.

Denominator and coverage. Two further differences are mechanical, and they motivate the two changes we adopt. The Census item is scaled by shipments, and the closest structural-business-

³⁶Istat Working Paper 3/2023, on the fashion supply chain, from 2019 structural business statistics: 81.1 percent of about 56,000 enterprises in NACE 13–15 are micro-enterprises, and they bill 12.9 percent of the sector’s turnover.

statistics analog is production value (V12120), not value added (V12150). Scaling by value added makes the treatment depend on industry margins, and in an earlier extension it made the treatment’s denominator identical to the nominal outcome. NACE Section C also includes repair and installation of machinery and equipment (division 33), which NAICS classifies outside manufacturing (NAICS 811). The division is contract-intensive: repair of ships, of aircraft, and of fabricated metal products are three of Germany’s ten most subcontracting-intensive four-digit classes. Its industries have no counterpart in the U.S. panel, so we drop it.

Data quality and a national measure. Eurostat flags every Polish 2017 value of V23110 as following a different definition and nearly every Italian, French, and Finnish value as a break in series. The third change drops Poland and takes the Italian, French, and Finnish treatment from the 2014 wave instead. The fourth replaces the German treatment with the Destatis cost-structure item (Kosten für Lohnarbeiten, GENESIS table 42251-0006), contract work on the firm’s own materials as a share of gross production value, averaged over 2015–2019. That is the U.S. concept on a nearly identical denominator. Cells Destatis suppresses are left missing, not imputed.

Results. Dropping repair raises the European employment estimate by about 40 percent and more than doubles it once textiles are excluded, because repair industries rank high on reliance while their employment held up in 2020. The production-value denominator raises the pooled production estimate from -2.6 to -3.6 percent. Dropping Poland and using the 2014 wave change nothing material. The German tolling treatment raises the German production estimate from -2.2 to -2.7 percent in 2020 and from -1.5 to -2.4 percent over the full window. No step moves the within-division estimates away from zero, and none makes the employment estimate without textiles significant. We adopt the first two changes as the baseline, since both move the European variable toward the U.S. definition. The third and fourth are reported as robustness.

German comparison. Measured before 2020, the harmonized and national German intensities rank four-digit industries at Spearman 0.96, and at 0.89 when the harmonized variable is divided by value added. They also agree in level, both averaging about 2.8 percent of production value. That agreement suggests Destatis compiles the German V23110 from the same cost-structure item. The comparison therefore shows that the harmonized variable records tolling in Germany; it cannot show the same for Italy, where subcontracting to specification is likely larger. An earlier version of this check used the 2020 national cross-section and gave a Spearman of 0.77. For a pre-determined treatment, the pandemic year is the wrong reference.

C.5 The European employment panel

Eurostat’s business statistics report persons employed at NACE four-digit detail in both statistical frameworks, V16110 through 2020 and EMP_NR from 2021, on the same definition. Joined, they give an annual four-digit employment panel for 2012–2024 in the 23 countries with at least 80 treated industries (22 report 2024). We keep industries observed in every year a country reports and use the baseline treatment. Three features of the panel shape the estimates in Table A11.

The 2021 framework change. The 2021 business-statistics reform changed the population of reporting units. Country-by-year effects absorb a uniform shift at the seam, but the within-industry change from 2020 to 2021 is mildly correlated with reliance (-0.010 per standard deviation,

Table A10. Making the European treatment comparable to the U.S. item, step by step

