Export Controls and Endogenous Upstream Price Incidence: The Chip War's Consequences for the AI-Chip Industry

Hyungjin Kim[1]

March 23, 2026

Abstract

Modern semiconductor production is vertical, granular, and concentrated. This paper shows that in a bilateral oligopoly, export controls force a renegotiation of upstream input prices — a channel absent from standard quantitative trade models that treat input costs as fixed. I develop a vertical production network model in which AI-chip designers negotiate with a concentrated set of High-Bandwidth Memory (HBM) suppliers via Nash-in-Nash bargaining. Using Korean customs data, I document destination-specific price divergences consistent with relationship-specific bargaining. In the calibrated model, targeted buyers' input prices fall while rivals' input prices rise — the upstream shock-absorber and cross-buyer spillover channels. The input-price adjustment is roughly one-fifth of the downstream markup response, and a frozen-input counterfactual that ignores this renegotiation overestimates the competitive reallocation caused by export controls. Network composition matters: adding a single qualified supplier dampens the targeted buyer's markup increase by approximately 46 percent.

JEL Codes: F12, F13, F14, L13, L14
Keywords: Trade policy, bilateral oligopoly, global value chains, bargaining, export controls, semiconductor supply chains

[1]Korea Development Institute; hyungjinkim@kdi.re.kr. I am deeply indebted to Jonathan Eaton and Stephen Yeaple for invaluable guidance. I thank James Tybout and Jingting Fan for their insightful discussions. I appreciate the insights and comments of Fernando Parro, Kala Krishna, Kai-Jie Wu, and Maria-Jose Carreras-Valle.

1 Introduction

Standard economic intuition treats export controls as simple output wedges that reduce downstream revenue while assuming upstream input costs remain invariant. However, this view overlooks a critical reality: powerful exporters often depend on an equally powerful tier of input suppliers. This paper shows that export controls force a renegotiation of the terms of trade between these tiers. In the calibrated equilibrium for the AI-chip industry, I find that this upstream renegotiation acts as a shock absorber: suppliers lower input prices for the targeted firm to preserve volume, thereby dampening the policy's intended effect (this direction holds when the restricted buyer's demand is sufficiently responsive to input prices and when the buyer's bargaining weight remains below an analytically derived threshold — approximately $\gamma \approx 0.85$ at the baseline elasticities; above this threshold the mechanism inverts (Remark 5); the baseline calibration satisfies both conditions, see Remark 3). Simultaneously, suppliers extract higher prices from unconstrained rivals facing surging demand, further attenuating the competitive reallocation the policy intends.

I quantify these mechanisms in the context of the U.S.-China "Chip War." In October 2022, the United States imposed sweeping controls on the export of advanced artificial intelligence (AI) chips to China. The AI supply chain is the ideal laboratory for this analysis: a handful of chip designers (e.g., NVIDIA, AMD) negotiates prices with a concentrated set of High-Bandwidth Memory (HBM) suppliers (e.g., SK Hynix, Samsung, Micron). In this bilateral oligopoly, a trade shock does not just reduce output; it fundamentally alters the outside options — and hence the effective distribution of surplus — between buyers and suppliers.

I show that this upstream repricing depends on the composition of the production network. Adding a single qualified supplier significantly amplifies the upstream sector's shock-absorbing capacity, further dampening the export control's effect on downstream prices. This composition non-neutrality implies that the supply chain's resilience to trade policy crucially depends on the specific configuration of firm-to-firm contracts.

To capture the upstream propagation, I leverage a unique empirical setting. Ideally, one would track HBM shipments directly; however, HBM is a novel intermediate that lacks a dedicated long-standing trade code. I therefore identify the trade shock through the explosion in Korean exports of Multi-Chip Packages (MCP-IC) — the specific customs classification under which HBM-attached accelerator modules are recorded. This deep-level classification distinguishes HBM-bearing packages from commodity memory chips, overcoming the limitation of conventional HS-6 codes, which aggregate DRAM, NAND, and HBM into buckets that obscure product-level heterogeneity. The HS 8542323000classification is stable throughout the sample: administrative customs records show the code reporting exports to 20–21 destinations in every year from 2007 to 2025, with no discontinuous jump in coverage around either policy event, and the commodity-memory control codes (DRAM: HS 8542321010; NAND: HS 8542321030) show no corresponding drop that would indicate product reclassification. Using high-frequency customs data and industry reports, I document three facts. First, the AI chip value chain is highly concentrated, with Herfindahl–Hirschman Index (HHI) values exceeding 3,000 in both upstream and downstream tiers. Second, in line with U.S. export controls, prices and quantities did not adjust uniformly; instead, there was a sharp reallocation of high-value components toward Taiwan. Third, the unit values of bottleneck inputs diverged significantly across destinations, consistent with negotiated, relationship-specific prices rather than a single market-clearing price.

To document the empirical footprint of the policy shock, I implement an event-study design around the two major control regimes. The event-study profiles show a sharp contraction of 15–20 log points in Taiwan-bound MCP–IC flows following the October 2022 controls, and a dramatic sign reversal — a large sustained expansion — following the October 2023 tightening. The reversal reflects supply-chain reorganization: by late 2023, the industry had fully rerouted advanced AI-accelerator assembly from Chinese buyers to the Taiwan CoWoS packaging hub, so tighter controls on China-bound chip exports further accelerated Korean HBM shipments to Taiwan rather than dampening them. This opposite-

signed response across regimes is the key descriptive pattern that the structural model is calibrated to replicate; the model's identification comes from equilibrium moment conditions rather than from quasi-experimental variation in the event study. An important caveat is that the customs data record aggregate destination-level unit values and cannot directly identify the bilateral link-specific price divergence the model predicts: the shock-absorber mechanism requires NVIDIA's negotiated input price to fall while AMD's rises, but these opposite-signed bilateral movements are mixed together in the aggregate customs record. The empirical evidence thus documents three reduced-form facts *consistent with* the mechanism; the bilateral price responses are recovered from equilibrium moment conditions in the structural calibration (Section 4).

To interpret these facts, I develop a vertical production network model characterized by oligopoly in both the upstream and downstream sectors. Downstream firms compete in quantities (Cournot), but input prices are determined via Nash-in-Nash bargaining with upstream suppliers. Baseline bargaining weights are set symmetrically at 0.5 — a conservative choice that isolates network-position effects from heterogeneous leverage; Section 4 sweeps $\gamma \in \{0.30, 0.50, 0.70\}$ to confirm results are directionally robust and to bound the sensitivity of the one-fifth ratio. In the calibrated quantification exercise, I find that the endogenous input-price adjustment is quantitatively significant — roughly one-fifth of the downstream markup response for the dominant buyer under the baseline calibration. The relevant comparison is a *frozen-input* counterfactual that holds Nash-bargained upstream prices at their pre-shock equilibrium values while downstream quantities adjust — the implicit assumption in hat-algebra quantification exercises (Caliendo and Parro, 2015) that treat input prices as exogenous. The gap between the full Nash-in-Nash equilibrium and this frozen-input benchmark isolates the contribution of endogenous HBM repricing: neither the deflationary channel for the targeted buyer nor the inflationary spillover to its rival appears in the frozen benchmark, so models that abstract from vertical bargaining substantially overestimate the competitive reallocation caused by export controls.

Although this paper focuses on the AI-accelerator supply chain, the mechanism I identify — upstream price incidence in bilateral oligopolies — applies to a wide range of high-value industries. Modern high-value manufacturing is increasingly characterized by "layers of powerful firms" rather than atomistic competition. In commercial aerospace, a duopoly of airframers (Boeing and Airbus) negotiates with a concentrated tier of engine manufacturers. In the pharmaceutical industry, a handful of large drugmakers source active ingredients from specialized contract manufacturers. In the automotive sector, large OEMs face equally powerful Tier-1 suppliers for critical components such as batteries. In all such settings, trade policy shocks will be mediated by the renegotiation of input prices between powerful upstream and downstream firms. Global value chain theory studies how such vertical relationships are organized (Antràs and Chor, 2013; Alfaro et al., 2019); my analysis complements this work by focusing on the incidence of policy shocks within an existing network structure.

