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

Philip A. Luck†

First draft: March 2026
This draft: July 2026

Abstract

Resilience policy seeks to bring production home, largely through import restrictions. We show that much of the domestic vulnerability to supply-chain disruption sits on an organizational margin those restrictions do not reach: reliance on outside contractors for production steps performed on materials a manufacturer owns. On a panel of 358 U.S. manufacturing industries for 2005–2023, this contract-production margin is predominantly domestic and nearly orthogonal to import penetration. It also carries a real cost. When COVID-19 hit, the more reliant half of manufacturing contracted about 6.4% relative to the rest ($p < 0.001$), or 2.4% per standard deviation ($p = 0.002$), with inference clustered at the six-digit level at which reliance varies. Coarser clustering leaves the estimate significant (wild cluster bootstrap $p = 0.029$ at NAICS-4, 0.022 at NAICS-3); only tests that treat the 21 three-digit sectors as the sample are borderline ($p = 0.068$), so we read the estimate as a bounded output cost. The loss fell on volume rather than prices, persisted through 2023, left a durable establishment shortfall running through depressed entry, and appears with the same sign in five European countries and Japan. A foreign-input channel of comparable size is identified in the same regression, and it is the one the import lever reaches. The cost is contingent: in normal times contract work is an ordinary input of about 1.6% of cost, with no effect under diffuse supply-chain pressure and a non-negative response in the demand-driven 2008–09 recession. That contingency also bounds the policy reading: the result identifies a vulnerability to broad, synchronized shocks, while the targeted disruptions that motivate reshoring—decoupling from one country, a strait closure, an input embargo—are the low-synchronization case, where the framework predicts the sign reverses.

JEL Classification: L23, L60, F14, D24, E32.
Keywords: contract manufacturing, firm sourcing, supply-chain resilience, outsourcing, productivity, COVID-19.

1 Introduction

Reshoring has become a standard policy prescription for supply-chain fragility, and the instrument that carries it is the import restriction. Yet on the shock that motivated the prescription, the industry attribute that best sorts which U.S. manufacturers lost output in 2020 is not exposure

*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.

to imports. It is how much of its production an industry hands to outside contractors, a margin the reshoring debate does not measure. When a manufacturer needs a production step performed, it faces a classic make-or-buy choice: carry it out in house, or pay an outside contractor to do it. The Census records the second 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 an industry's *contract-production reliance* (CPR). It is a specific sourcing margin, distinct both from buying finished inputs and from owning a supplier.

This margin carries a real cost under disruption. When COVID-19 hit, the more reliant half of manufacturing contracted about 6.4% relative to the rest ($p < 0.001$), or 2.4% per standard deviation ($p = 0.002$, clustered at the six-digit level at which reliance varies); the estimate survives every coarser clustering level and is borderline only when the 21 three-digit sectors are treated as the sample, so we read it as a bounded output cost.

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.

This paper makes 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 the import-reduction lever that dominates resilience policy[1] reaches it only incidentally. 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 framework below predicts the sign reverses, so a vulnerability of the kind found here need not arise. To test these arguments, we build a panel of 358 six-digit NAICS manufacturing industries observed annually from 2005 to 2023, joining contract-work and real-activity data from the Annual Survey of Manufactures, the 2022 Economic Census, and the 2023 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.

A simple framework formalizes the trade-off (Section 5.2). 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 (Proposition 1). 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.

Three results follow, each with a stated limit. 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% of total contract-work spending; because the filing is optional and duty-driven, that share is a floor,

[1] By 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).

and the partner-mirror and receipts checks point the same way. A shift-share design built on foreign suppliers' export capacity to other high-income countries finds no evidence that foreign sourcing displaced the margin, though it is weakly identified and corroborates the domesticity evidence rather than carrying it. Reshoring policy that targets imports therefore does little about the vulnerability domestic contract manufacturing carries.

Second, contract-production reliance carries a real output cost under synchronized disruption. We take pre-pandemic reliance as a pre-determined treatment and interact it with the 2020 shock: more reliant industries lost 2.4% more output per standard deviation, and the loss concentrated at the top of the distribution, where the above-median contrast is 6.4%. Pushed inside three-digit sectors, the within-sector floor, the estimate is 1.8%; under the full composition controls, 2.8%. The contraction runs entirely through physical volume rather than profit margins, relative prices, or per-worker productivity. It persists, remaining visible through 2023, and leads to a subsequent fall in establishment counts that a decomposition attributes to depressed entry rather than elevated exit: the lost capacity is durably not replaced.

This domestic-contracting channel is identified *alongside* a foreign-input-dependence channel of comparable size; the two exposures coexist, and import-focused policies act on only one. Benchmarked against every measured exposure on the same panel and shock, contract reliance accounts for more of the cross-industry dispersion in 2020 output losses than import penetration or teleworkability, even though the import channel carries the larger per-standard-deviation response (Section 4.6). Nor is the pattern an American peculiarity: the same empirical design yields negative acute-2020 estimates across five European countries and Japan.

Third, the cost appears only under acute, synchronized shocks. In normal times, contract work is an ordinary, minor input that imposes no steady-state performance drag. It accounts for only about 1.6% 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 is therefore contingent on acute, synchronized supply-chain collapse; it is absent in calm times, under diffuse, idiosyncratic supply-chain pressures, or during the demand-driven 2008–09 recession.

Three boundaries guide the reading. First, the output cost is identified from a single, highly synchronized shock; because the margin shows no effect under continuous, low-level supply-chain friction, we bound the claim strictly to acute, systemic disruptions. This bound matters for policy. The targeted disruptions that motivate reshoring are the opposite, low-synchronization case, and the framework predicts the contract-reliance sign reverses there (Proposition 1; Section 6); the vulnerability documented here belongs to pandemic-type events and, on the paper's own logic, does not govern the geopolitical contingencies the reshoring debate is mostly about. Second, the cost of contract work captures a specific organizational margin, tolling and commission work on owned materials, and not the entire make-or-buy boundary, part of which is recorded as material input purchases. Third, because contract-reliant industries cluster at the standardized, low-R&D end of manufacturing, an industry-level design cannot cleanly separate the organizational frictions of the contracting relationship from the baseline fragility of the commodity production typically routed to those contractors. A battery of empirical horse races narrows this reading but cannot fully resolve it; the final separation remains a task for firm-level microdata.

Related literature. The paper's measure sits in the firm-boundaries and domestic-outsourcing tradition (Antràs and Helpman, 2004; Grossman and Helpman, 2002; Antràs and Chor, 2013; Alfaro et al., 2019; Atalay et al., 2014; Goldschmidt and Schmieder, 2017); the make-or-buy boundary we read this margin against is 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, and not a measure of the vertical boundary. The mechanism 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 cost-minimizing sourcing can build networks that are efficient and fragile at once (Elliott et al., 2022), and on the theory of supply-chain resilience as an under-provided investment (Grossman et al., 2024). 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. The two exercises are complements rather than substitutes, and it is worth being precise about how. 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. Neither delivers the other: 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. Because the persistence documented below runs through establishment entry rather than exit, the paper also touches the literature on business dynamism and firm formation, in which entry rates are a slow-moving margin that responded sharply and positively to the pandemic (Dinlersoz et al., 2021; Decker and Haltiwanger, 2023); the contract-reliant shortfall is a deviation from that aggregate. The redundancy-versus-flexibility and correlated-supplier-risk trade-offs at the center of the mechanism 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). 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 its empirical counterpart, placing the domestic vulnerability on a margin the import lever does not reach (the broader reallocation debate is surveyed by Alfaro and Chor, 2023). The framework in Section 5.2 adapts a relocation-option logic in the spirit of Bergin et al. (2009), in which outsourcing is valuable because assembly can be shifted across locations in response to shocks (closest in structure to the portfolio diversification of Acemoglu et al., 2012), and augments it with a synchronized shock that disables that option. 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. 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 shows why that cost is contingent on the synchronization. The two margins are separable in time as well as in the cross-section: the domestic response arrives in 2020–2021 and the import channel's in 2022–2023. Within the domestic response, the acute loss is production at establishments that already existed, and only later does it become a shortfall of establishments that were never opened. The measurement side is reusable independently of the shock: Section 3.1 establishes the margin's domesticity through four checks that other work can apply to the same Census item, a customs-based floor on offshore processing, a European balance-of-payments mirror, a Mexican mirror, and the receipts side of the Economic Census.

The shock evidence joins fast-growing work on supply chains and the pandemic. In Guerrieri et al. (2022), a sectoral supply contraction of the pandemic's kind can induce a demand shortfall larger than the shock itself, so the aggregate response is not bounded by the size of the disruption; the cross-industry gradient we estimate is one input into which sectors that logic starts from. Lafrogne-Joussier et al. (2023) show that French firms exposed to China's early lockdown lost sales, and 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. Bonadio et al. (2021) quantify the cross-country transmission of the pandemic through global supply chains in a many-country model; the margin studied here is the within-border counterpart their framework does not separate. 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. Relative to Fort (2017), who identifies ICT as a determinant of domestic versus foreign fragmentation, Wu et al. (2023), who study contract manufacturing and steady-state productivity, and Li et al. (2025), who tie vertical-integration choices to how disruptions reach firm value, the contribution is a shock-identified consequence of contract reliance together with a production-function account of why it is contingent.

Section 2 builds the panel and defines contract-production reliance, which turns out to vary far more across industries than within them. Section 3 establishes the margin's domesticity. Section 4 is the paper's core, the COVID stress test, with the European and Japanese replications in Section 4.7. Section 5 shows that contract work is an ordinary input in calm times and develops the framework that orders shocks by how synchronized they are. Incidence and policy are taken up in Section 6, and Section 7 concludes.

2 Data and Measurement

The estimation panel is a set of 358 six-digit NAICS manufacturing industries observed annually from 2005 to 2023, 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 ASM covers 2005–2021 through two API endpoints; the 2022 Economic Census covers the 2022 benchmark year; and the Annual Integrated Economic Survey, which absorbed the ASM beginning with reference year 2023, covers 2023. Contract-production reliance is

$$\mathrm{CPR}_{it} = \frac{\text{cost of contract work}_{it}}{\text{value of shipments}_{it}}. \tag{1}$$

The ratio answers a simple 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. We verify this directly: the 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, so the years pool under industry and year fixed effects. The 2023 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.

Eight industries, including automobile and petrochemical manufacturing, have fully suppressed 2023 survey records and drop from the 2023 outcome sample.

What the measure captures, and what it does not. 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, not as the inverse of vertical integration. 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. They cannot contaminate the panel, though the measure understates outsourcing at that extreme. Bounding how many producers sit near that boundary is a firm-level question this industry panel cannot answer, and it bears only on the most extreme cases. The truncation runs in the direction Houseman et al. (2011) identify for these data more generally: when production moves outside the surveyed frame, manufacturing statistics understate the extent of sourcing and overstate the output and productivity of what remains. Both biases work against finding the gradient we find, since the industries most likely to cross into the factoryless frame are the contract-reliant ones.

The item also counts work performed by *other establishments of the same firm*, so it 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. Two facts bound the concern. 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. A direct multi-unit-versus-single-unit split of contract-work spending, which would size the intra-firm share, is a firm-level tabulation left to the Census microdata agenda.