	(0) Value-added denominator, NACE 10–33	(1) Drop NACE 33 repair	(2) Production- value denominator	(3) Drop Poland	(4) 2014 wave for IT, FR, FI	(5) German national tolling
<i>Panel A. Acute window, 2018–2020</i>						
Employment, all countries	−0.88*** (0.34)	−1.24*** (0.38)	−1.22*** (0.36)	−1.21*** (0.36)	−1.20*** (0.36)	−1.20*** (0.36)
Employment, within division	−0.42 (0.34)	−0.58 (0.43)	−0.28 (0.38)	−0.31 (0.39)	−0.27 (0.38)	−0.25 (0.38)
Employment, without NACE 13–15	−0.28 (0.37)	−0.63 (0.46)	−0.59 (0.45)	−0.59 (0.46)	−0.57 (0.45)	−0.57 (0.45)
Production, five countries	−2.72*** (0.85)	−2.61*** (0.87)	−3.62*** (0.96)	−3.62*** (0.96)	−3.16*** (0.96)	−3.29*** (0.92)
Production, within division	1.59 (1.85)	1.68 (1.78)	1.28 (1.84)	1.28 (1.84)	1.29 (1.68)	1.51 (1.72)
Production, Germany	−1.48 (1.57)	−1.27 (1.40)	−2.20 (1.53)	−2.20 (1.53)	−2.20 (1.53)	−2.67* (1.43)
<i>Panel B. Full window, 2018 to latest year</i>						
Employment, all countries	−1.33** (0.53)	−1.96*** (0.57)	−2.33*** (0.54)	−2.30*** (0.55)	−2.27*** (0.54)	−2.29*** (0.54)
Production, five countries	−1.90 (1.22)	−1.89 (1.19)	−2.13* (1.18)	−2.13* (1.18)	−1.89 (1.23)	−2.12* (1.24)
Production, Germany	−1.08 (1.18)	−0.96 (1.07)	−1.54 (1.23)	−1.54 (1.23)	−1.54 (1.23)	−2.36* (1.32)

Notes: Percent change per one within-country SD of pre-COVID subcontracting intensity, interacted with post-2020. Steps are cumulative: each column adds its change to all columns to its left. Column (0) is V23110 over value added (V12150), 2017, NACE 10–33; column (2) is the baseline used in Section 6. Employment: Eurostat SBS persons employed, NACE four-digit, 23 countries (22 from column 3), pre-period screen of Appendix C.5; the similarity groups of Appendix C.5 are not used here. Production: STS production index, the five countries that publish it at four digits. Country-by-industry and country-by-year fixed effects (industry and year for Germany alone); “within division” replaces country-by-year with country-by-NACE-2-by-year effects. CRV1 on country-by-NACE-2 (NACE-2 for Germany alone). The full window ends in 2024 for employment and 2025 for production. Because the employment panel here keeps the industry set of the earlier build, column (2) differs slightly from Table A11 (−1.22 against −1.24 in 2020). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

$p = 0.001$). Estimates for 2018–2020 use the old framework alone and do not cross the seam, which is why we lead with them.

A pre-period screen. Without screening, the pooled estimate is −0.6 percent and insignificant. A handful of industries with explosive growth before 2020 drive that result, above all Hungarian battery manufacturing (NACE 27.20). It is Hungary’s most subcontracting-intensive industry, and its employment rose from 53 in 2018 to 4499 in 2021 as electric-vehicle battery plants opened. We drop, in every country, industries whose employment moved by more than 50 log points between 2018 and 2019. The rule uses only pre-2020 data and removes 65 of about 3270 country-industries. We chose it after inspecting the Hungarian case, so its sensitivity matters: every cutoff from 25 to 150 log points gives −1.0 to −1.2 percent across the 23 countries, each significant at 1 percent, and dropping only the Hungarian battery industry gives −1.2 percent.