First, I contribute a small-n Nash-in-Nash framework in which a buyer-specific trade wedge endogenously reprices upstream inputs and generates quantifiable cross-buyer spillovers. The closest paper is Alviarez et al. (2025), who apply Nash-in-Nash bargaining in a general equilibrium model of firm-to-firm trade to document large bilateral price wedges; their paper characterizes the *cross-sectional* distribution of prices under two-sided market power, whereas I study how a buyer-specific policy *shock* endogenously reprices upstream inputs — the shock-absorber and cross-buyer spillover channels are absent from their framework by design. I also complement Baqaee and Farhi (2024), whose linearized production-network framework characterizes aggregate incidence of markup distortions; I operate at the small-n, link-specific level, which permits the nonlinear incidence and cross-buyer spillovers formalized in Corollary 1 that are obscured by aggregation. The broader Nash-in-Nash foundations trace to Rey and Vergé (2004) and de Fontenay and Gans (2014) on passive-beliefs bargaining in networks with externalities, Stole and Zwiebel (1996) on multilateral bargaining, and Inderst and Wey (2003) on buyer power in bilaterally oligopolistic industries. Applications to

domestic industries include Crawford and Yurukoglu (2012) (television) and Collard-Wexler et al. (2019) (healthcare), neither of which studies how trade wedges interact with bargaining weights. Amiti et al. (2014) show that large importers absorb exchange-rate shocks via markup adjustment, a mechanism analogous to the upstream shock-absorber I identify; Juárez (2025) documents empirically that buyer market power shapes exchange-rate pass-through in a bilateral oligopoly setting.

Second, I contribute to the literature on trade policy with firm heterogeneity (Eaton et al., 2022b; Gaubert and Itskhoki, 2021). Most quantitative trade models treat input prices as exogenous or determined by competitive clearing (Caliendo and Parro, 2015). I show that ignoring the endogenous renegotiation of input prices leads to a substantial mismeasurement of policy incidence. The theory of cost pass-through under imperfect competition (Weyl and Fabinger, 2013) and under bilateral Nash bargaining with vertical contracts (Gaudin, 2016) provides the analytical foundation for this channel; Adachi and Ebina (2014) extend the analysis to double-marginalization settings. On the trade policy side, Brander and Spencer (1985) provide the canonical oligopoly trade policy model; I study export restrictions — the inverse instrument — in a bilateral bargaining setting.

Finally, this paper offers a rigorous structural evaluation of the AI chip industry. While recent work has examined semiconductor subsidies and learning-by-doing (Goldberg et al., 2024) and structural dynamics in semiconductor-adjacent concentrated markets (Igami, 2017), this paper is the first to model the specific bottleneck bargaining between AI chip designers and memory suppliers. I find that the effectiveness of export controls depends critically on the *network structure*—specifically, whether rival firms share the same suppliers as the targeted firm. Concurrently, Crosignani et al. (2025) document the supply-side costs of the same October 2022 controls, finding substantial market-capitalization losses for U.S. semiconductor suppliers unable to redirect exports; my analysis focuses on the upstream price incidence mechanism for targeted buyers and the cross-buyer spillover to their rivals.

The paper is organized as follows. Section 2 documents stylized facts on multi-stage

concentration and the bargaining granularity in AI/HBM. Section 3 lays out the vertical-link model and equilibrium. Section 4 presents counterfactuals for recent export-control scenarios, decomposes mechanisms, and discusses policy implications and external validity. Section 5 concludes.

2 Stylized Facts on the AI/HBM Supply Chain: Market Power and Trade Patterns

This section documents empirical patterns that map to the key primitives of the model: bilateral oligopoly with Nash-in-Nash input bargaining and Cournot competition downstream. The facts below are used as calibration points for policy experiments.

2.1 Industry Structure: A Bilateral Oligopoly

This subsection describes HBM and AI accelerator producers and documents concentration in each segment. The AI-accelerator supply chain is characterized by extreme concentration at both the upstream and downstream tiers.

Table 1 summarizes the production ecosystem. Panel A lists the three firms—SK hynix, Samsung, and Micron—that account for virtually the entire global supply of High-Bandwidth Memory (HBM). Panel B reports the leading accelerator designers. NVIDIA is the primary buyer of HBM for data-center GPUs, with AMD and a limited set of cloud-native designers (e.g., Google, AWS) constituting the residual demand. Crucially, the production network is highly globalized yet distinct from commodity memory chains. HBM production requires tight integration with advanced packaging capacity (e.g., TSMC's Chip-on-Wafer-on-Substrate (CoWoS)), which creates high switching costs and qualification barriers between buyers and suppliers.

Table 2 translates this structure into concentration metrics. Panel A reports Q2 2025 HBM market shares from Counterpoint Research, chosen to reflect the quarter when Samsung

Table 1: HBM and AI accelerator value chain: producers, partners, and concentration (indicative)

Panel A: HBM producers and production/assembly sites

Chip maker	HQ (Country)	Major HBM fab (City, Country)
SK hynix	South Korea	Icheon, Cheongju (KR); Wuxi (CN)
Samsung Electronics	South Korea	Hwaseong, Pyeongtaek (KR); Xi'an (CN)
Micron Technology	USA	Taichung (TW, back-end); Hiroshima (JP)*

Panel B: AI accelerator designers and manufacturing/packaging partners

Designer	Foundry (node)	Advanced packaging (type)
NVIDIA (H100/H200)	TSMC (5/4/3 nm)	TSMC CoWoS (2.5D with HBM)
AMD (MI300 series)	TSMC (5/4 nm)	TSMC CoWoS
Intel (Gaudi 3)	Intel/TSMC (var.)	EMIB/CoWoS/OSAT mix
Broadcom (custom XPUs)	TSMC (advanced)	CoWoS/interposer

Panel C: Concentration snapshot (2024–2025, indicative)

Segment	Recent concentration snapshot
HBM suppliers	SK hynix, Samsung, Micron—three-firm oligopoly
Advanced packaging	TSMC CoWoS is a key capacity bottleneck
AI accelerators	NVIDIA is the dominant AI accelerator designer and HBM buyer
HBM market size	Rapid growth from $\sim$ \$18B (2024) to $\sim$ \$34B (2025e)

* Linked ecosystem/supply-chain site rather than a dedicated HBM fab.

Notes: Panel A sites reflect public reporting of representative DRAM/HBM front-end or back-end locations; specific qualified HBM3/3E lines are subsets of broader DRAM facilities. Panel B emphasizes that most leading accelerators rely on HBM and advanced packaging. Panel C entries reflect public trackers and company statements; see main text for sources. In the model, I treat the upstream side as an HBM triopoly.

had not yet cleared NVIDIA's HBM3E qualification: SK hynix (62.0%), Samsung (17.0%), and Micron (21.0%). This implies an HHI of approximately $4,574$, more than double the threshold for a "highly concentrated" market under U.S. merger guidelines. Panel B reports a lower bound for downstream concentration: industry estimates place NVIDIA's share of the AI accelerator market near 80%, implying a minimum HHI of $6,400$.

This structure motivates the modeling choice in Section 3. Unlike markets with atomistic firms, where prices are determined by aggregate clearing, the AI/HBM vertical is defined by granular market power. Each side has a significant outside option, and prices are determined via bilateral negotiation. Consequently, a trade shock that hits a specific buyer (e.g., NVIDIA) does not merely shift a demand curve; it fundamentally alters the outside options of the few players involved, shifting the effective balance of bargaining power and activating the upstream price incidence channel.[1]

Table 2: Concentration in the HBM upstream and AI accelerator segments (indicative)

Panel A: HBM suppliers — market shares and concentration (Q2 2025)	
Supplier	**Share (%)**
SK hynix	62.0
Samsung	17.0
Micron	21.0
HHI	4,574
Panel B: AI accelerators — lower-bound concentration (2024–2025)	
Leader share (s_L, NVIDIA)	≥ 80
HHI lower bound	6,400

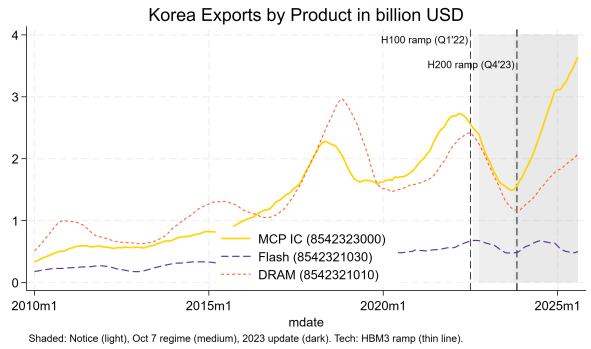
Notes: Panel A shares follow Counterpoint Research (Q2 2025), a quarter when Samsung had not yet qualified for NVIDIA's HBM3E program; $HHI = 62.0^2 + 17.0^2 + 21.0^2 = 4{,}574$. Panel B reports a lower bound for accelerator concentration using the leader share $s_L \approx 80\%$; with only a leader identified, $CR1 \geq 80$ and $HHI_{\min} = s_L^2 = 6{,}400$. Panel A figures are used as calibration targets for upstream market shares in the static model.