Other measures. Total factor productivity through 2016, and real output, employment, materials, and value added through 2018, come from NBER-CES; real activity through 2023 comes from the Census instruments above. R&D intensity is proxied by the non-production worker share and validated against NSF Business Enterprise R&D expenditure (National Center for Science and Engineering Statistics, 2024). Foreign-input exposure is the NAICS-6 import penetration of intermediate inputs, built from the HS10-to-NAICS6 concordance of Pierce and Schott (2012), with a China-specific component used below. Table 9 reports summary statistics. One magnitude recurs throughout: contract work is a minor input, averaging about 1.6% of total manufacturing cost (the same spending is 1.2% of shipments in Table 9); Section 5.1 develops what follows from that share.

A cross-sectional industry attribute. A variance decomposition of contract-production reliance (Table 1) attributes 81% of its variation to differences across industries and only 19% to within-industry movement; the largest single component, 50%, 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, while the credible variation is cross-sectional or comes from a sharp shock. It is the reason the central design below uses pre-shock reliance as a cross-sectional treatment.

3 A Domestic Margin, Orthogonal to 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 The margin is domestic

The Census definition does not restrict the contractor's location, so we establish the margin's domesticity empirically, with five checks.

First, and remarkably, 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 (Appendix Figure A1).[2]

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 about \$2–3 billion of value added abroad per year over 2017–2021 (roughly \$4–5 billion of gross 9802 value), against roughly \$53 billion of total manufacturing contract-work spending, a directly measured offshore share of 5–9%. 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, 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, but the partner mirror 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% of contract-work spending. The receipts side adds up: domestic establishments *selling* contract manufacturing services reported receipts of \$42 billion in 2017 and \$47 billion in 2022 against contract-work purchases of \$49 and \$51 billion, positively rank-correlated across industries—though because seller receipts also include work for wholesale-classified factoryless principals, the match corroborates a domestic seller population of the right order rather than sharply bounding the offshore remainder. And the Mexican mirror addresses the largest offshore counterpart: Mexican contractors earn about 85% 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

[2] A 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.

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 (Appendix C).

These findings are consistent with Fort (2017), who finds domestic fragmentation to be the larger and more ICT-sensitive margin, and with the broader evidence that outsourcing within U.S. manufacturing is substantial and imperfectly captured in the standard statistics (Houseman, 2007; Houseman et al., 2011). The foreign-input margin is real and we estimate it directly (Section 4); it simply lives in a different variable.

3.2 No substitution away from the domestic margin

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—finds no such displacement: the point estimate runs the other way, with more China-sourced input availability associated with *higher* contract-production reliance. The first stage is weak (a NAICS-3-clustered F near 5.7), so we read it as evidence against substitution rather than as an estimated complement magnitude and report Anderson and Rubin (1949) weak-instrument-robust intervals: the share-level interval excludes zero under the original NAICS-3 construction ($[0.004, 0.254]$) but includes it when the instrument is rebuilt at NAICS-6 ($[-0.011, 0.144]$), so the test corroborates the domesticity evidence rather than carrying it.[3] Either way, there was no substitution for import restrictions to reverse, and the margin is in any case predominantly domestic, so the import lever acts mostly on the other exposure.

4 The Output Cost of Disruption

Contracting a production step out diversifies idiosyncratic supplier risk but loads on a common supply factor (Section 5 develops this), so 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 tests whether the domestic-contracting margin carried a real cost when it hit, sizes that cost against the other exposures competing to explain the same collapse, and asks in Section 4.7 whether the response is particular to the United States.

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:

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

[3] A Rotemberg-weight decomposition (Goldsmith-Pinkham et al., 2020) confirms China carries the identification (weight 1.16, own first-stage $F \approx 48$); 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 ($\widehat{\beta} = 0.039$). Full diagnostics are in the replication package.

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. One standard deviation of reliance is 1.6 cents of contract work per dollar shipped, against a median industry at 0.7 cents and a ninetieth percentile at 2.8 (Table 9). 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. The design uses the dominant cross-sectional variation in reliance as the treatment and the shock as the timing. 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). The above-median contrast (−6.4%; Table 3) and the quartile dose-response (Appendix B) report the same object without that restriction. Identification rests on conditional parallel trends (tested below), on contract reliance being pre-determined, and on 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 4.4. 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 4.4 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 Two exposure channels

Table 3 reports estimates of equation (2). Contract- reliant industries lost more output: 2.4% per standard deviation of pre-COVID contract reliance ($p = 0.002$; the conservative NAICS-3 wild bootstrap gives 0.022); on the linear counterfactual developed in Section 6, that differential comes to roughly 1.2% of manufacturing shipments per year, about $68 billion on the panel's $5.7 trillion 2019 shipments base. The foreign-input-dependence channel is comparable in size over the full window (−1.8% for output, −2.4% for employment), and the two exposures are identified separately in the same regression; because its timing differs from the acute reliance break (taken up below), we read γ as a full-window average rather than an acute-disruption response.[4] Both sourcing margins, contractors at home and suppliers abroad, shaped which industries held output. The contraction did not 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 4.4 shows.

The effect is persistent. Figure 1 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 remains negative in 2023; employment follows the same path. Estimating both channels' year interactions jointly shows the two exposures also move on different timetables: the acute 2020–21

[4]The acute import disruption is documented at monthly frequency: French firms exposed to China's early lockdown lost sales in the spring of 2020 (Lafrogne-Joussier et al., 2023), and U.S. production fell most where Chinese-input exposure was highest (Meier and Pinto, 2024). That disruption had largely reversed within 2020, and an annual panel cannot see a V-shaped within-year interruption, so the near-zero 2020–21 China event-study coefficients are consistent with it. What the annual data add is that the differential loss still visible at annual frequency in 2020–21 belongs to the organizational margin; the foreign channel's annual footprint arrives with the 2022–23 goods-cycle unwind.

break belongs to contract reliance alone (China-exposure coefficients of -0.005 and -0.002 in 2020–21, against -0.023 and -0.031 for reliance), the China-exposure loss arrives only in 2022–23, years whose attribution the demand-rotation checks of Section 4.4 complicate, and where the reliance pre-coefficients sit at zero the China series drifts upward in 2015–16 (estimates in the replication package). The two exposures are separable in time, not only in the cross-section.

County Business Patterns establishment counts give the persistence an extensive-margin counterpart. Because the counts reference March of each year, the 2020 count predates the lockdowns; it shows no differential movement (-0.6%, insignificant). The gap then opens and widens: contract-reliant industries had 1.3% fewer establishments per standard deviation by March 2021, 2.2% by 2022, and 2.5% by 2023, relative to 2019. Dating the post period at 2021 and dropping the transition year, the loss is 2.2% per standard deviation ($p < 0.001$; wild NAICS-3 bound 0.009), and it survives removing each industry's own pre-2020 trend (-1.4%, $p = 0.026$, with a flat detrended pre-period). The persistent output gap is therefore not only 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 nearly all of the remaining gap (Appendix B).

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 (-2.6% per standard deviation, $p < 0.001$ clustered at NAICS-4, the level at which its aggregated treatment varies), and the split then attributes the loss to entry: post-2021 entry rates fall by 0.46 percentage points per standard deviation of reliance ($p = 0.005$), about 6% of the 7.2% post-period entry rate for manufacturing as a whole. 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. Remarkably, the exit margin does not move—the exit rate is statistically flat ($p = 0.11$) 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 B 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 4.4: the gap survives that control essentially unchanged.

The persistence also holds outside the Census collection system. 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 2.3% per standard deviation in 2020 and stays 2.8–3.1% below its 2019 level through 2023 ($p \leq 0.002$ in every post year). 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 leaves a flat pre-period and a gap of 1.7–2.3% in every post year ($p \leq 0.017$), the numbers we quote. BLS payroll employment at the four-digit level moves the same way, drift included. Appendix B places this cross-system agreement, together with the one output measure that does not show the persistence, in a single measurement decomposition.

4.3 Where the contraction concentrates

The output cost is not spread evenly. Estimating equation (2) within the cells of an R&D-by-export-control-exposure grid, the point estimates concentrate in *low-R&D* industries at both levels of export-control exposure (-0.030 at low and -0.049 at high) and are near zero in high-R&D industries (Table 4); the heterogeneity runs on the R&D dimension, not on the export-control axis. The cells are not statistically distinguishable—the pooled triple interaction has $p = 0.64$—so we read the low-R&D concentration as interpretation: the commodity margin most easily contracted out. Reliance itself is essentially uncorrelated with R&D intensity across industries ($r = -0.02$), and a direct horse race leaves the reliance effect unchanged while R&D-by-post carries nothing (Section 4.4): the concentration marks where the reliance effect bites. This low-R&D footprint is what Section 6 takes up.

4.4 Robustness and inference

Inference: the clustering ladder. Clustering is a design choice, and we treat it as one (Abadie et al., 2023). The panel is the population of six-digit manufacturing industries, so sampling considerations do not force sector-level clustering; what remains is the correlation of treatment and residuals across industries, and 64% of the variance of pre-COVID reliance sits inside three-digit sectors. Primary inference therefore clusters at NAICS-6, the level at which the treatment varies, and Table 6 reports the full ladder of coarser choices, none suppressed. 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 $p = 0.001$ (Table 5), and within-sector randomization inference, which holds sector composition fixed, gives 0.030. (Unrestricted randomization inference gives 0.003, but it permutes reliance across all industries and so ignores the sector clustering of both treatment and shock; we do not lean on it.) Only procedures that treat the 21 sectors as the sampling unit are borderline: block-resampling whole NAICS-3 sectors for the heterogeneity-robust estimand gives $p = 0.068$, and Ibragimov–Müller, which t-tests the 21 sector-level estimates, does not reject ($p = 0.23$). The honest summary is that the output effect is significant at every clustering level up to the level of treatment assignment and borderline only when the sectors themselves are the sample; we therefore report a bounded output cost, and rest the strongest statements on the unanimous sign (all 21 leave-one-sector-out refits in $[-0.028, -0.020]$) and the above-median contrast (-0.064, $p < 0.001$; wild NAICS-3 $p = 0.007$). The employment effect is a shade weaker under the conservative bound (wild $p = 0.029$ at baseline, 0.072 under controls), so we treat output as the clean outcome and employment as corroborating. Robustness rows here and elsewhere continue to report the NAICS-3 wild bootstrap, the ladder's conservative rung. Table 5 collects the remaining checks and Appendix B reports them, with a reliance-quartile dose-response, in full.

Heterogeneity-robust estimand. The per-standard-deviation coefficient is a conservative linear summary; the magnitude is best read off the nonparametric dose-response. Because 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; the only continuous-treatment concern is that the per-standard-deviation slope averages those responses under a linearity restriction (Callaway et al., 2024). Estimating the Callaway et al. (2024) 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 (-0.027 in the acute window, where it is not individually significant, $p = 0.31$), larger in magnitude than the -0.024 per-standard-deviation slope because the loss concentrates at the top of the reliance distribution (quartile effects 0.000, -0.004, -0.047, -0.083). Resampling whole NAICS-3 sectors, this 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 dose-response *curvature* is not separately identified. What carries is the sign and the individually significant top-quartile and above-median (-0.064, wild $p = 0.007$) contrasts.