Pre-trends, hours, and similarity. Extended back to 2012, the screened panel does not reject flat pre-trends as a whole ($p = 0.43$), but it does reject for the ten countries most similar to the United States ($p = 0.004$). Removing each country’s pre-2020 reliance trend, estimated on 2012–2019 and projected forward, leaves the pooled estimate at -1.0 percent in 2020 and -1.7 percent over the full window. A separate linear trend for every industry gives identical coefficients, because in a balanced panel only the part of each industry’s trend that lines up with reliance can move the estimate. Hours worked (V16130) move one for one with headcount, so hours per person show no response to reliance (-0.0001 , standard error 0.0008) and cannot reveal whether short-time work cut hours while preserving jobs. Finally, we rank the 18 countries without a four-digit production index by the rank correlation between their division-level subcontracting intensity and the U.S. division profile, which uses treatment data only. The ten most similar countries show -1.6 percent in 2020 and the eight least similar -0.5 percent. The continuous triple difference is not significant (-0.4 percent per standard deviation of similarity, standard error 0.3), so we treat the split as descriptive.

Table A11. The European employment panel

	(1) All 23, no screen	(2) All 23	(3) 10 most U.S.-like	(4) 8 least U.S.-like	(5) Production- index countries: employment	(6) Production- index countries: production
Baseline, 2020	−0.65 (0.63)	−1.24*** (0.35)	−1.58*** (0.55)	−0.48 (0.52)	−1.18 (1.04)	−3.59*** (0.94)
Baseline, 2018–2024	−1.42* (0.85)	−2.25*** (0.52)	−3.14*** (0.75)	−1.01 (0.87)	−2.25 (1.46)	−1.84* (1.07)
Pre-2020 reliance trend removed, 2020	−0.31 (0.70)	−0.96** (0.41)	−1.13** (0.56)	0.11 (0.63)	−0.91 (1.16)	−3.11*** (0.91)
Pre-2020 reliance trend removed, 2018–2024	−0.73 (0.95)	−1.65*** (0.64)	−2.19*** (0.73)	0.01 (1.18)	−1.66 (1.68)	−1.19 (1.02)
Pre-trend test, p	0.628	0.434	0.004	0.328	0.262	0.001
Countries	23	23	10	8	5	5
Country-industries	3268	3203	1364	907	663	663

Notes: Percent change per one within-country SD of subcontracting intensity (baseline definition) interacted with post-2020. Outcome in columns (1)–(5) is log persons employed at NACE four-digit (SBS V16110 through 2020, EBS EMP_NR from 2021); column (6) is the log STS production index on the same country-industry-years as column (5). Columns (2)–(6) apply the pre-period screen of Appendix C.5. Columns (3) and (4) split the 18 countries without a four-digit production index by the similarity of their division-level subcontracting profile to the U.S. profile. Country-by-industry and country-by-year fixed effects, CRV1 on country-by-NACE-2. The trend rows remove, within each country, the linear trend in the outcome associated with reliance, estimated on 2012–2019 only and projected forward; the rows for a separate linear trend in every industry are identical by construction and omitted. The pre-trend test is the joint F -test that the 2012–2018 event-study coefficients are zero. Counts are for the 2020 baseline. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.6 Composition of the European result

Two checks separate contracting from industry composition, and Table A12 runs both in each region. One adds country-by-division-by-year fixed effects in Europe and NAICS-3-by-year effects in the United States, the closest U.S. analog of a NACE division, so that only industries inside the same broad sector are compared. The other drops textiles, apparel, and leather (NACE 13–15; NAICS 313–316).

Within divisions, both European outcomes are near zero, with standard errors close to their baseline values, so these are informative nulls rather than a loss of power. Textiles, apparel, and

leather contribute about three-fifths of the European employment estimate. Without them the employment estimate is insignificant, and the 2020 production estimate survives at -2.8 percent while the full-window production estimate does not. The U.S. panel behaves differently on the second check. There textiles, apparel, and leather contribute about a fifth of the 2020 shipments estimate, and dropping them raises it. Within three-digit sectors, the U.S. estimate for 2020 alone is -0.8 percent and insignificant, while over 2018–2023 it is -1.7 percent, and -2.6 percent without textiles.