[1] Jon Peddie Research (Q3 2024) and Mercury Research discrete-GPU shipment estimates both place NVIDIA's data-center AI accelerator share at approximately 80% for 2024 training workloads. The lower bound $HHI_{\min} = s_L^2 = 6{,}400$ treats the remaining 20% as distributed across infinitely many firms (the minimum-concentration scenario); a two-firm residual (NVIDIA 80%, AMD 20%) yields $HHI = 80^2 + 20^2 = 6{,}800$.

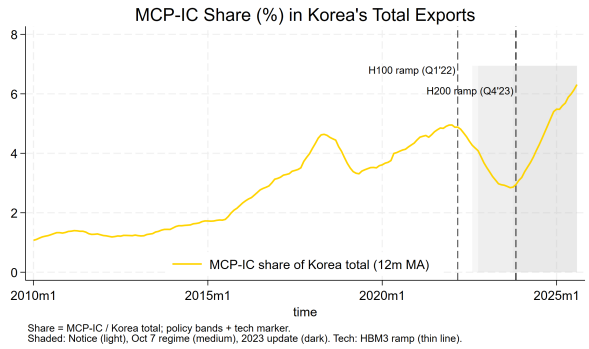
The inference that these structural conditions support bilateral bargaining rather than posted-price schedules is consistent with industry practice. NVIDIA's 10-K filings disclose long-term supply commitments with HBM vendors tied to specific product generations, and industry analyst reports document that HBM pricing is negotiated on a per-customer, per-generation basis rather than set at a single market-clearing price (see, e.g., Omdia, 2024). These multi-quarter supply agreements, together with the absence of a spot market for HBM, distinguish MCP-IC from commodity DRAM — where short-term contracts and benchmark pricing are the norm — and provide the institutional foundation for modelling upstream prices as bilaterally negotiated outcomes. The assumed disagreement payoffs follow the standard Nash-in-Nash convention: each side earns its outside option — the payoff from trading with remaining partners at the hypothetically removed link's pre-shock prices (Remark 2).

2.2 Trade Patterns in HBM-Related MCP–IC Exports

Remark 1 (Aggregate unit values vs. bilateral link prices: three caveats)**.** *The customs data record destination-level unit values (USD per kg shipped from Korea to a given country), not the bilateral negotiated prices p^U_{du} of the model. Three caveats govern interpretation. (i) Buyer-flow mixing: the model predicts that NVIDIA's link-specific input price* falls *while AMD's* rises *(Corollary 1); customs data aggregate across all buyers, so when NVIDIA is constrained and AMD demand surges, the AMD-weighted upward pressure dominates the aggregate unit value — these predictions need not contradict the observed rising trend. (ii) Generational upgrade: the HBM upgrade from HBM2E to HBM3E roughly doubles per-die memory bandwidth and the value-per-kilogram ratio, independently driving the surge to ≈200,000 USD/kg by 2025. (iii) AI investment boom: expanding total Taiwan-bound volume shifts product mix toward highest-generation chips. The model is calibrated to the cross-sectional policy component — the divergence between Taiwan-bound and China-bound unit values within the post-October 2022 window — not to the level trend. The bilateral link prices themselves are unobserved and are recovered from equilibrium moment conditions in*



(a) Korea exports by product (WR totals, 12m MA). Shaded = policy; thin line = HBM3/H100 ramp.



(b) MCP–IC share of Korea's total exports (WR HS=0, 12m MA).

Figure 1: Scale and macro footprint of the MCP–IC boom.

Section 4.

To quantify the upstream response to downstream shocks, I track Korea's exports of Multi-Chip Packages (MCP–IC, HS 8542323000). This category records HBM-attached accelerator modules, distinct from commodity memory. I contrast these flows against DRAM (HS 8542321010) and NAND Flash (HS 8542321030), which serve as controls: they share upstream wafer inputs and macro-demand drivers but do not face the specific integration requirements of the AI-accelerator vertical.[2]

Using monthly customs records, I first document the decoupling of AI memory from the broader cycle. Figure 1 shows that while standard memory exports (DRAM and Flash) follow cyclical macro-demand patterns, MCP–IC exports exhibit a structural break beginning around the HBM3/H100 technology ramp (Panel A). By 2025, this single tariff line approaches 6% of Korea's total exports (Panel B), confirming that HBM has transitioned from a niche intermediate good to a macroeconomically significant export. This divergence validates the use of MCP–IC as a proxy for AI-specific hardware: its volume dynamics are

[2]I use monthly Korean customs export data by HS10 code and partner country (Korea Trade Statistics Promotion Institute), from 2010m1 to the latest available month. See Appendix A for variable construction and index normalization details. Exynos and other application processors are classified under HS 8542.31 (processors and controllers), not HS 854232 (memories), and therefore do not enter HS 8542323000. The residual contamination risk is from mobile memory packages (eMCP/uMCP), which ship primarily to China and Vietnam for handset assembly — not to Taiwan — and trade at unit values several orders of magnitude below the observed Taiwan-bound premium.

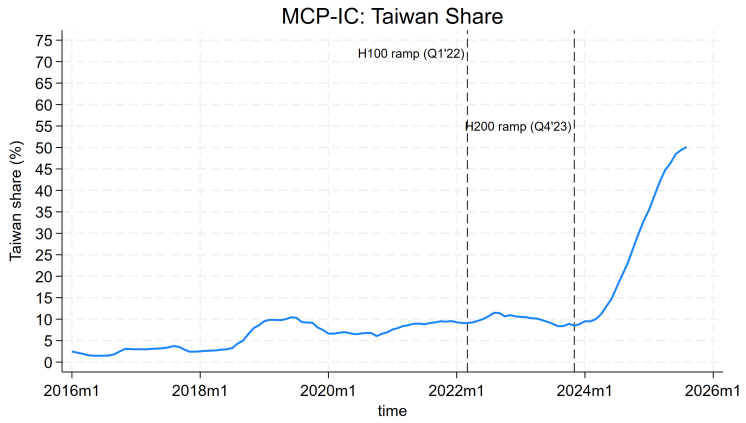
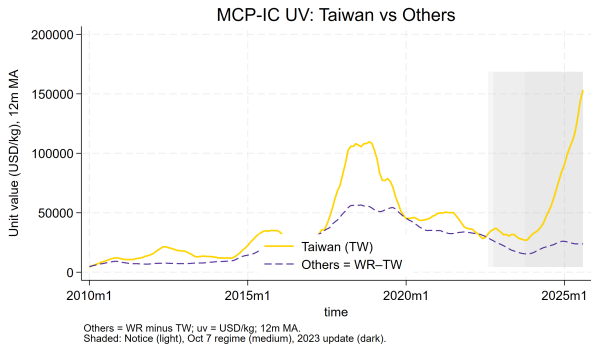


Figure 2: Reallocation toward Taiwan around the accelerator ramp. MCP–IC: Taiwan's share of world exports (12m MA). Tech markers: H100/H200.

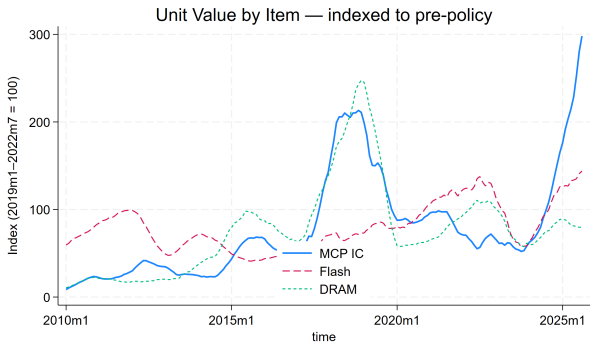
orthogonal to the broader memory cycle.

Fact 2. The destination structure of these exports reveals the rigidity of the production network. Figure 2 plots Taiwan's share of Korean MCP–IC exports. The share remains stable through 2021 but bends sharply upward following the launch of HBM3-enabled accelerators, reaching 52% by 2025. This upward trend reflects two distinct forces. The pre-2022 structural shift toward Taiwan reflects TSMC's emergence as the primary CoWoS advanced-packaging hub for AI accelerators: as NVIDIA's CoWoS demand scaled from 2020, so did Korean HBM shipments to Taiwan as an upstream intermediate. The post-October 2022 *acceleration* is the enforcement-driven component: tighter controls redirected AMD's and other firms' sourcing away from Chinese fabs toward TSMC CoWoS, magnifying Taiwan's share beyond its structural trajectory. The event-study in Section 2.3 targets this second component by conditioning on the pre-event trend. In the model, this geographic rigidity is captured by the sparse adjacency matrix of active links—buyers cannot costlessly switch sourcing locations but are tethered to specific production hubs.