Within-sector identification. Half the treatment variation 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% per SD (-2.4% once composition controls are added). We read this within-sector estimate as the within-sector floor: it compares more and less reliant industries inside the same sector-year, not contract-heavy sectors against unlike ones, and the headline -2.4% is the between-inclusive upper estimate. Sector-by-year fixed effects push the identifying variation inside NAICS-3 clusters, where the 21-cluster wild bootstrap is known to lose power ($p = 0.21$); under the primary NAICS-6 clustering the estimate is significant ($p = 0.044$), at NAICS-4 it is borderline (CRV1 $p = 0.062$, wild 0.090), and within-sector randomization inference, the specification's design-based test, sits at the conventional threshold ($p = 0.055$). The within-sector result is thus significant under the primary clustering and at the margin under the coarser bounds. Pushing the fixed effects one level finer still (NAICS-4-by-year) attenuates the coefficient to -0.013 and out of significance at any clustering level. A split-sample check shows this 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 (first-stage $F > 190$) corrects classical measurement error. The halves correlate at 0.73 within four-digit families, implying a within-family reliability of 0.84: attenuation alone predicts a coefficient of -0.022, and the error-corrected estimates are -0.013 to -0.019 (the larger with $p = 0.082$). The within-family gradient is therefore genuinely flatter, and the between-family component of the variation carries a steeper slope; the finer contrast leaves the within-family commodity gradient of Section 4.5 doing real work at that margin. The same correction strengthens the within-sector floor, from -0.018 to an instrumented -0.024 ($p = 0.024$), so measurement error, if anything, understates it. That the effect holds in sign inside sectors, where the standardized, commodity character of a three-digit family is differenced out, is the strongest industry-level evidence that reliance is not merely a relabeling of sector type, though it does not isolate the contracting relationship from the finer within-sector commodity gradient (Section 4.5).

Composition. The first 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 unchanged and, if anything, larger (-0.024 to -0.028). Because the controls push the coefficient away from zero rather than toward it, its stability across specifications is itself the reassurance: selection on unobservables would have to run opposite to selection on observables to overturn it.[5] Three of these are what a demand- or lockdown-based alternative leans on most: the consumer-demand share, against a pure-demand reading; teleworkability, against the reading that these industries simply could not send their work home; and physical proximity (uncorrelated with reliance, $r = -0.09$, and null on its own), against the reading that their work happens at close quarters and was thinned by distancing.[6]

A direct control for the 2020 shutdown designation points the same way: classing industries essential or not under the CISA critical-infrastructure guidance and interacting essential status with the post period (and, in a second variant, with year fixed effects) holds the reliance effect at -0.022 (wild $p = 0.033$) while the essential interaction is itself null. Essential industries are, if anything, *less* contract-reliant ($r = -0.31$). The shutdown margin works against the estimate. A natural mediator, thin inventory buffers, runs the other way: reliant industries hold if anything *more* inventory before the shock, and the interaction leaves the effect intact (Table 5), so it does not account for the contraction. The index construction, the inventory coefficients, and the in-window inventory dynamics the pre-2018 series cannot reach are in Appendix B.

Industry type. That the loss concentrates in low-R&D industries (Section 4.3) raises a sharper version of the threat: that reliance 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, and racing the two interactions leaves reliance unchanged (-2.4%, wild $p = 0.023$) while R&D-by-post carries nothing ($+0.4\%$, $p = 0.73$). Under the full composition controls the R&D interaction even turns positive ($+0.036$, CRV1 $p = 0.008$; high-R&D industries fare relatively better once demand-side attributes are held fixed, mirroring the low-R&D concentration of Section 4.3) and reliance stays intact (-2.6%, wild $p = 0.047$). Pre-period industry scale is null and leaves the estimate unmoved (-2.4%, wild $p = 0.020$; Table 5), and a placebo sharpens the point: another minor input's cost share—energy, of comparable size—interacted with post-2020 yields nothing (-0.1%, $p = 0.95$). The contraction attaches to the contracted-production margin, not to low-R&D status, size, or small input shares in general.

Confound horse races. Four one-issue races, each a row of Table 5 with detail in Appendix B, close the remaining measurement readings. First, supply-chain position: adding a pre-determined upstream COVID-exposure control (the input-weighted 2019–2020 contraction of an industry's supplier sectors, ADB-MRIO, interacted with post-2020) leaves the effect at -0.021 (wild $p = 0.018$), and a finer NAICS-6 upstreamness measure (BEA Detail) is null. Network position does not carry the contraction. Second, embodied import content, the foreign content that reaches an industry indirectly through its domestic suppliers, measured from the 2017 benchmark Use table through the Leontief inverse, which cumulates the imports used at every upstream stage of the domestic chain: 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 (wild $p = 0.025$) while

[5] The Oster (2019) δ that would drive the coefficient to zero is -4.8 at $R_{\max} = 1.3\tilde{R}$, but the bound is sensitive to the assumed $R_{\max}$: -1.4 at $2.0\tilde{R}$, -1.0 at $2.5\tilde{R}$, and near zero at the absolute $R_{\max} = 1$. Because the controls barely move the coefficient or raise the within-R^2, the δ extrapolation is highly leveraged, so we lean on the direction and stability of the coefficient rather than on a single δ.

[6] A second composition question concerns the response rather than the sample: whether the loss is establishments producing less or fewer establishments. The two margins separate exactly, since log shipments are log establishments plus log shipments per establishment. In the acute window the intensive margin carries it, -0.018 of the -0.022 detrended response, while the establishment count does not move; the extensive share climbs to 79% by 2023. Appendix B reports the split and its limits.

the embodied channel itself is present and correctly signed (about -0.023, $p \approx 0.05$), so the two exposures are separable. Third, foreign-exposure measurement: substituting total import penetration of intermediates for the China measure leaves the contract-reliance coefficient unchanged (-2.4%, wild $p = 0.035$) and carries the foreign channel itself (-2.8%, $p = 0.001$); the total measure absorbs the China component, so the foreign side of Table 3 should be read as general import dependence. The source of that measure does not matter either. Rebuilding it from public CEPII BACI flows rather than the licensed Trade Data Monitor extract reproduces it at $r = 0.99$ and leaves the contract-reliance coefficient at -0.024. Fourth, real output: deflating by industry-specific producer price indexes (BLS PPI, 300 of the 358 industries) leaves the contraction essentially unchanged on the common sample, so it is a real-quantity effect, not a price artifact. Whether any residual sits in the contracting relationship or in an individual contractor's own sourcing is a firm-level question this industry panel cannot answer.

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 ($\bar{M} = 0.5$); the breakdown values are identical whether the event-study variance is the primary NAICS-6 clustering or the conservative NAICS-3 one. The employment effect, whose pre-period drifts up, does not survive the same relaxation in its first post-year—again the clean-outcome/corroborating split. Removing each industry's own linear trend (fit on 2015–2019 and extrapolated forward) gives a flat detrended pre-period and a break to -2.3% at the shock, -2.2% in the acute window and -2.5% over the full window (Table 5; Appendix B). The contraction is a break, not the continuation of a differential trend.

Demand rotation. The persistence through 2023 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. The composition controls interact fixed attributes with a single post dummy, absorbing a level shift but no post-period path, so two controls address the path directly: a realized demand-rotation index (each industry's 2017 final-use mix weighted by the corresponding national real final-demand paths) and the final-use shares interacted with year fixed effects (constructed in Appendix B). Reliance barely loads on the final-use composition, and realized rotation if anything ran *against* it, so neither control moves the estimate: the full-window coefficient goes from -0.024 to -0.023, and the drop-2020 persistence contrast from -0.025 to -0.024, with the 2021–2023 event-study coefficients unchanged (Table 5). Conditioning instead on the industry's own realized apparent-consumption path removes most of the coefficient, but that variable contains the outcome—total supply moved one-for-one with shipments and imports never filled the gap (Section 4.5)—so it conditions on the effect itself. The persistence survives the demand paths industries actually faced. The event-study timing points the same way: the reliance gap opens in 2020–21, before the rotation years, while the channel whose decline coincides with the rotation window is the import one (Section 4.2).

4.5 Interpretation

The cost is contingent on disruption, and the output swing it produces is large relative to the input's steady-state weight. A margin that averages just 1.6% of cost, with an output elasticity near 0.05 in normal times (Section 5.1), carries a point-estimated 2.4% output swing per standard deviation once the shock hits. That a 1.6-point input should move output by more than its own share is suggestive of complementarity between contracted and in-house tasks—a disruption to the contracted step propagating beyond the step itself—though the industry data cannot identify the elasticity of substitution at this resolution, so we read the magnitude as suggestive rather than as an estimated technology parameter and leave pinning it down to firm-level data.

The loss fell on volume, not on margins or productivity (Table 3), the split the framework delivers by construction: a disruption removes a fraction of contracted output at unchanged unit economics, so it surfaces as lost quantity rather than compressed margins (Panel B). These industries lost output they could not source or produce while the business they retained ran normally. The near one-for-one transmission of disrupted inputs to output documented for offshore linkages (Boehm et al., 2019) is the firm-level analog: an input that cannot be re-sourced quickly shows up directly as lost volume, whether contracted at home or imported from abroad. The two exposure channels in Table 3 are estimated separately; the reliance channel's acute timing is consistent with a contingent loss of the room to re-sequence and reprioritize that in-house production affords, while the foreign channel's loss arrives only in 2022–23 at annual frequency and pools disruption with the goods-cycle unwind (Section 4.2). 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 distinction is a firm-level question.

Relative prices complete the picture. Had contract-reliant industries been rationed while their customers' demand held firm, their prices should have risen; they did not move. Estimating equation (2) with the log industry producer price index as the outcome gives a precise zero (-0.1% per standard deviation, wild $p = 0.93$; -0.5%, $p = 0.34$, with the full controls), and the price event study is flat through the window (Appendix B). 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, consistent with the null here. That is what one expects at the low-R&D, standardized end of manufacturing, where the contraction concentrates (Section 4.3), and it makes the establishment losses above economically legible: 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 2023 (+0.001 per standard deviation, wild $p = 0.55$; the penetration series behind the foreign channel, available through 2021, agrees), and competing imports *fell* in the acute year (-3.7% per standard deviation, wild $p = 0.008$; -3.8%, $p = 0.064$, restricted to industries with import data in every year). The synchronized shock hit foreign producers of the same goods, which is what Section 4.7 documents directly.[7] Apparent consumption, shipments plus competing imports, falls roughly one-for-one with the output loss (-2.4% against a -2.8% shipments reference on the same panel, wild $p = 0.062$ and 0.043). 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.

[7]The 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).

4.6 The size of the effect in context

A per-standard-deviation coefficient is hard to weigh without a yardstick, so we benchmark it two ways, both on the same panel and shock with each standardized exposure entered one at a time, and Figure 2 reports the comparison. On a common footing, contract reliance moves output about as much as the foreign-sourcing exposure the pandemic literature has concentrated on (Lafrogne-Joussier et al., 2023; Meier and Pinto, 2024)—-2.7% per standard deviation, against -3.0% for total import penetration and -2.2% for Chinese-input exposure—through a margin that literature does not measure, and 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% of the total (against 1.6% for import penetration and 0.6% for teleworkability), the largest single measured contributor and 64% of what the three exposures jointly explain, and it ranks first whichever foreign measure it is benchmarked against. Entered jointly with foreign exposure, teleworkability, physical proximity, and upstreamness on one fixed sample, the reliance coefficient holds at -0.025 (wild $p = 0.021$).

Two limits discipline this. The three exposures together explain only about 6% of the cross-industry dispersion in 2020 output losses, so the shares above are shares of that explained portion and most of who lost output is left unexplained by any measured exposure. And the ranking is within a single shock: it speaks to which attribute best predicts who lost output in 2020, not to how contract reliance would rank under a disruption of another kind (Section 5 bounds that).