Two readings fit the European null within divisions. Italy’s decree of 22 March 2020 suspended most of NACE 13–15 by four-digit code while exempting nonwovens (13.95), technical and industrial textiles (13.96), and workwear (14.12), so within those divisions the 2020 variation partly reflects which classes the decree allowed to operate.³⁷ A broad concept measured with error also loses more signal within divisions, where true differences in reliance are smaller. On either reading, the European evidence supports the between-industry claim of Section 6 and leaves the within-sector claim to the U.S. panel after 2020, where Table A7 shows that it depends on how customer exposure is measured.

Table A12. Contracting or composition: the same two checks in both regions

	Baseline	Within sector or division	Without textiles, apparel, leather
<i>2020 (acute window)</i>			
U.S. shipments	-2.26^{***} (0.63)	-0.83 (0.84)	-2.53^{***} (0.87)
U.S. employment	-0.94^{**} (0.44)	0.12 (0.42)	-0.79^* (0.46)
EU production, five countries	-3.62^{***} (0.96)	1.28 (1.84)	-2.84^{***} (1.03)
EU employment, 23 countries	-1.24^{***} (0.35)	-0.44 (0.38)	-0.64 (0.43)
<i>Full window</i>			
U.S. shipments	-2.59^{***} (0.72)	-1.72^* (0.88)	-3.16^{***} (0.87)
U.S. employment	-2.48^{***} (0.68)	-0.93 (0.68)	-2.68^{***} (0.73)
EU production, five countries	-2.12^* (1.18)	4.01 (3.30)	-0.14 (0.91)
EU employment, 23 countries	-2.25^{***} (0.52)	-0.78 (0.56)	-0.91 (0.63)

Notes: Percent change per one SD of pre-COVID reliance interacted with post-2020. U.S.: NAICS-6 panel, single-treatment specification, NAICS-6 and year fixed effects, CRV1 on NAICS-6; “within sector” adds NAICS-3-by-year effects (NACE-division-by-year effects, from the concordance of NAICS-3 to NACE divisions, give estimates within 0.01 of these); the textiles group is NAICS 313–316; the full window is the paper’s 2018–2023. Europe: baseline treatment; production is the STS index for Germany, Italy, France, Spain and Greece through 2025; employment is the screened 23-country panel of Table A11 through 2024; “within division” uses country-by-NACE-2-by-year effects; the textiles group is NACE 13–15. CRV1 on country-by-NACE-2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

³⁷Decree of the President of the Council of Ministers of 22 March 2020 and its annex of permitted activity codes. Export-oriented firms in these divisions could reopen on 27 April and all manufacturing on 4 May.

D Summary statistics

Table A13. Summary statistics, NAICS-6 manufacturing panel

	Mean	SD	p10	p50	p90
<i>Panel A. Industry-year panel, 2005–2024</i>					
Contract-production reliance (CPR)	0.012	0.015	0.001	0.007	0.029
TFP (NBER-CES, 2012=1)	1.00	0.12	0.87	1.00	1.10
Non-production worker share	0.31	0.12	0.18	0.28	0.46
log employment (thousands)	3.0	0.9	1.9	3.0	4.2
<i>Panel B. Pre-COVID industry cross-section</i>					
Contract-production reliance, 2015–19 mean	0.012	0.016	0.001	0.007	0.028
China-input exposure, 2015–19 mean	0.067	0.119	0.000	0.019	0.179
Import penetration of intermediates, 2015–19 mean	0.230	0.232	0.000	0.166	0.541
Inventory intensity, 2015–18 mean	0.133	0.065	0.061	0.128	0.207
Teleworkable employment share, 2019	0.222	0.135	0.110	0.171	0.399
Physical proximity (O*NET, 1–5), 2019	3.25	0.11	3.13	3.25	3.36
Establishments per industry, 2015–19 mean	810	1651	100	395	1501
Industries: 358 Years: 2005–2024 (2017 gap) Industry-years: 5584					