Fact 3. Crucially, the data reject the hypothesis of a global market-clearing price for HBM. Figure 3 compares the unit values of exports to Taiwan against the rest of the world. A persistent, widening premium emerges for Taiwan-bound shipments during the AI boom (Panel A). Furthermore, within Taiwan, the unit value of MCP–IC diverges sharply from



(a) Unit values of Korean MCP–IC exports to Taiwan vs. others (12m MA, policy shading).



(b) Unit values to Taiwan by item (MCP–IC, DRAM, Flash; USD/kg, 12m MA).

Figure 3: Reallocation toward Taiwan and rising prices in the Taiwan-directed channel. MCP–IC (HS 8542323000) records recent HBM-bearing packages. Panel A unit values (USD/kg) peak above 200,000 USD/kg for the Taiwan-bound MCP–IC stream; the scale reflects the extreme value concentration of HBM-attached accelerator modules.

that of DRAM and Flash (Panel B).

If the upstream market were perfectly competitive, arbitrage would equate quality-adjusted prices across destinations. Instead, the persistent Taiwan premium suggests that prices are determined through relationship-specific bargaining, in which high-value buyers (accelerator designers using the Taiwan hub) face different terms of trade than other users. This violation of the Law of One Price motivates the Nash-in-Nash assumption in Section 3, where upstream prices p_{du}^U are negotiated link by link based on each buyer's marginal value and outside options. These destination-level unit values are aggregate proxies for the underlying bilateral link prices; they do not constitute direct observations of the link-specific negotiated prices p_{du}^U, which are unobserved (see Remark 1).

Three alternative explanations for the Taiwan premium deserve acknowledgement. *First, quality composition:* if Taiwan-bound shipments systematically contain higher HBM generations (e.g., HBM3E vs. HBM2E), the premium partly reflects a generational mix shift. This is mitigated by the within-code panel structure — the DRAM and NAND controls show no comparable Taiwan premium at the same time, and generational transitions affect all destinations simultaneously. *Second, packaging costs:* assembled HBM modules with more chiplet layers are mechanically more expensive per kilogram; the premium partly reflects

a cost passthrough. This confound motivates using *relative* unit values (MCP-IC minus DRAM within the same destination-year cell) rather than levels when constructing calibration targets. *Third, arbitrage is legally constrained:* EAR export restrictions and bilateral supply agreement clauses prohibit resale of restricted chips, so destination-level price dispersion is consistent with competitive posted pricing even absent market power. The premium inference therefore identifies not a competitive distortion but *relationship-specific* pricing — suppliers can maintain differentiated negotiated prices across buyers precisely because resale is prohibited, regardless of the competitive structure of upstream supply.

In sum, the trade data deliver three robust messages. First, MCP–IC exports grow explosively and reach a macroeconomic scale in Korea's export basket. Second, this growth is accompanied by a sharp reallocation of MCP–IC flows toward Taiwan, consistent with the emergence of a bottleneck hub for HBM-attached accelerators. Third, the Taiwan-directed channel exhibits pronounced and persistent unit-value premia relative to other destinations and other memory lines. These patterns motivate treating HBM as the key upstream bottleneck in the model, targeting HBM and accelerator concentration in the calibration, and focusing the subsequent event-study analysis on the Taiwan×MCP–IC cell as the primary treated margin through which downstream export controls propagate upstream.

2.3 Event-Study Evidence: Timing and Magnitude of the Policy Shock

To document the timing, direction, and magnitude of the policy shock on the Taiwan–MCP–IC corridor, I implement an event-study design around the two major policy regimes: the initial controls in October 2022 and the comprehensive update in October 2023. The design controls for confounds through rich fixed effects and residualization (described below), but causal identification rests on the structural model in Section 3, which recovers supply-chain

primitives from equilibrium moments rather than requiring a clean quasi-experiment.[3]

I define the treated unit as the Taiwan–MCP–IC export flow. Although the statutory incidence of U.S. export controls falls on Chinese destinations, the upstream incidence in this concentrated vertical network runs through the critical manufacturing bottleneck where HBM is integrated into AI accelerators. As identified in Section 2.2, this integration is overwhelmingly clustered in Taiwan; thus, the Korea-to-Taiwan corridor represents the upstream margin directly exposed to the policy shock. Figure 4 documents the destination divergence directly: Korea's MCP–IC exports to China and Hong Kong together accounted for 29–37% of the total in 2020–2022, while Taiwan's share was only 7–11%. Following the October 2022 controls, this pattern reversed sharply — Taiwan's share surged to 36% by 2024 and 52% by 2025, while the China-plus-Hong-Kong share fell to 14% and then 9%. This simultaneous expansion of the Taiwan corridor and contraction of the China channel confirms that the Taiwan–MCP–IC cell captures the active upstream margin through which the policy shock propagates.

Comparison items are DRAM (HS 8542321010) and NAND/Flash (HS 8542321030). These are natural controls: they share upstream wafer inputs and macro-demand drivers but do not face the specific advanced-packaging constraints of the AI vertical. The identifying assumption is that DRAM and NAND exports to Taiwan would have tracked MCP-IC absent the policy shock. While all three products are manufactured by the same Korean firms, DRAM and NAND trade on shorter-term contracts than MCP-IC. Appendix C.3 reports two alternative control designs that do not rely on cross-product comparisons: Design A uses within-MCP-IC variation across non-Taiwan destinations as the counterfactual (removing the cross-product assumption entirely), and Design B uses within-Taiwan variation across non-AI semiconductor HS codes (preserving the Taiwan-specific demand environment). Both designs produce ATT estimates consistent in sign and order of magnitude with the baseline,

[3]For an overview of U.S.–China technology competition and export control policy, see Branstetter (2024). A comparable identification setting arises from Japan's 2019 export restrictions on semiconductor materials to Korea, studied by Kim et al. (2024).

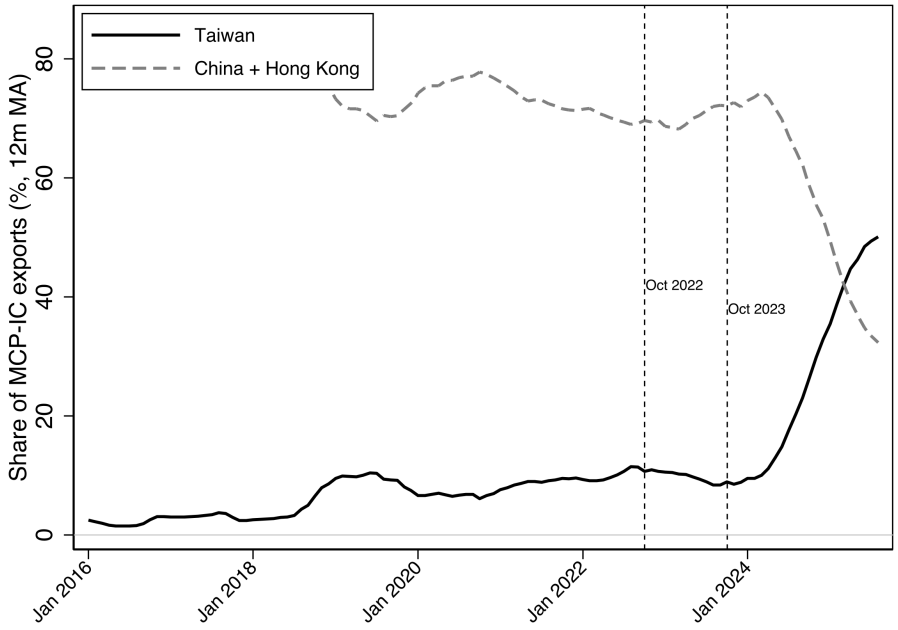


Figure 4: Korea's MCP–IC exports by destination: Taiwan vs. China + Hong Kong (share of total, 12-month moving average). Vertical dashed lines mark the October 2022 and October 2023 U.S. export-control events. The divergence confirms that the Taiwan corridor is the active upstream margin exposed to the policy shock, while the China channel contracts simultaneously.

supporting the parallel trends assumption for the main specification.