4.7 The same shock in other economies

The estimate so far rests on one country's statistical system, which raises a specific worry rather than a general one. The Census contract-work item is an establishment-level accounting category, and a gradient estimated from it could reflect how U.S. manufacturers book tolling arrangements rather than what happened to their production. The check that speaks to this is not replication for its own sake but re-estimation where the accounting differs. It differs sharply. 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 the European denominator is value added rather than shipments. An artifact of one collection convention should not survive a switch to the other. The design travels because it 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 value added at factor cost from the harmonized Structural Business Statistics (2017 wave, NACE four-digit), 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 (2019) separately reports the cost of consigned production on furnished materials—the closest foreign analog of the Census contract-work item—at four-digit industry detail, and the outcome is the METI production index over 75 hand-concorded industry cells. 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 differences, and what the fixed-effects structure does and does not absorb.

Figure 3 and Table 7 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 −2.7% per standard deviation ($p = 0.002$), Japan −2.8% ($p = 0.020$), both close to the U.S. −2.2%; Italy (−6.2%) and Greece (−3.5%) 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 (−2.2% pooled Europe, $t \approx 2.8$; −2.4% Japan, $p = 0.050$; Appendix Figures A2–A3). Where the economies differ is afterward: the U.S. loss persists through 2023, while the European and Japanese gaps close by 2022–2023, so the full-window estimates abroad are negative but attenuated (pooled Europe −1.6%, $p = 0.14$; Japan −0.6%, not significant). At the country level the full-window signs are mixed: Germany, Italy, and Greece stay negative while Spain and France turn mildly positive (+0.010, +0.006) as recovery overshoots the acute dip (Table 7), which is why the comparable object is the acute-window estimate, where every economy is negative. As 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 full-window estimate moves from $p = 0.14$ to $p = 0.066$, while Germany, Spain, and France remain individually imprecise at every clustering level (Appendix C).

The right reading stops at sign and shape. The pooled European full-window estimate is borderline and depends on Greece and Italy; dropping either roughly halves it. The pooled event study carries a small positive pre-period coefficient in 2018 (0.010). And Japan's acute estimate rests on the pooled fine-cell variation: restricting to the cells whose industry concordance is exact attenuates it to −1.7% ($p = 0.18$), and a between-sector-only panel shows nothing. Nor are the magnitudes comparable across countries, since the foreign subcontracting concepts are broader than the U.S. item and the institutional levels differ by an order of magnitude (Appendix C). What the exercise establishes is narrower than replication and more useful than it. Seven statistical systems, collecting subcontracting on different units of observation and against different denominators, produce the same sign at the same moment, and the enterprise-based European measure that nets out intra-firm tolling behaves like the establishment-based measures that keep it. A gradient generated by U.S. booking conventions would not do that. The estimate is still identified from one shock; what it is no longer identified from is one country's way of recording the margin.

Mexico offers a further, weaker check, and the one economy where both sides of the market are observed: the annual manufacturing survey reports what industries pay for contract production and what they earn performing it. That evidence stays out of the pooled exhibits—the payer-side treatment carries a pre-existing relative decline, so its acute estimate is directional only, and the seller-side split is carried by one sectoral bloc whose lockdown status is close to collinear with the treatment—but the payer-side reliance concept, averaging 1.2% of gross output and domestic to first order, verifies from both sides of the market (Appendix C).

5 Why the Cost Is Contingent

Why should a margin that is only about 1.6% of cost move output by more than its own share, and only under some shocks? Because in normal times it is an ordinary input, and its cost is contingent on how synchronized the disruption is. This section establishes both: 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.

5.1 An ordinary input in normal times

Is contract work simply an unproductive input, so that contract-reliant industries pay a steady-state penalty for using it? Two facts say no, and both say it is a minor input. Its cost share averages about 1.6% of total cost,[8] and its output elasticity is small in absolute terms: a gross-output production function that treats contract work as an endogenous input—estimated by the control-function approach of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015) cast as one-step GMM by Wooldridge (2009)—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). That 0.05, while small in absolute terms, runs about three times the 1.6% cost share, so contract work already contributes a little more to output than its expenditure share implies, a mild hint of complementarity between contracted and in-house tasks, though the industry data do not identify it (Section 4.5). A direct test reinforces the reading: regressing R&D on contract reliance is null across NSF R&D expenditure, the non-production share, and non-production employment, so the steady-state cost is not the capability atrophy a hollowing-out account would predict (Pisano and Shih, 2009). 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.

5.2 A simple framework

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 adapts the relocation-option logic of Bergin et al. (2009), 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, with Mexico/U.S. standard-deviation ratios of 2.03 on the production-worker wage bill and 2.21 on employment, against 1.53 and 0.77 for aggregate manufacturing. 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. Appendix A states the model formally and proves these results.

What the asymmetry stands in for. The inequality $\rho > \rho_h$ is a reduced-form stand-in that the industry data identify only in that form, and several distinct micro-mechanisms deliver it. A contractor may be shut down outright while the principal keeps partial in-house operation (a blackout); it may stay up but ration scarce capacity toward its own priority customers and away from these buyers (priority rationing); the principal may lack the process capability to bring the step back in house at all, which is often why the step was tolled to begin with (no in-house fallback); the disruption may reach the principal later and less visibly through a supplier it does not directly observe (a visibility lag); or the physical hand-off of work in process across sites may add freight and requalification legs that stall (added logistics). Each raises the contracted loss relative to the in-house loss under a common shock, and the reduced-form coefficient does not separate them.

One coarse discriminator is available. Aggregate receipts of the domestic establishments that

[8] Scaled by the value of shipments, which exceeds total cost, the same spending is the 1.2% reliance ratio of Table 9; each denominator is used where it is natural, shipments for the reliance measure and total cost for the input's expenditure share.

sell contract manufacturing did not collapse across the shock (\$42 billion in 2017, \$47 billion in 2022): the contractor population as a whole kept operating. That weighs the balance toward the rationing and capability mechanisms, in which the contractor stays up while the reliant buyer still loses output, and away from a pure blackout, and it fits the concentration of the loss in low-R&D industries with the least in-house fallback. Which mechanism dominates, and how intra-firm tolling (which the contract-work item does not separate from arm's-length work) figures in it, is a firm- and contractor-level question the industry panel cannot resolve (Section 4.5). These mechanisms have a natural home in the incomplete-contracts theory of the firm boundary (Grossman and Helpman, 2002; Antràs and Helpman, 2008), where non-contractible priority allocation and ex-post bargaining under an aggregate shock can lead an arm's-length contractor to ration differently than an in-house division. We do not derive $\rho > \rho_h$ from that primitive: the framework is a device that orders shocks by their synchronization, and we read only its sign predictions.

Formally (Appendix A), a producer contracts a share of its tasks across substitutable contractors and performs the rest in house, and disruptions carry a common (synchronized) component alongside idiosyncratic ones. A thick contractor market diversifies the idiosyncratic outages but not the common component, so expected contracted loss has a floor that a synchronized shock sets and that market thickness cannot lower (Lemma 1); we test that null below. Netting the diversifiable idiosyncratic risk against the contractor's excess exposure to the common factor, the marginal value of reliance in a thick market is $\theta - q(\rho - \rho_h)$. The first term is the in-house idiosyncratic risk a contracted task escapes; the second is the contractor's excess exposure to the common factor that the task takes on. The expression falls in the synchronization q and changes sign at the threshold $q^* = \theta/(\rho - \rho_h)$, positive below and negative above (Proposition 1). Reliance is a bet on how correlated disruptions turn out to be. This is the portfolio signature of outsourcing—it diversifies idiosyncratic supply risk but loads on the aggregate supply factor—and the sign-flip needs only that contractors are more exposed to the common factor than in-house production, not that in-house is immune to it. In words, routing a task to a contractor trades plant-specific risk for a share of the economy-wide risk. The mapping to the estimated shocks follows (Corollary 1): 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; and a demand-driven or idiosyncratic shock predicts a non-negative coefficient, the 2008–09 case.

Section 4 estimates the COVID case; the other two shocks are read through the same lens below. Because q is not measured directly, the framework *orders* the shocks by synchronization rather than testing a point prediction: the data line up with that ordering, and the one out-of-sample sign check the data permit—a low-synchronization shock that should turn the coefficient non-negative—is the 2011 Tōhoku episode, underpowered and reported only in Appendix B. Complementarity between in-house and contracted tasks would sharpen the prediction—the loss rising with reliance—and the quartile dose-response does rise with reliance, though the industry data do not separately identify the elasticity of substitution.

5.3 The framework's ordering in the data

Supplier-market thickness. The common-floor result (Lemma 1) predicts that a thick contractor market should not buffer the loss under a broad, synchronized shock, and it does not: interacting reliance with the pre-determined 2015–2019 mean count of establishments per NAICS-6 (County Business Patterns), the triple interaction is null (-0.006, $p = 0.24$), and the effect is -0.025 in thin against -0.027 in thick markets. Here the null is the prediction: thickness insures against idiosyncratic supplier risk, which a synchronized shock does not engage.

Other shocks: the claim is bounded. 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 relationship (on the full window or on the pre-COVID window that isolates the 2008–09 crisis, the 2011 Tōhoku earthquake, and the 2017–18 pressures), and the 2008–09 recession, a demand shock, if anything moved reliant industries the other way (its pre-trends are not clean). An acute, economy-wide collapse and a smooth pressure index are different objects, so these nulls bound the COVID estimate rather than overturn 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 4.2, not ongoing disruption. The sign pattern is what the framework predicts (Proposition 1): negative under a broad synchronized collapse, muted under diffuse pressure, non-negative under a demand shock. A dated check of the sharper sign-flip prediction—Tōhoku 2011, where the estimate turns positive but imprecise—is in Appendix B.

6 Incidence and Policy

The exercise of this section bounds a relative gap, not a national loss. Section 4.6 sized the margin against the other exposures that competed to explain the 2020 collapse; here we size it in output and ask where it falls. Reading the size off the reliance-quartile gaps and quartile shipment shares—the object the heterogeneity-robust estimand supports, rather than the linear per-standard-deviation slope—if every industry had entered 2020 at the bottom-quartile mean of reliance, the post-2020 differential output gap comes to about 2.6% of manufacturing shipments per year; the linear counterfactual is smaller (1.2%, roughly \$68 billion on the panel's \$5.7 trillion 2019 shipments base) and we treat it as the conservative end. Both translations apply the cross-industry gradient uniformly across the size distribution; repricing the linear counterfactual with size-tercile coefficients moves it only from 1.2 to 1.3% of shipments, so the uniformity is not doing hidden work. What remains is the measurement decomposition's finding (Appendix B) that the persistent *real* gap concentrates within four-digit families and is muted in value-weighted aggregates, a reason to read these figures as upper ends for the aggregate.