Notes: Contract-production reliance is the cost of contract work over shipments. Panel A pools industry-years; TFP is NBER-CES through 2016, the non-production share through 2018, and employment is in thousands. Panel B reports each treatment and control in the pre-determined cross-sectional form in which it enters the COVID design (Section 4): China-input exposure and import penetration of intermediates are built from the HS10-to-NAICS6 concordance of Pierce and Schott (2012); inventory intensity is NBER-CES inventories over shipments; the teleworkable share and the physical-proximity index weight OEWS occupation mixes by Dingel and Neiman (2020) and Mongey et al. (2021) scores and match 249 of 358 industries at NAICS-4 (the estimation merges add coarser fallbacks); establishment counts are County Business Patterns.

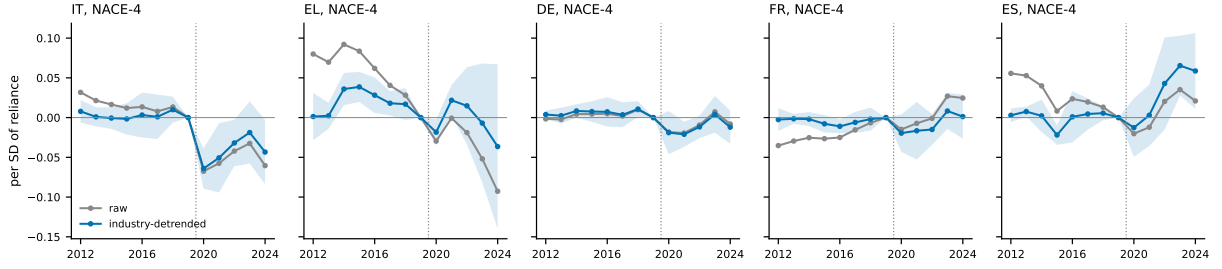


Figure A1. European event studies from 2012, raw and industry-detrended

Notes: Coefficients on pre-COVID subcontracting intensity (one within-country SD, Eurostat SBS 2017 wave over production value) interacted with year dummies, base year 2019, for the five countries that publish the STS production index at NACE four-digit; log production index, industry and year fixed effects, 95% intervals clustered on NACE-2. Pre-years run 2012–2018. The detrended series removes each industry’s own linear 2012–2019 trend before estimation. The pooled five-country estimate at the 2018 base opens at 2020 (-0.031 , $t \approx 3.5$) and closes by 2023; the sixteen-country division-level extension is in Section 6.

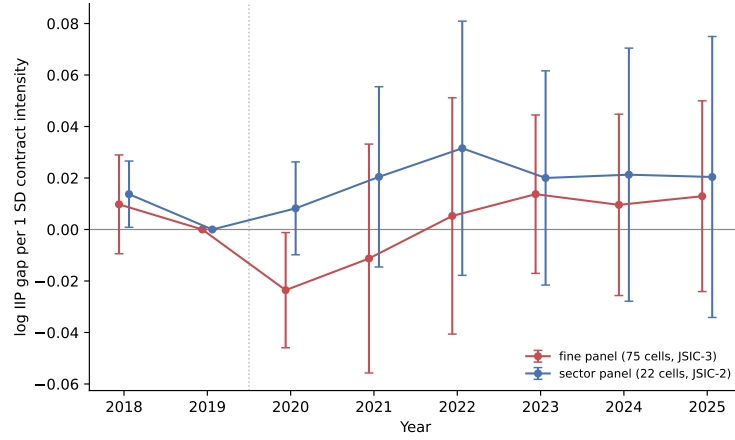


Figure A2. Japan event study

Notes: Coefficients on pre-COVID consigned-production intensity (one SD, final Census of Manufactures, 2019) interacted with year dummies, base year 2019; log METI production index, 95% intervals clustered on JSIC-2. The fine panel (75 cells, JSIC-3 concordance) shows the 2020 break and recovery; the 22-cell sector panel, which retains only between-sector variation, is flat—the identifying variation is across industries within sectors, as in the U.S. panel (Section 2).