A key interpretive concern is that MCP–IC exports might follow a different trend than commodity memory due to technology-specific shocks (e.g., the generative AI boom, see Appendix C.1 for a detailed timeline) or distinct seasonality. To sharpen the descriptive comparison, I include a rich set of fixed effects and covariates that purge common confounds and isolate the Taiwan–MCP–IC cell's differential response. First, I include destination-by-month fixed effects to absorb local demand shocks in Taiwan. Second, I include item-by-month fixed effects to strip out global shocks common to the memory industry. Finally, to control for the AI-specific technology cycle, I residualize all outcomes against a global AI demand shifter — proxied by NVIDIA data center revenue — and interact it with pre-period exposure.[4] This ensures that the results capture the *differential* response of the Taiwan hub

[4]Formally, I estimate residuals $\tilde{y}_{it}$ from the projection $y_{it} = \alpha_i + \mu_t + \gamma_i t + \delta_{im} + \beta(\text{AI}_t \times W_i) + \varepsilon_{it}$, where α_i and μ_t are unit and time fixed effects, $\gamma_i t$ is a unit-specific trend, and δ_{im} captures monthly seasonality. The event study is then performed on these residuals. Appendix C presents robustness specifications, including

to U.S. policy, net of the secular technology cycle.

Figure 5 (Top Panels) plots the dynamic response around the October 2022 controls. The results show a sharp, immediate contraction: relative to controls, Taiwan–MCP–IC flows fall by approximately 15–20 log points in the initial post-event months (the twelve-month average treatment effect, reported in Appendix Table 10, is larger in magnitude as enforcement deepens over the post-period). Joint F-tests of the eleven pre-period coefficients ($k = -12$ through $k = -2$) reject the null of joint equality to zero: $F(11, \cdot) = 4.63$ ($p < 0.001$) for export value and $F(11, \cdot) = 18.86$ ($p < 0.001$) for quantity. The rejected pre-trends partly reflect the HBM3 development ramp at SK hynix documented in Appendix C.1, which differentially elevated MCP–IC relative to commodity memory before any policy intervention. This pre-existing divergence prevents strict causal attribution; the event-study coefficients are best interpreted as descriptive decompositions of the post-shock patterns rather than as clean treatment effects. Appendix C.4 shows that the post-event ATT is virtually unchanged when the pre-period is shortened to $k = -7$ to -2 (excluding the HBM3 ramp months), confirming that the early pre-trend divergence does not contaminate the average treatment effect estimates. For export value, a restricted test over the immediate pre-period ($k = -7$ through $k = -2$) yields $F(6, \cdot) = 1.25$ ($p = 0.294$), indicating a relatively flat pre-policy path in that narrower window. For export quantity, the corresponding $F(6, \cdot) = 6.58$ ($p < 0.001$) reflects anticipatory front-loading of shipments ahead of the October 2022 deadline; this pre-event inventory build-up contextualizes the sharp post-shock contraction as the partial unwind of accumulated stocks.[5] This pattern is consistent with an initial "compliance chill," in which uncertainty about the new Foreign Direct Product Rule led to a temporary freeze on high-end shipments.

stacked event studies (Appendix C.2) and alternative control groups (Appendix C.3), which confirm that the results are not driven by the residualization choice.

[5]The 2023 event exhibits a similar pattern: $F(11, \cdot) = 6.62$ ($p < 0.001$) for export value and $F(11, \cdot) = 5.60$ ($p < 0.001$) for quantity over the full pre-period. Restricted to the immediate window ($k = -7$ through $k = -2$), $F(6, \cdot) = 4.37$ ($p < 0.01$) for value and $F(6, \cdot) = 2.28$ ($p = 0.049$) for quantity, consistent with anticipatory pre-positioning ahead of the announced 2023 update. Because pre-trends are present, these estimates are descriptive magnitudes rather than causal treatment effects; they serve as empirical targets for the structural model's calibration in Section 3 rather than as standalone causal estimates.

The calibration in Section 4 targets the *post-shock average* treatment effects (Appendix Table 10) rather than the dynamic pre-policy path, so the pre-trend non-flatness does not contaminate the moments used to identify the structural parameters. Specifically, the calibration uses the direction and relative magnitude of the price and quantity responses across the two regimes, not the level of the pre-shock path.

The pattern reverses dramatically following the October 2023 update (Bottom Panels). Despite tighter statutory thresholds, the Taiwan–MCP–IC margin exhibits a large, sustained expansion relative to the control group. Both quantities and export values surge, driving the unit-value premium documented in Fact 3. This opposite-signed response highlights the non-neutrality of the supply chain adjustment: by late 2023, the industry had reorganized to aggressively reallocate capacity to the compliant Taiwan hub, creating the scarcity rents that drive the upstream bargaining results in Section 4. Appendix Table 10 tabulates the average post-event treatment effects for both regimes and both outcomes. Because the design has a single treated unit, standard asymptotic inference may be unreliable; Appendix J reports Fisher permutation p-values obtained by reassigning treatment to each of the 53–54 control units in turn. The October 2022 quantity contraction is significant at the 5% level ($p = 0.019$); the remaining outcomes have $p \in [0.13, 0.23]$, consistent with the wide confidence bands visible in Figure 5. Figure 15 in Appendix J plots the full placebo distribution for each event–outcome pair, showing that the October 2022 quantity estimate lies at the extreme left tail of the distribution.

These reduced-form patterns map directly to the primitives of the bilateral oligopoly model in Section 3. The destination-specific unit-value premia in Fact 3 are the empirical fingerprint of relationship-specific pricing: they arise when each buyer–supplier pair negotiates bilaterally rather than taking a posted price. The divergence in prices and quantities across destinations confirms that the upstream market is not globally integrated, and Section 3 formalizes this as a Nash-in-Nash equilibrium with link-specific negotiated prices p^U_{du}. Furthermore, the sharp reallocation observed in Regime 2 provides the empirical basis for

the counterfactuals in Section 4, where I model export controls as buyer-specific wedges (τ_d) that shift the outside options of HBM suppliers, generating the endogenous price incidence. Causal identification of the upstream price incidence mechanism comes from the equilibrium moment conditions used to calibrate the structural model in Section 4 — not from the event-study design, which documents the observable policy shock but does not recover the structural parameters driving it.

3 A Vertical Bargaining Model of the AI/HBM Chain

This section presents a bilateral oligopoly model in which a few producers dominate each stage of vertical production. A small set of downstream accelerator designers purchase HBM from a small set of upstream suppliers and compete in quantities in the accelerator market. For any given vector of negotiated HBM link prices, each downstream firm produces with a CES technology. I introduce Nash-in-Nash bargaining over the HBM link prices $\{p^U_{du}\}$ between each accelerator designer d and its qualified HBM suppliers indexed by u. The equilibrium prices and quantities at both tiers are jointly determined by (i) downstream Cournot competition and (ii) upstream bilateral negotiations that split surplus according to bargaining weights $\boldsymbol{\gamma}$. Export controls enter through changes in buyer–specific output wedge τ_d (market access or effective quality in controlled destinations). The framework deliberately abstracts from dynamic and extensive-margin decisions[6] to isolate the input-market incidence channel: how a buyer-specific export wedge feeds back into upstream input prices and downstream markups in a concentrated, small-n bargaining environment.

[6]Firms make a wide range of dynamic choices, including firm and production-network dynamics (Lim, 2018; Oberfield, 2018), innovation and R&D (Klette and Kortum, 2004; Arkolakis et al., 2018), foreign direct investment and multinational production (Helpman et al., 2004; Ramondo and Rodríguez-Clare, 2013), inventory management over the business cycle and in trade (Alessandria et al., 2010, 2011, 2013), long-term and relational contracting (Antras and Helpman, 2004; Antras and Foley, 2015), and capacity choice (Kreps and Scheinkman, 1983), among others

3.1 Notation and Sequence of Moves

Although the model is static, it is useful to structure decisions in three stages:

1. For each active upstream–downstream pair (d, u), firms bargain over the link price p^U_{du}, taking all other link prices as given. A Nash-in-Nash equilibrium is a fixed point of these bilateral problems.

2. Given the negotiated upstream prices $\{p^U_{du}\}$, unit costs of downstream producers become common knowledge.

3. Downstream producers simultaneously choose quantities in the accelerator market.

To focus on intensive-margin adjustments in upstream prices, I treat the set of active links between upstream and downstream producers as fixed. I start from the last stage (Cournot competition with unit costs), map unit costs to upstream prices, and then define the Nash-in-Nash bargaining problem over link prices. Table 3 summarizes the main notation.