Both are relative gaps. Each nets out anything common to all industries and by construction cannot see demand that reallocated across product lines, so the economy-wide loss is smaller than either figure, and may be much smaller. Relative prices did not move (Section 4.5) and demand for this standardized output is elastic, so a buyer who lost a reliant supplier could substitute toward rivals and other product lines; to that extent the gap measures which industries, establishments, and places absorbed the shock rather than foregone national output. The one offset the data do rule out is the most direct: the gap was not filled from abroad (Section 4.5) and apparent consumption of these products fell roughly one-for-one, so within these product markets the shortfall is real even as its economy-wide footprint is bounded above by the relative gap. 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 mapping that follows is descriptive: the cell contrast behind it is not statistically distinguishable (the triple interaction has $p = 0.64$). The output cost falls more heavily on low-R&D industries (Section 4.3), and the export-control axis does not carry the heterogeneity (Table 4). The largest such group, the low-R&D, low-export-control set, holds 31% of manufacturing shipments and 32% of employment and clusters in Southern and Plains states: Alabama, Arkansas, Mississippi, and

Georgia hold 40–44% of their manufacturing employment in it, against 24–31% in Massachusetts, New Hampshire, and California (Figure 4). The footprint overlaps the regions most exposed to the 2001–2011 China trade shock (Autor et al., 2013) and is largely disjoint from the recipients of CHIPS Act awards and Section 232 protection: about 4% of the set's employment lies in a covered industry, and none in the semiconductor industries CHIPS targets. Across states the set's 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. Massachusetts, the least contract-reliant state, is the most exposed to CHIPS; the one industry in both maps is steel.[9]

Two policy implications follow, at the altitude the evidence supports. 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 rather than a counter-example. 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 not only poorly aimed at the contract-production margin but 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; which instrument fits—inventory, dual sourcing, capacity reservation, each with its own cost—is a separate question that firm-level data would inform.

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 $q \to 1$ case of the framework, 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. In the framework those are low-q, targeted shocks, for which the sign of the contract-reliance effect should reverse (Proposition 1); the one out-of-sample test the data permit—the 2011 Tōhoku episode—is consistent with that reversal but is underpowered and cannot establish it (Appendix B). We therefore do not claim contract reliance is protective against targeted

[9] Cell shares are the NAICS-6 assignment of Section 4.3 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.

shocks, only that the vulnerability identified here is specific to broad, synchronized disruptions—a pandemic or a systemic financial shock—and need not govern the targeted geopolitical contingencies that dominate the reshoring debate. 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.

7 Conclusion

We began by asking whether the import lever that dominates resilience policy reaches the margin through which disruption cuts domestic output. It largely does not. The margin that carries the domestic vulnerability, reliance on outside contractors for production steps, is predominantly domestic, nearly orthogonal to import penetration, and shows no sign of having been displaced by foreign sourcing. In normal times it is an ordinary, minor input. When COVID-19 hit, industries one standard deviation more contract-reliant lost 2.4% more output, a contraction that fell on physical volume, persisted through 2023, was followed by a durable net decline in establishment counts driven by depressed entry, and appears with the same sign in five European countries and Japan; a foreign-input channel of comparable size runs alongside it in the same regression, and that channel is the one the import lever can touch. The kind of shock matters, and the framework says why: contracting a step out diversifies idiosyncratic supplier risk while loading on a common supply factor, and the estimates line up with that ordering, negative under the broad 2020 collapse, absent under diffuse supply-chain pressure, non-negative in the demand-driven 2008–09 recession. Most of the disruptions that motivate reshoring are targeted rather than synchronized, so the vulnerability identified here belongs to broad, synchronized events and need not govern the contingencies the reshoring debate is mostly about.

Three boundaries frame these results. The output cost is identified from one acute, synchronized disruption, though no longer from one economy. The contract-work item measures a specific organizational margin, tolling and commission work on owned materials, and not the whole make-or-buy boundary. And the design is at the industry level: it identifies the output response of the contract-reliant margin, which cannot be separated here from the fragility of the standardized production that tends to be routed to contractors. 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. The mechanism is ultimately a firm-level adjustment story, and the next step is establishment-level Census microdata, linking the contract-work item to firm boundaries, supplier structure, and the disruption response observed directly.

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 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, open-data area slices), 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 (statInfId 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 2 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.

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, 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.

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 and Elhanan Helpman. Contractual frictions and global sourcing. In Elhanan Helpman, Dalia Marin, and Thierry Verdier, editors, *The Organization of Firms in a Global Economy: Economics, Politics, and Institutions*, pages 9–54. Harvard University Press, 2008.

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.

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.

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.

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.

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.

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.

Yifei Li, Anni Wang, Qun Wu, and Zejiang Zhou. Vertical integration, supply chain disruptions, and corporate yield spreads. *The British Accounting Review*, page 101639, 2025. doi: 10.1016/j.bar.2025.101639.

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.

Natália 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.

Dazhong Wu, Michael Tannen, Julius Anyu, Sergey Ivanov, and Feng Xu. Contract manufacturing, market competition, and labor productivity in US manufacturing industries. *Operations Management Research*, 16(1):377–390, 2023. doi: 10.1007/s12063-022-00317-0.

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
Total cross-industry	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, small and insignificant except under fixed effects, is included in every specification and reported in the replication package. Returns to scale is the sum of the input elasticities. Control-function GMM follows Wooldridge (2009)/Ackerberg et al. (2015); standard errors are industry-block bootstrapped. 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	$\text{CPR}^{\text{pre}}\times$ post		$\text{Foreign}^{\text{pre}}\times$ post	
	$\widehat{\beta}$ (SE)	p	$\widehat{\gamma}$ (SE)	p
Panel A. Real quantities				
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
Panel B. Margins and factor shares				
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

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 4.4); the conservative NAICS-3 wild cluster bootstrap bounds (Webb weights, null imposed, $B = 9999$) are 0.022 for output, 0.029 for employment, and 0.007 for the above-median contrast, with the full clustering ladder in Table 6. Panel A shows the cost falls on real volume; Panel B shows it does not run through profitability or factor shares for either channel.

Table 4. The output contraction concentrates in low-R&D industries

Cell	log output $\text{CPR}^{\text{pre}}\times$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	−0.006	−0.017

Notes: Equation (2) estimated within each cell of the R&D×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 ($p = 0.64$ 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 (Section 4.4), 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.

Table 5. Robustness of the output contraction

Specification	$\widehat{\beta}$ (CPR$^{\text{pre}}$×post)	p
Baseline (nominal shipments, full sample)	−0.024	0.022^b
Panel A. Composition and confound controls		
+ full composition controls	−0.028	0.032^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
Panel B. Measurement		
Foreign channel = total import penetration	−0.024	0.035^b
Real output (PPI-deflated, common sample) + controls	−0.034	0.022^b
Panel C. Fixed-effects structure and trends		
NAICS-3 × year fixed effects	−0.018	0.055^r
NAICS-4 × year fixed effects	−0.013	0.22^c
Industry pre-trends removed (fit 2015–19)	−0.025	0.051^b
Panel D. Inference variants		
Collapsed pre/post cross-section (two periods)	−0.024	0.001^h
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$^{\text{pre}}$×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 (Table 6). 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). The construction of each added control, and its own coefficient, are given in Section 4.4 and Appendix B. 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 coincide, so the gap to the −0.024 baseline is the narrower sample and the added controls, not deflation (Appendix B). 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. Dropping each of the 358 industries one at a time leaves every estimate negative and within −0.028 to −0.022, with only three moving the coefficient by more than a tenth; 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 (influence diagnostics in the replication package).

Table 6. The clustering ladder: inference matrix for the output contraction

Procedure	Baseline NAICS-6 + year FE	Sector × year + NAICS-3×year FE
$\widehat{\beta}$ ($\text{CPR}^{\text{pre}}\times$post)	−0.024	−0.018
Panel A. Industry-level variation as the effective sample		
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
Panel B. Sector as the sampling unit		
Sector block bootstrap (heterogeneity-robust estimand)	0.068	—
Ibragimov–Müller ($G = 21$ sectors)	0.23	0.23

Notes: p-values on the $\text{CPR}^{\text{pre}}\times$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 × 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 × 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 4.4. Ibragimov–Müller t-tests the 21 sector-level coefficients and does not reject; it is the ladder's most conservative member. 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.

Table 7. The acute COVID output response of contract-production reliance across economies

Economy	Treatment source	Outcome	Cells	Acute 2018–20		Full
				$\widehat{\beta}$ (SE)	p	$\widehat{\beta}$
United States	ASM contract work, 2015–19	shipments	358	−0.022 (0.007)	0.007	−0.024
EU-5 pooled	SBS subcontracting, 2017	prod. index	713	−0.027 (0.009)	0.002	−0.016
Germany	(same)	prod. index	172	−0.015 (0.016)	0.36	−0.010
Greece	(same)	prod. index	91	−0.035 (0.011)	0.006	−0.046
Spain	(same)	prod. index	152	−0.019 (0.016)	0.23	0.010
France	(same)	prod. index	150	−0.005 (0.010)	0.58	0.006
Italy	(same)	prod. index	148	−0.062 (0.025)	0.020	−0.039
Japan	Census of Mfg. consigned prod., 2019	prod. index	75	−0.028 (0.011)	0.020	−0.006

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 last column is the full-window estimate (2018–2024 for Europe and Japan; 2018–2023 for the U.S.). U.S.: NAICS-6, log nominal shipments, two-channel specification of equation (2), CRV1 on 21 NAICS-3 clusters for comparability with the sector-level clustering of the foreign rows; under the primary NAICS-6 clustering the acute coefficient has $p = 0.001$, and the NAICS-3 wild cluster bootstrap gives 0.038 (dropping the import channel gives −0.023). EU: treatment is Eurostat SBS payments to subcontractors over value added at factor cost (NACE four-digit, 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.

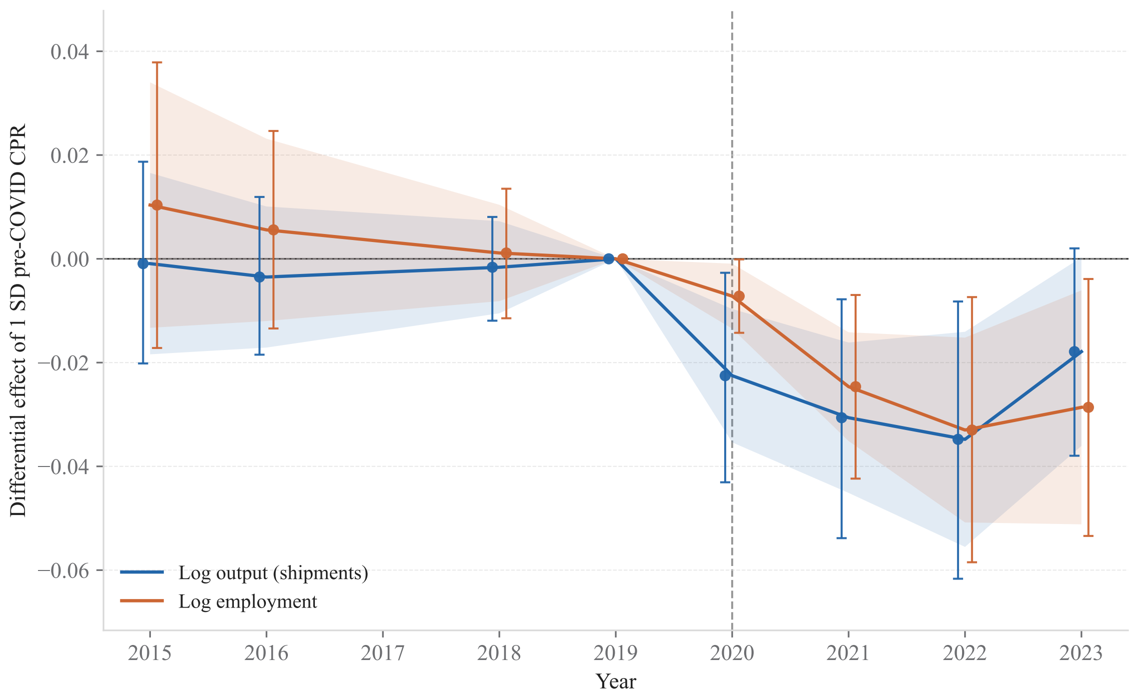

Figure 1. Event study: persistent contraction in contract-reliant industries

Notes: Coefficients on pre-COVID contract reliance (one SD, standardized) interacted with year dummies, base year 2019, window 2015–2023 (2017 Census gap), 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. 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 Table 6). Pre-2020 coefficients are small and, for output, insignificant; the employment series shows a mild pre-2020 drift discussed in Section 4.4.