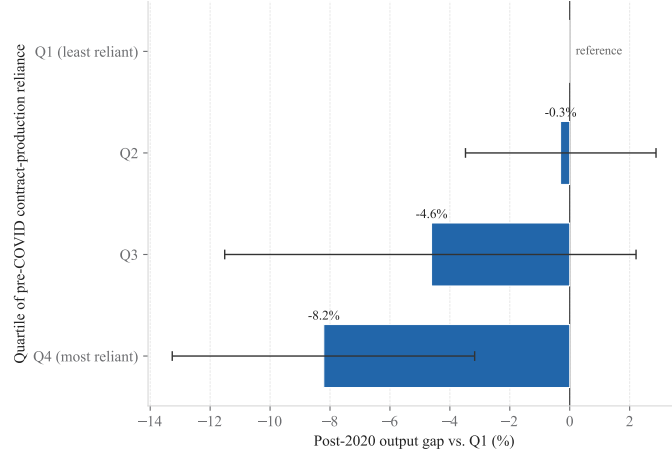


Figure A3. Dose-response: post-2020 output gap by pre-COVID reliance quartile

Notes: Post-2020 log-output gap relative to the least-reliant quartile (Q1), from equation (2) with quartile-by-post interactions, 2018–2023. Whiskers are wild cluster bootstrap 95% intervals (21 NAICS-3 clusters). Only the top quartile is individually significant under the primary NAICS-6 clustering, and no quartile is individually precise once whole NAICS-3 sectors are resampled (Appendix A), which is why the exhibit is read for the direction of the gradient rather than for its curvature.

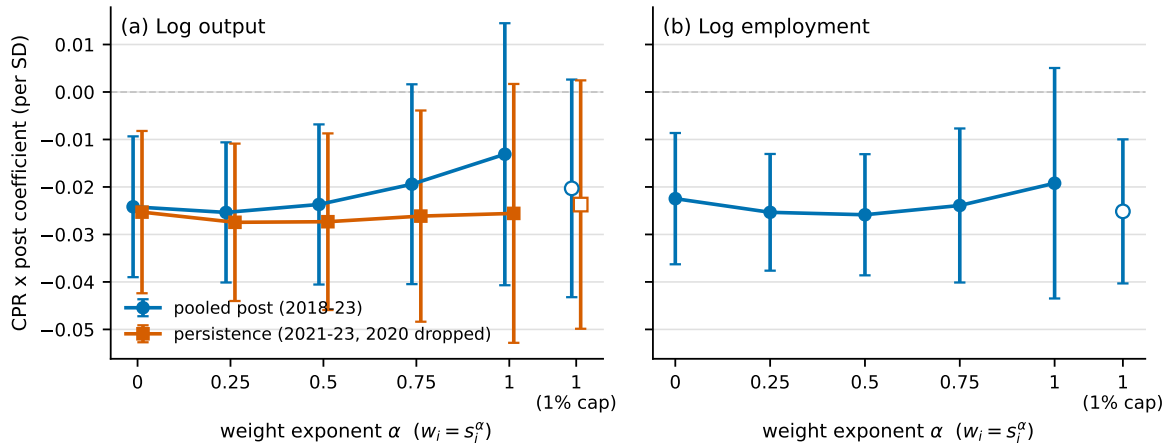


Figure A4. The weighting path: from the average industry to the average dollar

Notes: The $CPR^{pre} \times post$ coefficient from equation (2) under weights $w_i = s_i^\alpha$, where s_i is 2019 shipments and α runs from 0 (the unweighted baseline) to 1 (dollar weights); whiskers are 95% intervals clustered at NAICS-6. Hollow markers cap any industry at one percent of total weight: the uncapped dollar weights concentrate (Kish effective $N \approx 61$; petroleum refining alone is a tenth of the weight), which widens the intervals without moving the persistence contrast. The acute 2018–2020 output cell is not drawn: under full dollar weights its nominal-shipments coefficient is an uninformative +0.023 (standard error 0.037), the footprint of the 2020 petroleum price collapse in the least reliant giant industries (Appendix A).