A. Indices and sets	
D, U	sets of downstream and upstream firms
$d \in \mathsf{D},\ u \in \mathsf{U}$	individual downstream / upstream firm
$\mathsf{q}, \mathsf{p}, \mathsf{C}$	generic vectors of quantities, prices, and costs
B. Variables (time index t implicit)	
$z_d,\ z_u$	productivities (down-/up-stream)
p^D_d, q^D_d, C^D_d	downstream price, quantity, and unit delivery cost
$p^U_{du}, q^U_{du}, C^U_{du}$	input price, quantity, and unit delivery cost from u to d
P^M_d	materials (composite input) price index for downstream firm d; the superscript M distinguishes the upstream input bundle from the downstream output price p^D_d
C. Parameters	
σ, ρ, η	substitution elasticities
γ_d	bargaining leverage of downstream firms

Table 3: Notation guide.

3.2 Technology and Cost Structure

The final-good sector demands accelerator varieties as imperfect substitutes. The inverse demand for firm d's output is

$$p_d^D = A\left(\frac{q_d^D}{Q^D}\right)^{-1/\sigma}(Q^D)^{-1/\rho}, \tag{1}$$

where $A > 0$ is a demand shifter (normalized to pin equilibrium scale; in counterfactuals A is held fixed at its calibrated value) and

$$Q^D = \left[\sum_{d' \in \mathsf{D}} \left(q_{d'}^D\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

is a CES aggregator across downstream firms. The parameter $\sigma > 1$ governs substitution across accelerator producers, while $\rho > 1$ captures substitution between the aggregate accelerator composite Q^D and other inputs used by final-good producers.[7]

Downstream firms aggregate inputs from upstream suppliers u using a CES technology with substitution elasticity η.

$$q_d^D = z_d\left(\sum_u (q_{du}^U)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}, \tag{2}$$

where z_d is downstream producer d's productivity. q_{du}^U is the quantity of the upstream input sourced by downstream producer d from upstream supplier u.

Define the materials price index for firm d:

$$P_d^M = \left(\sum_u \left(p_{du}^U\right)^{1-\eta}\right)^{\frac{1}{1-\eta}}. \tag{3}$$

where p_{du}^U denotes the price of upstream firm u's products supplied to downstream firm d.

[7] A standard microfoundation has monopolistically competitive final-good firms with a nested CES technology: an inner nest over accelerator varieties with elasticity σ, and an outer nest over the accelerator composite and other inputs with elasticity ρ.

The cost function dual to (2) is

$$C_d^D(q_d^D; \{p_{du}^U\}) = \frac{\tau_d}{z_d} P_d^M \, q_d^D,$$

where $\tau_d > 0$ captures buyer-specific output frictions (e.g., export controls, market-access wedges), which are the empirical objects studied in Section 2. P_d^M is the CES materials price index defined in (3). Normalizing by q_d^D yields the unit delivery cost

$$C_d^D = \frac{\tau_d}{z_d} P_d^M. \tag{4}$$

By Shephard's lemma, the conditional demand for input u is

$$q_{du}^U = \frac{q_d^D}{z_d} \left(\frac{p_{du}^U}{P_d^M} \right)^{-\eta}. \tag{5}$$

Hence, the expenditure share on link (d, u) is $\omega_{du} = \left(p_{du}^U / P_d^M\right)^{1-\eta}$, and own-price elasticity is $-\eta$.

The upstream producers use labor as the only input — that is, upstream unit costs depend only on productivity and pair-specific delivery frictions, not on intermediate inputs purchased from other upstream firms. This is the natural assumption for HBM production, where the binding bottleneck is advanced-packaging capacity at the downstream assembly stage rather than the availability of wafer inputs at the HBM fabs themselves. The unit delivery cost of an upstream firm u supplying to a downstream firm d depends on the pair-specific trade frictions and the productivity that pins down the cost at the factory gate

$$C_{du}^U = \frac{f_{du}}{z_u} \tag{6}$$

where z_u is the productivity of u, f_{du} is the pair-specific trade friction between d and u (*e.g.*, logistics, reliability, or qualification costs). The notation f_{du} distinguishes this upstream

delivery friction from the buyer-specific output wedge τ_d in (4) and from the iceberg trade costs $\kappa_{n'i}$ in the multi-country extension (Section 3.6).

3.3 Firm Values and Bargaining Surpluses

Because the value chain has few producers, each upstream–downstream pair generates surplus by agreeing to trade. The terms of trade, upstream product prices p^U_{du}, determine how they divide the surplus.

Downstream surplus Let boldface characters denote the vectors of downstream variables: $\mathsf{q}^D = (q^D_d)_d$ and $\mathsf{C}^D = (C^D_d)_d$ are vectors of production quantity and unit cost. The downstream producers' surplus from the value chain arises from the oligopolistic competition.

$$\Pi^D_d = (p^D_d(\mathsf{q}^D) - C^D_d)q^D_d \tag{7}$$

Each downstream producer chooses its quantity, treating competitors' unit production costs as given. I write the equilibrium downstream quantity as a function of unit production cost C^D.

Definition 1. *The downstream Cournot equilibrium given unit production cost* C^D *is* $\mathsf{q}^{D*}(\mathsf{C}^D)$ *where* $\mathsf{q}^{D*}_d(\mathsf{C}^D)$ *maximizes* (7) *with* $\mathsf{q}_{d'} = \mathsf{q}^*_{d'}(\mathsf{C}^D)$ *for* $d' \neq d$.

Let $\Pi^D_d(\mathsf{q}^D; \mathsf{C}^D)$ denote firm d's Cournot profit in (7). Given a cost vector C^D, a (pure-strategy) Cournot equilibrium is a profile $\mathsf{q}^{D*}(\mathsf{C}^D)$ satisfying

$$q^{D*}_d(\mathsf{C}^D) \in \arg\max_{q^D_d} \Pi^D_d\left(q^D_d, \mathsf{q}^{D*}_{-d}(\mathsf{C}^D); \mathsf{C}^D\right) \quad \text{for all } d. \tag{8}$$

Firm d's downstream value as a function of unit costs is

$$\tilde{\Pi^D}_d(\mathsf{C}^D) := \Pi^D_d\left(\mathsf{q}^{D*}(\mathsf{C}^D); \mathsf{C}^D\right).$$

Combining this with the cost map (4) yields

$$\Pi_d^D(\mathsf{p}^U) = \tilde{\Pi^D}_d\Big(\mathsf{C}^D(\mathsf{p}^U)\Big), \tag{9}$$

so that downstream profits depend on upstream prices only through the cost vector $\mathsf{C}^D(\mathsf{p}^U)$.

Upstream surplus Upstream profits depend on derived demand from downstream producers. $q_d^D(\mathsf{C}^D(\mathsf{p}^U))$ is the production in the downstream pure strategy Nash equilibrium in Definition 1.

Combining the upstream producer's delivery cost in (6) and the factor demand of downstream producer in (5), the profit of upstream producer u is

$$\Pi_u^U(\mathsf{p}^U) = \sum_d (p_{du}^U - C_{du}^U) q_{du}^U(\mathsf{p}^U) \tag{10}$$

3.4 Bilateral Oligopoly Equilibrium

I model the bilateral oligopoly using the Nash-in-Nash bargaining framework between the upstream and downstream firms (Alviarez et al., 2025; Bagwell et al., 2020).[8] Each pair maximizes the geometric average of contributions to each other by choosing upstream prices. I do not allow side payments; in particular, I rule out global transfers across links. Each upstream–downstream pair bargains over the price on its own link, and surplus is split locally rather than through firm-wide transfers.[9]

Let $\mathsf{p}^U_{-(d,u)}$ denote the vector of upstream prices between firms other than pair (u,d). Namely, $\mathsf{p}^U_{-(d,u)} = (p_{d'u'} : d' \neq d \text{ or } u' \neq u)$. Define the Nash product of pair (d,u) as the Nash product of the incremental surpluses, weighted by the bargaining power of d, denoted

[8] Bargaining is a standard way to model two parties in production networks that split the surplus generated by the contributions from both. Dhyne et al. (2023); Eaton et al. (2022a,b); Lamadon et al. (2022) feature producers and their factor suppliers dividing the surplus by bargaining.