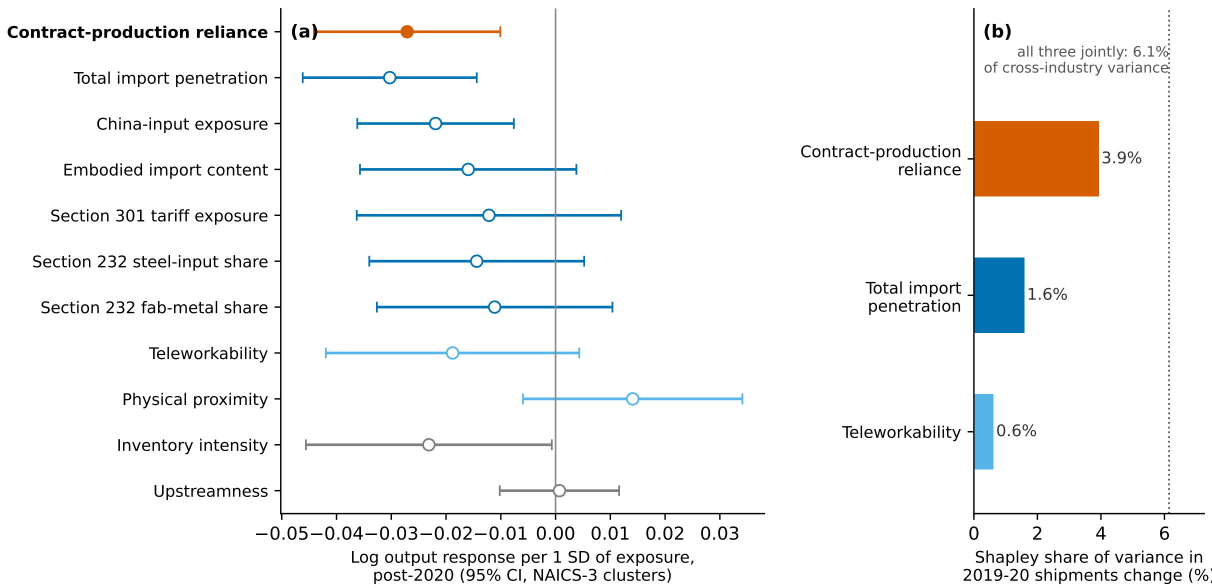

Figure 2. Contract reliance against the measured exposures, on a common footing

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. 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.

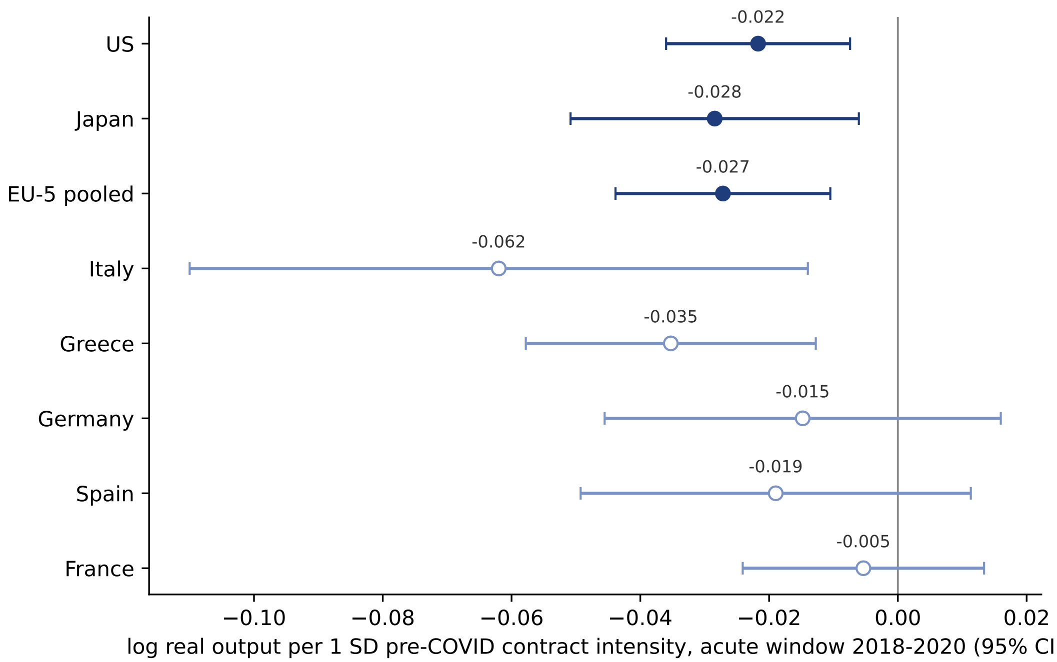

Figure 3. The acute-2020 output response of contract-production reliance, by economy

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 7. 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.

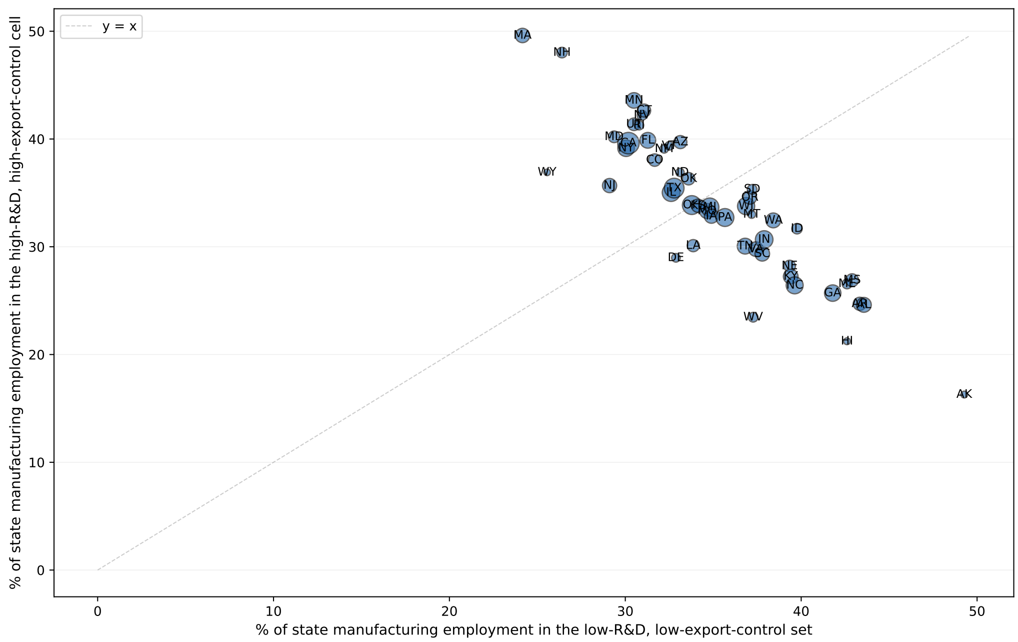

Figure 4. State manufacturing employment in the low-R&D, low-export-control set and the high-R&D, high-export-control cell

Notes: Each bubble is a state; bubble size is total manufacturing employment (2018–2022 average). The horizontal axis is the share of state manufacturing employment in the low-R&D, low-export-control set (the largest low-R&D group; Section 4.3); the vertical axis is the share in the high-R&D, high-export-control cell. Southern and Plains states (AL, AR, MS, GA) sit at the lower right; high-R&D coastal and northern states (MA, NH) at the upper left.

A A Simple Framework: Model and Proofs

This appendix states and proves the framework summarized in Section 5.2 and develops an amplification extension. The model augments the relocation-option mechanism of Bergin et al. (2009)—in which outsourcing is valuable because assembly can be shifted across alternative locations in response to shocks—in two ways their setting does not contain: a common (synchronized) factor that disables all suppliers at once, and the control–flexibility asymmetry between in-house and contract production (Assumption 1). These turn the relocation option, merely *inert* under a common shock, into a source of *exposure*.

A.1 Setup and notation

Assumption 1 (Control–flexibility asymmetry). *(i)* 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, while a facility-specific (idiosyncratic) failure cannot be re-sourced. *(ii)* Contract production is spread over $N \geq 2$ substitutable contractors: an idiosyncratic outage at one contractor is re-sourced to others, while a common disruption disables all contractors at once, costing a fraction $\rho > \rho_h$ that can be neither re-sourced nor backfilled in house. The content of the assumption is that contractors are *more* exposed to the common factor than in-house production ($\rho > \rho_h$), not that in-house is immune to it.

A producer completes a unit measure of tasks, contracting a share $c \in [0, 1]$ across $N \geq 2$ substitutable contractors and performing $1 - c$ in house; c is pre-determined when a disruption strikes. A common factor strikes with probability q, costing a fraction ρ of contracted output and a smaller fraction $\rho_h < \rho$ of in-house output; independently, each contractor suffers an idiosyncratic outage with probability $\pi < 1$, and the in-house facility loses an expected fraction θ to its own idiosyncratic failure (Assumption 1). The parameter q is the synchronization of supplier shocks.

Lemma 1 (Buffering and the common floor). *Expected contracted loss per unit is* $(1 - q)\pi^N + q\rho$*. As the market thickens (*$N \to \infty$*) the idiosyncratic term* $(1 - q)\pi^N$ *vanishes and the loss converges to the common-factor floor* $q\rho$*: substitutable contractors diversify idiosyncratic outages, but not the synchronized component.*

Proof of Lemma 1. Condition on the common factor. If it fires—probability q—it costs a fraction ρ of contracted output. If it does not—probability $1 - q$—contractors fail only idiosyncratically and independently, so a contracted task and all $N - 1$ alternates are down with probability π^N, a full loss on that task; the spare capacity of the contractors that remain up absorbs the rest. Expected contracted loss per unit is therefore $q\rho + (1 - q)\pi^N$. Since $\pi < 1$, $\pi^N \to 0$ as $N \to \infty$, leaving the common-factor floor $q\rho$. □

Lemma 1 is the resilience counterpart of the diversification of idiosyncratic shocks in production networks (Acemoglu et al., 2012): a thick, substitutable supplier market washes out independent outages but leaves the correlated component.

Combining the buffered contracted bundle with the in-house bundle gives expected output

$$\mathbb{E}[Y] = (1 - c)(1 - \theta - q\rho_h) + c\big[\,1 - (1 - q)\pi^N - q\rho\,\big], \tag{3}$$

so the marginal value of reliance is

$$\frac{d\,\mathbb{E}[Y]}{dc} = \theta - (1-q)\pi^N - q(\rho - \rho_h) \xrightarrow[N\to\infty]{} \theta - q(\rho - \rho_h). \tag{4}$$

Moving a task to a contractor escapes in-house idiosyncratic risk θ but takes on the contractor's *excess* common-factor exposure $q(\rho - \rho_h)$, net of the diversifiable idiosyncratic term. Whether reliance helps or hurts therefore turns on the synchronization q.