Data and Code Availability

The estimation panel is built from public data. Contract work, shipments, employment, and value added come from the Census Bureau’s Annual Survey of Manufactures, 2022 Economic Census, and 2023 and 2024 Annual Integrated Economic Survey (api.census.gov); establishment counts from County Business Patterns, establishment entry and exit flows from Business Dynamics Statistics, and competing-goods imports by NAICS from the international-trade series, via the same API and census.gov flat files; productivity and real quantities from the NBER-CES Manufacturing Industry Database (data.nber.org); producer prices and the occupation-based teleworkability and physical-proximity inputs from the Bureau of Labor Statistics and the O*NET database (Work Context, physical proximity); administrative employment from the BLS Quarterly Census of Employment and Wages (national annual averages by six-digit industry through 2025, open-data area slices), end-of-year inventories from the AIES historical timeseries (2012–2024), payroll employment from the BLS Current Employment Statistics API, and the monthly G.17 industrial production indexes of the Federal Reserve via FRED; input–output structure, the imported-input content built from the 2017 benchmark detail Use table, and the real final-demand component paths behind the demand-rotation controls (NIPA Table 1.1.6) from the Bureau of Economic Analysis; R&D intensity from the National Science Foundation Business Enterprise R&D survey; the supply-chain pressure index from the Federal Reserve Bank of New York; and upstream propagation weights from the ADB Multi-Regional Input–Output tables. The HS10-to-NAICS6 concordance is from Pierce and Schott (2012). The cross-country panels use only free, openly downloadable sources: Eurostat Structural Business Statistics (`sbs_is_subc_r2`, `sbs_na_ind_r2`), the Eurostat short-term-statistics production index (`sts_inpr_a`), and the balance-of-payments services detail (`bop_its6_det`) via the Eurostat API; Japan’s Census of Manufactures and METI production index via anonymous e-Stat file downloads (`statInflId 000032112513` and `000040172370`); the Destatis cost-structure survey (GENESIS table 42251-0006) for the German validation; Mexico’s EAIM annual manufacturing survey, EMIM volume indexes, and IMMEX program statistics via INEGI’s open-data archives (`eaim_anual_2018_csv`, `emim_indices_nac_csv`, `immex_manu_act_csv`); and the OECD–WTO BaTIS balanced services panel for the Mexican balance-of-payments verification. The receipts-side validation uses the Economic Census NAPCS product-line APIs (`ecnnapcsprd 2017` and `2022`; `ecnipa 2012`). The firm-based import cross-check in Section 3.1 uses the Census experimental Trade by Industry and Product Statistics (imports by four-digit NAICS, 2021–2023). The headline contract-production-reliance results use only these public sources. The foreign-input-exposure channel is constructed from bilateral trade flows licensed from Trade Data Monitor; rebuilding it from public CEPII BACI flows (U.S. HS-6 imports mapped to NAICS-6 through the same Pierce–Schott crosswalk) reproduces it closely. The BACI- and TDM-based industry measures correlate at $r = 0.99$, and re-estimating the two-channel model on the BACI measure leaves the contract-reliance coefficient at -0.024 and the foreign coefficient at -0.020 (against -0.018 from TDM), so the headline does not depend on the licensed source. Census, BLS, and BEA pulls require free API keys, documented in the replication repository.

Replication code (data construction, estimation, and figures) is available from the author on request and will be posted to a public repository upon publication. The Census/BEA/BLS/NBER-CES/NSF inputs are redistributable or freely downloadable; the Trade Data Monitor extract is not, and the repository documents how to reconstruct the foreign-exposure variable from Comtrade or BACI.