[9] Specifically, there are no additional lump–sum transfers and no firm–wide or cross–link transfers that could rebalance surplus independently of the marginal input price. All surplus splitting on a given link is therefore implemented through the negotiated price p_{du}^U itself.

by γ_d

$$NP_{du}(p^U_{du}; \mathsf{p}^U_{-(d,u)}) = (\Pi^D_d - \Pi^D_{d,-u})^{\gamma_d} \times (\Pi^U_u - \Pi^U_{-d,u})^{1-\gamma_d} \tag{11}$$

where Π^D_d and Π^U_u are defined in (9) and (10). $\Pi^D_{d,-u}$ and $\Pi^U_{-d,u}$ are the value of d and the periodic profit without trade between u and d, obtained by removing link (d,u) from the set of active links (equivalently, letting $p^U_{du} \to \infty$) so that the bilateral share of d within u $\omega_{du} \to 0$ and the link carries no quantity in (9) and (10).

Definition 2. *The solution to the Nash-in-Nash bargaining between upstream-downstream pairs is the prices of upstream products $\mathsf{p}^{U*} = (\mathsf{p}^{U*}_{du})_{du}$ that maximizes the Nash products of upstream-downstream pairs in (11), taking the rest of upstream prices as given. The Nash-in-Nash equilibrium is defined as the fixed point of these simultaneous bilateral bargaining problems. Namely,*

$$\begin{aligned} & p^{U*}_{du} = \arg\max NP_{du}(p^{U*}_{du}; \mathsf{p}^{U*}_{-(d,u)}) \\ & \textit{where } NP_{du}(\mathsf{p}^{U*}_{du}; \mathsf{p}^{U*}_{-(d,u)}) = (\Pi^D_d - \Pi^D_{d,-u})^{\gamma_d} \times (\Pi^U_u - \Pi^U_{-d,u})^{1-\gamma_d} \\ & \textit{s.t. } \Pi^D_d \geq \Pi^D_{d,-u} \textit{ and } \Pi^U_u \geq \Pi^U_{-d,u}. \end{aligned} \tag{12}$$

The bargaining outcome p^{U*} pins down the downstream unit production costs $\mathsf{C}^{D*} = \mathsf{C}^D(\mathsf{p}^{U*})$.

Definition 3 (Bilateral oligopoly equilibrium)**.** *Suppose exogenous firm-level productivities $\mathsf{z} = (z_u, z_d)_{u \in \mathsf{U}, d \in \mathsf{D}}$ and bargaining weights $\gamma = (\gamma_d)_{d \in \mathsf{D}}$.*

1. ***Upstream–downstream bargaining.*** *The input-price vector $\mathsf{p}^{U*} = (p^U_{du})_{d,u}$ solves the Nash-in-Nash problem in (11), with downstream values Π^D_d and upstream profits Π^U_u as in (9) and (10), and I adopt* passive beliefs *(Horn and Wolinsky, 1988; Collard-Wexler et al., 2019; de Fontenay and Gans, 2014): upon breakdown of link (d,u), all other pairs maintain their equilibrium agreements and quantities. Disagreement payoffs therefore correspond to removing link (d,u) from the active set only, holding all other agreements fixed.*

2. ***Unit-cost mapping.*** *Given* p^{U*}*, downstream unit costs are* $\mathsf{C}^D = \mathsf{C}^D(\mathsf{p}^{U*})$ *as in* (4).

3. ***Downstream Cournot competition.*** *With costs* C^D*, each downstream firm chooses* q_d^D *to maximize* (7)*, taking rivals' quantities as given. Let* $\mathsf{q}^{D*} = \mathsf{q}^{D*}(\mathsf{C}^D)$ *denote the (unique) Cournot profile characterized by* (8).

A collection $\left\{\mathsf{p}^{U*}, \mathsf{C}^D, \mathsf{q}^{D*}, \right\}$ *is a bilateral oligopoly equilibrium if 1–3 hold.*

Remark 2 (Passive vs. active beliefs)**.** *Definition 3 adopts* passive beliefs*: upon breakdown of link* (d, u)*, all other pairs maintain their equilibrium agreements. An alternative is* active beliefs*, where parties anticipate that other bilateral pairs will renegotiate if the current link collapses (de Fontenay and Gans, 2014). Passive beliefs are standard in the empirical NiN literature (healthcare, cable television) because they yield a tractable fixed-point problem with unique comparative statics. Under active beliefs the renegotiation of other links upon breakdown introduces a recursive system that is typically not solvable in closed form, making it difficult to derive the shock-absorber sign analytically. Because my goal is a sharp comparative-statics result for buyer-specific trade wedges, passive beliefs are the appropriate choice. Quantitatively, Collard-Wexler et al. (2019) show that passive and active beliefs deliver similar equilibrium price levels in cable-TV bargaining; the directional comparative statics I derive — that a constrained buyer's input price falls while rivals' prices rise — rely on the M-matrix structure of the passive-beliefs Jacobian and would require separate verification under active beliefs. Note, however, that the shock-absorber sign depends on the log-submodularity of buyer incremental surplus in* (p_L, τ_d) *— a condition governed by cost structure and demand curvature, not by the belief protocol itself. Since the buyer's incremental surplus function is determined by the Cournot block and CES technology (Remark 3), the log-submodularity condition is the same whether beliefs are passive or active; the main effect of switching belief protocols is to change the quantitative level of equilibrium prices, not to overturn the sign of the comparative static.*

Intuitively, at a bilateral oligopoly equilibrium, upstream prices, downstream costs, and

quantities are jointly determined by (i) bilateral bargaining over link prices and (ii) Cournot competition in the accelerator market. Numerically, I compute this equilibrium by nesting a Cournot solver inside a link-by-link Nash-in-Nash update over upstream prices; see Appendix I for algorithmic details and convergence diagnostics.

3.5 Theoretical Properties: Uniqueness and Comparative Statics

I characterize the model in two steps. First, I exploit the nested-CES structure, which yields strictly concave bilateral bargaining problems and a unique joint equilibrium price vector. I then characterize how "Chip War" shocks — modeled as exogenous increases in a buyer's output-market friction τ_d — propagate through the system.

3.5.1 Uniqueness and Comparative Statics of the Nash-in-Nash System

Fix the set of active links and focus on equilibria in which all active links have nontrivial intensive shares (i.e., $\omega_{du} \equiv (p^U_{du})^{1-\eta} / \sum_{u'} (p^U_{du'})^{1-\eta} \in (\varepsilon, 1-\varepsilon)$ for some $\varepsilon > 0$).

Assumption 1 (Active-link interiority ("if you use it, use it"))**.** *For every buyer d and every* active *link (d,u) at the equilibrium,*

$$\varepsilon \leq \omega_{du} \leq 1-\varepsilon \quad \textit{and} \quad p^{U*}_{du} - C^U_{du} \geq m,$$

for some constants $\varepsilon \in (0, 1/2)$ and $m > 0$.

Assumption 1 rules out knife-edge cases where an active link carries negligible weight or zero markups. I pair it with ε–uniform curvature/small–gain conditions that ensure each link's own–price effect in the log–Nash FOC dominates cross–link feedback effects via the CES aggregator and Cournot block.

Assumption 2 (Curvature and bargaining-weight ceiling ($\gamma_d < \overline{\gamma}$))**.** *Maintain Assumption 1 with the same $\varepsilon \in (0, 1/2)$ and $m > 0$; hence every active link satisfies $\varepsilon \leq \omega_{du} \leq 1-\varepsilon$. For every active buyer d:*

(a) ***Downstream Demand Elasticity.***

$$-\frac{\partial \ln q_d^{D*}}{\partial \ln C_d^D} \;\geq\; \eta\,\frac{1-\varepsilon}{\varepsilon}. \tag{13}$$

(b) ***Bargaining–weight ceiling.***

$$\gamma_d \;<\; \overline{\gamma}, \qquad \overline{\gamma} = 1 - \frac{1}{\eta}\Big(\frac{1}{\rho} - \frac{1}{\sigma}\Big).$$

Assumption 2 (a) says the downstream quantity is not too inelastic in its own cost (strong downstream curvature), and (b) is an upper bound on the buyer's bargaining weight that ensures the Gershgorin diagonal-dominance condition holds: cross–link feedback effects (via the CES index P_d^M and Cournot spillovers) are dominated by own–link curvature. Since $\sigma > \rho > 1$, the ceiling $\overline{\gamma} = 1 - (1/\eta)(1/\rho - 1/\sigma) \in (0,1)$ is a well–defined interior point (given $\eta > 1/\rho - 1/\sigma$, which holds whenever $\eta > 1$ and $\rho \geq 1$, as maintained throughout). This assumption furnishes a sufficient, purely primitive condition under which (a) guarantees log–concavity of the buyer's incremental surplus uniformly over active links, while (b) ensures the own–link curvature of the log–Nash FOC dominates cross–link feedback effects from CES aggregation and Cournot feedback.[10] The next result combines the demand curvature, Cournot uniqueness, and one–link log–Nash concavity, and shows that stacking the link FOCs yields a P-matrix Jacobian and a unique joint price vector. A brief proof sketch follows; full details are in Appendix E.