Proposition 1 (Synchronization threshold)**.** *$d\,\mathbb{E}[Y]/dc$ is strictly decreasing in the synchronization q whenever $\rho - \rho_h > \pi^N$ (always satisfied in a thick market). In a thick market it equals $\theta - q(\rho - \rho_h)$, so the contract-reliance effect is positive for $q < q^*$ and negative for $q > q^*$, with threshold $q^* = \theta/(\rho - \rho_h)$, which lies in $(0, 1)$ when the in-house idiosyncratic loss is below the contractor's excess common-factor loss ($\theta < \rho - \rho_h$). Reliance raises resilience to weakly correlated disruptions and lowers it to strongly correlated ones.*

Proof of Proposition 1. From (4), $d\mathbb{E}[Y]/dc = \theta - (1-q)\pi^N - q(\rho - \rho_h)$. Its derivative in q is $\pi^N - (\rho - \rho_h)$, negative whenever $\rho - \rho_h > \pi^N$, so the marginal value of reliance is strictly decreasing in synchronization. Setting the expression to zero gives $q^*(N) = (\theta - \pi^N)/(\rho - \rho_h - \pi^N)$; as $N \to \infty$ the buffered term vanishes and the condition becomes $\theta - q(\rho - \rho_h) \gtrless 0$, i.e. $q^* = \theta/(\rho - \rho_h)$, which lies in $(0, 1)$ when $\theta < \rho - \rho_h$. The effect is positive below the threshold and negative above it. □

Corollary 1 (Mapping to the shocks)**.** *A broad, synchronized supply collapse (q near one) predicts a negative contract-reliance coefficient—the COVID case. Diffuse pressure that realizes little synchronized loss (q small) predicts roughly zero—the GSCPI case. A demand-driven or idiosyncratic shock ($q < q^*$) predicts a non-negative coefficient—the 2008–09 case.*[10]

Figure A6 shows the two results together for illustrative, uncalibrated parameters. The marginal value of reliance $d\,\mathbb{E}[Y]/dc$ is a line decreasing in the synchronization q: positive at low q, where routing a step to substitutable contractors diversifies idiosyncratic risk, and negative past the threshold q^*, where the common factor dominates. A thicker supplier market (larger N) raises the line at low q by buffering idiosyncratic outages, but the two lines converge as $q \to 1$, where the buffering term vanishes (Lemma 1): thickness does not insure against a synchronized shock. The mapping to the estimated shocks—diffuse pressure and Tōhoku at low q, COVID at q near one—is the content of the figure; the parameter values are illustrative, and no result in the paper is calibrated to them.

Relation to the literature. The re-sourcing option of Lemma 1 is the relocation option of Bergin et al. (2009), here rendered inert by the common factor; its diversification of idiosyncratic risk is that of Acemoglu et al. (2012). The fragility at $q \to 1$ in Proposition 1 is the efficient-yet-fragile network of Elliott et al. (2022). The framework's distinct content is Assumption 1, the control–flexibility asymmetry, which a symmetric-location model does not contain and which the empirical results identify only in reduced form.

[10] The framework is a model of supply disruption. A pure demand contraction carries no common *supply* factor, so it maps to $q = 0$, where (4) reduces to $\theta - (1-q)\pi^N \geq 0$ in a thick market. The 2008–09 result is thus a placebo for the proposition that not every downturn loads on reliance, not a point on the synchronization ordering.

B Extended Robustness and Inference

This appendix details the checks summarized in Table 5 and Section 4.4. The production-function elasticities of Section 5.1 (Table 2) and the cell-by-cell COVID estimates of Section 4.3 (Table 4) are reported with the other tables, which follow the references and precede this appendix.

Dose-response. 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 A4): 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.

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 is sample composition, not deflation. The industries 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 (293 industries, 2018–2023), the real series reproduces the nominal path: −3.1%, −3.2%, −2.2%, and −2.5% per SD for 2020–2023, each $p \leq 0.015$, so deflation preserves the year-by-year persistence as well as the pooled coefficient. A 2019-shipment-weighted variant shows no significant real gap in any year, but the weights concentrate heavily (Kish effective $N \approx 61$ of 358; petroleum refining alone carries a tenth of the weight), so we read it as underpowered; the measurement decomposition below locates where the real persistence sits.

The persistent gap across measurement systems. The Federal Reserve's G.17 industrial production index offers an independent monthly output measure 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 (2015–2023, the twelve months of 2019 pooled as base), the reliance gradient is absent in the lockdown months and opens as the aggregate reopens, widest in July–August 2020 (−3.1% and −2.6% per SD; June–December pooled −2.1% to −2.3%, p between 0.08 and 0.12), months by which the monthly import-shock literature's China-input effect had already dissipated (Meier and Pinto, 2024; Lafrogne-Joussier et al., 2023). The within-year timing separates the organizational margin from lockdown incidence and from the import channel, and its 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, and the difference decomposes cleanly. At the identical 63-group aggregation, nominal shipments retain the persistent path (−2.0%, −2.9%, −2.2% for 2021–2023) while PPI-deflated shipments match the production index, both within a point of zero after 2020; at six-digit resolution the real persistence is fully present (preceding paragraph). The persistent real gap therefore lives in the unweighted six-digit cross-section, within four-digit families; between family aggregates, what persists is nominal, carried by relative prices. A value-weighted, between-family real index is the one cut of the data in which no source shows persistence, ours included, and the production index is exactly that cut. The labor margin needs no such qualification: administrative QCEW employment (Section 4.2) 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 (-3.1% by 2023) 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 8)—the formal statement of the commodity tilt that the leave-one-industry-out exercise flags. The concentration is in leverage, not in outcome. 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.

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 A5 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 do move. Fully dollar-weighted estimates lose precision because the weights themselves concentrate (Kish effective $N \approx 61$; capping any industry at one percent of total weight restores both precision and the coefficient), and the dollar-weighted 2020 cell of nominal shipments is contaminated by the petroleum price collapse, which makes the least reliant giant industries crash in dollars (the dollar-weighted acute coefficient is an uninformative $+0.023$, standard error 0.037; capped, -0.012). 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, $p = 0.031$), and repricing the Section 6 counterfactual with tercile-specific coefficients gives 1.3% of shipments a year against the linear 1.2%. 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.

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 4.4, 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

Table 8. The effective sample of the baseline regression

Panel A. Ten largest identifying weights (industries)					
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

Panel B. Identifying weight by sector (six largest)				
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.

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 3.5% 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, −0.022, −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 4.4 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 (86 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 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 flat pre-trends on both margins (the largest pre-period entry coefficient, -0.039 in 2017, carries $p = 0.19$). 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.

Which margin carries the output loss. 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 is what makes the acute year legible. 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. 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 most of the response (83% undetrended, 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 6 documents throughout. Extensive shares are ratios of two estimated coefficients and carry no standard error; inference rests on the intensive coefficient.

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 (the pre-period carries one small negative coefficient, -0.006 in 2018, which 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 throughout.

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 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 ($r = -0.09$). 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). What the data cannot resolve is in-window inventory dynamics—whether contract-reliant firms drew down or failed to rebuild stocks during 2020–2023—because the NBER-CES series ends in 2018 and the post-2018 Census instruments do not report inventory; that, with establishment size, is left to firm-level data. Under the same composition controls the employment effect attenuates by about a quarter, so part of it is composition.

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 ($r = -0.16$ with the 2022 index), which is why neither control moves the estimate (Section 4.4). The industry-specific detrending fits each industry's linear trend on 2015–2019 and extrapolates it forward; fitting instead on 2012–2019 gives the same 2020 break with a smaller full-window estimate (-0.019).

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 (Proposition 1). 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. The estimate stays insignificant at every rung of the clustering ladder ($p = 0.20$–0.26), and the design is underpowered for the sharp question: the minimum detectable effect under the primary NAICS-6 clustering is roughly 0.038, half again the COVID magnitude. What the interval does support is a difference between the shocks: the 95% confidence interval, $[-0.011, 0.042]$, excludes the -0.024 COVID point estimated on the same industries, so the Tōhoku response, while not established as positive, is unlikely to be the COVID response reappearing.[11]

[11] The contract-reliance estimate is positive but imprecise—one narrow event on 188 industries—so we read it as the mechanism's predicted sign, not a second magnitude. 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.

Weak in-house fallback. The framework makes a heterogeneity prediction beyond the sign: the loss should be larger where a contracted step is hardest to bring back in house, since weak fallback raises the contracted loss ρ under a common shock (Assumption 1(ii)). We test it with three pre-2020 proxies for weak fallback, each nearly orthogonal to the R&D axis on which the loss concentrates ($|r| \leq 0.16$ with the non-production-worker share), so they do not merely relabel the low-R&D margin: capacity utilization (Federal Reserve G.17), capital intensity (NBER-CES real capital per worker), and average establishment size (County Business Patterns). Interacted with $\text{CPR}^{\text{pre}} \times \text{post}$, the triple coefficient carries the predicted negative sign for capacity utilization and for the composite index ($\widehat{\phi} \approx -0.004$) but is far from significant (wild cluster $p > 0.5$); splitting industries at the median, the per-standard-deviation loss roughly doubles in the weak-fallback half on the capital and scale proxies (-0.039 against -0.019 on establishment size) and is flat on capacity utilization. With 21 clusters this triple interaction is underpowered, as the cell contrast of Section 4.3 was ($p = 0.64$), and the directional pattern does not survive holding the R&D moderator fixed. The prediction is consistent with the data but not identified at the industry level, the firm-level question the paper leaves open; the build and estimates are in the replication package.

C Cross-Country Data and Comparability

This appendix documents the data behind Section 4.7: 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 (variable V23110) come from the Structural Business Statistics, compiled by national statistical offices under a common EU regulation, largely from a mix of business surveys and administrative or tax records, at the enterprise (not establishment) level, and disseminated at NACE four-digit. The variable was collected triennially; the 2017 wave is the pre-COVID cross-section, divided by value added at factor cost (V12150) from the same framework. 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 V23110 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. 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 correlate at Spearman 0.77 (Pearson 0.87), so the harmonized variable preserves the ranking the design uses.

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. The index's industry categories are goods-based rather than JSIC-based, so we construct a concordance of 75 cells: where 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; each cell carries a flag for whether the mapping is exact or approximate. A conservative 22-cell panel at the sector level, where the mapping is one-to-one by construction, is estimated alongside.

Mexico. Mexico is the one economy observed from both sides of the market, and its evidence is reported here rather than in the pooled exhibits. The annual manufacturing survey (EAIM, 2018 series) reports, for 206 six-digit SCIAN industries, 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, averaging 1.2% of gross output—and a seller-side item, income from performing such work, split by client location (foreign principals account for 85% of it, the maquiladora institution); the outcome is the EMIM physical-volume production index for the same industries, so the merge needs no concordance. No payee-location split of the expense item exists in the Mexican system, 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 collapses 89% over 2019–2023 with Pemex's crude-toll-processing wind-down while industries' contract spending stays nearly flat (−7%); and domestic contractors' income from domestic principals is roughly twice manufacturers' total contract spending. The estimates are held out of the pooled comparison: the payer-side treatment carries a pre-existing relative decline, so its acute −3.5% per SD reads as directional only, and the seller-side 2020 loss attaches to domestic-principal contractors (−3.0% detrended, randomization $p = 0.002$) but is carried by the textile–apparel–leather bloc (subsectors 313–316, 26 of 206 industries holding 45% of domestic exposure), whose lockdown status is close to collinear with the treatment—so "domestic broke and cross-border did not" and "non-essential broke and essential did not" are near observationally equivalent here. Mexico therefore verifies the payer-side reliance concept from both sides of the market without adding a ninth acute estimate; separating organizational form from industry composition would need establishment-level IMMEX-versus-non-IMMEX access through INEGI's microdata laboratory.

C.2 What the variables measure, and how their processes differ

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. And the resulting distributions sit 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 value added, reaches ten times that in Italy—with a long right tail in every country. The 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), 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 What the empirical model absorbs

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; the German Spearman of 0.77 bounds how much reranking the broad concept introduces.

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.

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 (full-window pooled $p = 0.067$). The clustering ladder of Section 4.4 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 0.002 to 0.0004 to 0.0002 and the full-window p from 0.137 to 0.102 to 0.066, in line with the within-country randomization p of 0.067; 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 ($p = 0.34$), Spain (0.34), and France (0.62) stay insignificant—so the sign pattern, and no single country, continues to carry the cross-country evidence. The pooled European point estimate is stable across fixed-effects structures (-0.015 to -0.022 under common-year effects, separable country effects, country trends, and weighting), but its full-window significance leans on Greece and Italy—dropping either roughly halves it—and the pooled event study has a small positive 2018 coefficient (0.010), both 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 -0.038 ($p = 0.065$), 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 B).

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 3 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.

D Summary statistics

Table 9. Summary statistics, NAICS-6 manufacturing panel

	Mean	SD	p10	p50	p90
Panel A. Industry-year panel, 2005–2023					
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
Panel B. Pre-COVID industry cross-section					
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: 2005–2023 (2017 gap) Industry-years: 5226					

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.

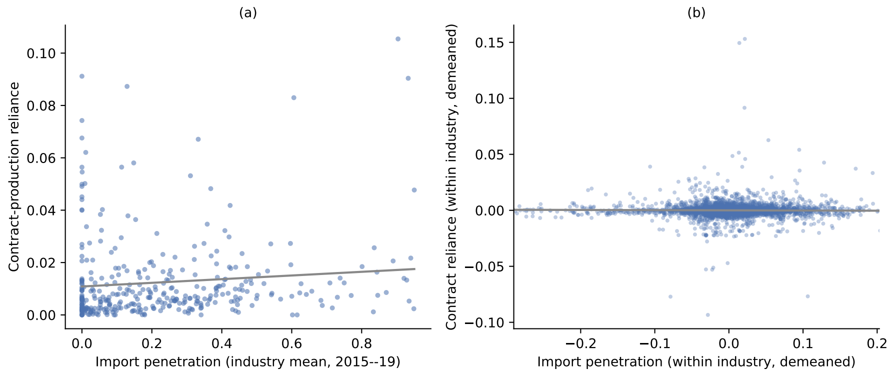

Figure A1. Contract reliance is statistically distinct from imports

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.

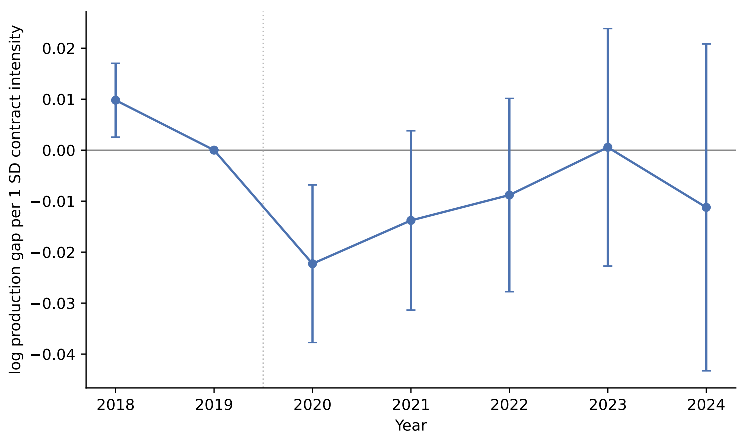

Figure A2. Pooled European event study

Notes: Coefficients on pre-COVID subcontracting intensity (one within-country SD, Eurostat SBS 2017 wave) interacted with year dummies, base year 2019; log STS production index, five countries at NACE four-digit, country-by-industry and country-by-year fixed effects, 95% intervals clustered on country-by-NACE-2. The gap opens at 2020 (-0.022, $t \approx 2.8$) and closes by 2023; the small positive 2018 coefficient (0.010) is discussed in Section 4.7 and Appendix C.

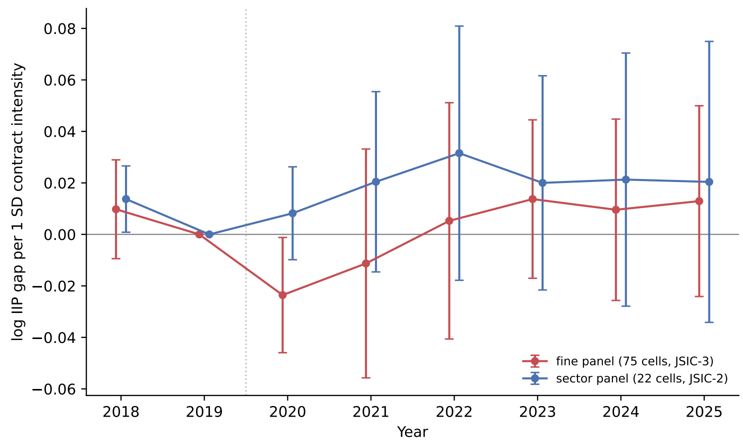

Figure A3. Japan event study

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).

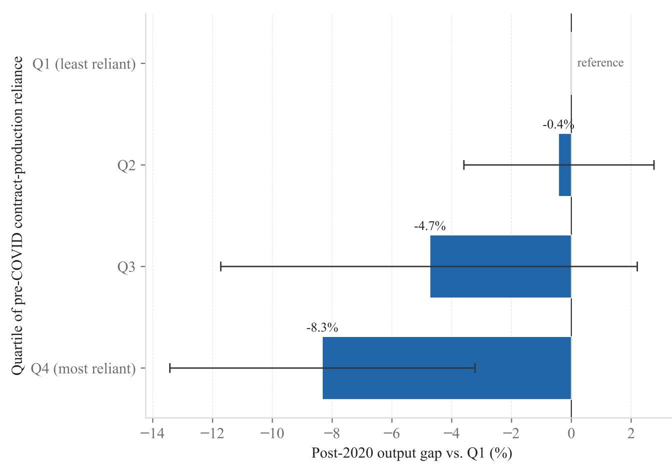

Figure A4. Dose-response: post-2020 output gap by pre-COVID reliance quartile

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.

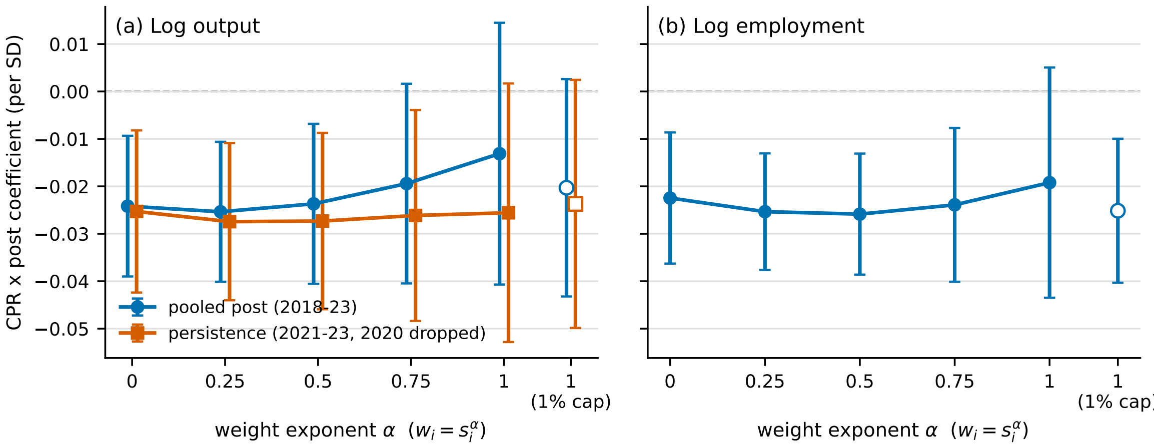

Figure A5. The weighting path: from the average industry to the average dollar

Notes: The CPR$^{\text{pre}}\times$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 B).

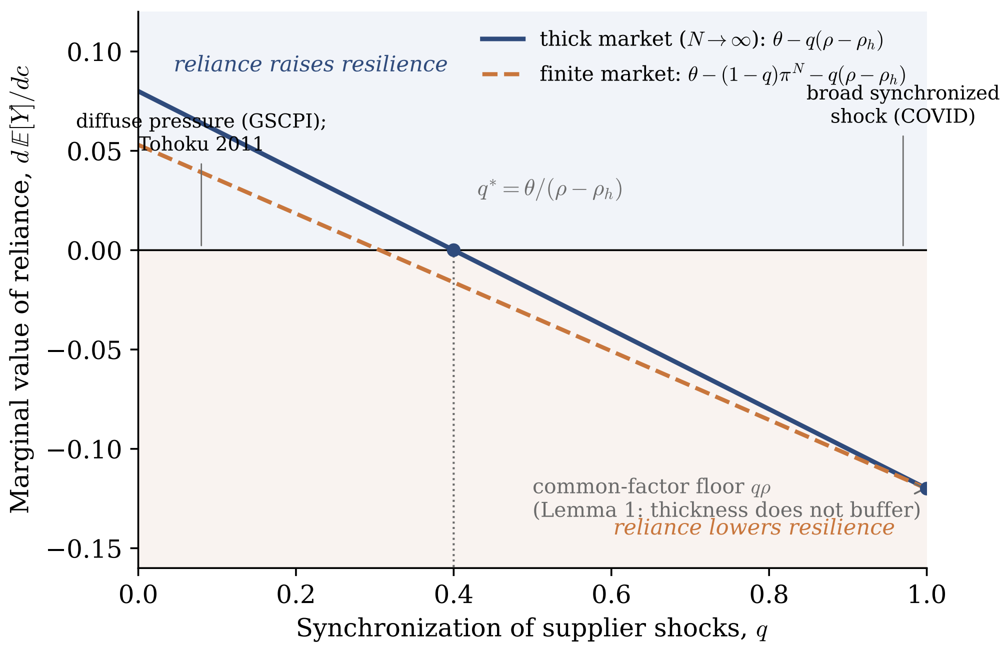

Figure A6. The marginal value of contract reliance and the synchronization threshold (illustrative)

Notes: The marginal value of reliance $d\,\mathbb{E}[Y]/dc = \theta - (1-q)\pi^N - q(\rho - \rho_h)$ as a function of the synchronization of supplier shocks q, for a thick market ($N \to \infty$, solid) and a finite market (dashed), under illustrative parameters $\theta = 0.08$, $\rho - \rho_h = 0.20$, and $\pi^N = 0.027$ (*not* calibrated). Reliance raises resilience for $q < q^*$ and lowers it for $q > q^*$, with $q^* = \theta/(\rho - \rho_h)$ in the thick market (Proposition 1); the two curves converge as $q \to 1$, where the buffering term $(1-q)\pi^N$ vanishes (Lemma 1). The figure illustrates the sign ordering the paper tests and identifies no parameter.