Proposition 1 (NiN system: uniqueness and comparative statics)**.** *Maintain Assumptions 1–2. Suppose the Cournot block satisfies*

$$\sigma \leq 2\rho. \tag{14}$$

[10]Three illustrative parameter tuples (σ, ρ, η) with the admissibility ceiling $\overline{\gamma}$ and an example admissible γ: $(6, 3, 2;\ \overline{\gamma} \approx 0.917)$, e.g. $\gamma = 0.75$; $(8, 4, 1.5;\ \overline{\gamma} \approx 0.917)$; $(5, 2, 3;\ \overline{\gamma} = 0.900)$. For the calibrated baseline (σ=6, ρ=4, η=2.5), the ceiling is $\overline{\gamma} \approx 0.967$; the calibrated γ=0.5 satisfies this bound, so Proposition 1 applies analytically. The equilibrium algorithm also converges smoothly across the full sensitivity grid in Table 9.

Then, the Nash-in-Nash price vector p^{U*} *over all active links exists and is the unique zero of the log–Nash pseudo-gradient* $g(\mathsf{p}^U) = 0$.

Proof sketch. Three steps: (i) CES demand curvature (Lemma 2) and $\sigma \leq 2\rho$ yield a Cournot block with diagonally dominant best replies (Lemma 3); (ii) holding rival link prices fixed, each link's log-Nash objective is strictly concave (Lemmas 4–5); (iii) stacking link FOCs yields a Jacobian J with strictly negative diagonals and bounded cross terms, so the Gershgorin argument on the symmetrized scaled matrix $S = (RJ + J^\top R)/2$ gives $S \prec 0$, making J a P- and M-matrix (Lemma 6); Rosen's theorem delivers the unique fixed point. Full details are in Appendix E. □

This result pins down a unique vector of negotiated link prices p^{U*}. Given p^{U*}, the cost map $\mathsf{C}^D(\mathsf{p}^{U*})$ and the uniqueness of the Cournot block imply a unique downstream quantity profile q^{D*} and associated profit vectors. Hence, the entire bilateral oligopoly equilibrium is unique.

3.5.2 Comparative Statics: Buyer-Specific Wedges

I model a "Chip War" episode as an exogenous increase in a downstream firm's output-market friction τ_d in (4). Unlike conventional trade models, upstream prices p^U adjust endogenously: as the shock alters downstream quantities, Nash-in-Nash bargaining reprices each active link.

Proposition 2 (Comparative statics in Nash-in-Nash)**.** *Maintain Assumptions 1–2 and the Cournot regularity condition in Proposition 1. Let* Θ *collect the primitives* $(\boldsymbol{\tau}, \boldsymbol{z}, \boldsymbol{\gamma}, \ldots)$ *and let* $\mathbf{p}^{U*}(\Theta)$ *be the unique Nash-in-Nash price vector from Proposition 1. Denote by* $g(\mathbf{p}^U, \Theta)$ *the stacked log–Nash pseudo–gradient and by* $J(\mathbf{p}^U, \Theta) := \partial g(\mathbf{p}^U, \Theta)/\partial \mathbf{p}^U$ *its Jacobian.*

(a) ***Regularity and derivative formula.*** *For any* Θ *in a neighbourhood of the baseline, the equilibrium upstream price vector* $\mathbf{p}^{U*}(\Theta)$ *is continuously differentiable in* Θ *and satisfies*

$$\frac{\partial \mathbf{p}^{U*}}{\partial \Theta_k} \;=\; -\,J\Big(\mathbf{p}^{U*}(\Theta), \Theta\Big)^{-1}\, \frac{\partial g(\mathbf{p}^U, \Theta)}{\partial \Theta_k}\bigg|_{\mathbf{p}^U = \mathbf{p}^{U*}(\Theta)}$$

for every scalar primitive Θ_k. Moreover, under Assumptions 1–2 the Jacobian $J\Big(\mathbf{p}^{U*}(\Theta),\Theta\Big)$ is a P– and M–matrix, so that

$$-J\Big(\mathbf{p}^{U*}(\Theta),\Theta\Big)^{-1} \quad \textit{is entrywise nonnegative.}$$

(b) ***Buyer–specific output wedge and productivity.*** *Fix a buyer d and an active link $L=(d,u)$ with expenditure share $\omega_{du}\in(\varepsilon,1-\varepsilon)$ and positive markup $p^{U*}_{du}-C^{U}_{du}>0$ at the baseline. Then the elasticities*

$$\frac{\partial \ln p^{U*}_{du}}{\partial \ln \tau_d}, \qquad \frac{\partial \ln p^{U*}_{du}}{\partial \ln z_d}$$

are well defined and can be written as

$$\frac{\partial \ln p^{U*}_{du}}{\partial \ln \tau_d} \;=\; \Psi^{(\tau)}_{d,u}(\Theta), \qquad \frac{\partial \ln p^{U*}_{du}}{\partial \ln z_d} \;=\; \Psi^{(z)}_{d,u}(\Theta),$$

for continuous functions $\Psi^{(\tau)}_{d,u}$ and $\Psi^{(z)}_{d,u}$ that depend on local demand curvature, the CES parameters, and the equilibrium shares.

For any rival buyer $d'\neq d$ and any of its active links $L'=(d',u')$ the corresponding elasticities $\partial \ln p^{U*}_{d'u'}/\partial \ln \tau_d$ and $\partial \ln p^{U*}_{d'u'}/\partial \ln z_d$ are given by the same implicit–function formula in part (a), with signs determined by the local shock vector $\partial g/\partial\Theta$ and the nonnegative matrix $-J^{-1}$.

(c) ***Bargaining weights.*** *For the same buyer d and any of its active links $L=(d,u)$,*

$$\frac{\partial p^{U*}_{du}}{\partial \gamma_d} \;<\; 0.$$

For any rival buyer $d'\neq d$ and any of its active links $L'=(d',u')$,

$$\frac{\partial p^{U*}_{d'u'}}{\partial \gamma_d} \;\leq\; 0,$$

with strict inequality whenever L' is directly or indirectly connected to L in the network.

Proof sketch. Part (a) applies the Implicit Function Theorem to $g(\mathbf{p}^{U*}(\Theta), \Theta) = 0$; invertibility of J and $-J^{-1} \geq 0$ follow from Lemma 6. Part (b) establishes the derivative formula; the deflationary direction (shock-absorber sign) is characterised in Remark 3 and established analytically for the nested-CES Cournot model in Lemma 7 (under the full-incremental assumption $\Pi^D_{d,-u} = 0$; the linear-Cournot case with the same assumption is Lemma 1). For part (c), the only nonzero shock-vector entry is

$$\frac{\partial g_L}{\partial \gamma_d} = \frac{\partial}{\partial p^U_{du}} \left(\log \Delta\Pi^D_d - \log \Delta\Pi^U_u \right) < 0,$$

strictly negative because $\Delta\Pi^D_d$ is strictly decreasing and $\Delta\Pi^U_u$ strictly increasing in p^U_{du} (Lemma 4); all rival entries are zero. Multiplying by $-J^{-1} \geq 0$ gives $\partial p^{U*}_{du}/\partial \gamma_d < 0$ for the focal link and $\partial p^{U*}_{d'u'}/\partial \gamma_d \leq 0$ for rivals, with strict inequality along connected paths. Full derivations are in Appendix F. □

Remark 3 (Shock-absorber sign and the log-submodularity condition)**.** *Proposition 2(b) establishes that $|\partial \ln p^{U*}_{du}/\partial \ln \tau_d|$ is positive but does not globally sign the direction. Since the export wedge τ_d affects only the buyer's incremental surplus $\Delta\Pi^D_d$ (the seller's surplus $\Delta\Pi^U_u$ does not depend on the buyer's delivery cost), the derivative of the NiN first-order condition decomposes as*

$$\frac{\partial g_L}{\partial \ln \tau_d} = \gamma_d \frac{\partial^2 \log \Delta\Pi^D_d}{\partial p_L \, \partial \ln \tau_d}.$$

Because $-J^{-1} \geq 0$, the sign of $\partial \ln p^{U*}_{du}/\partial \ln \tau_d$ equals the sign of $\partial g_L/\partial \ln \tau_d$, and hence the sign of the cross log-derivative $\partial^2 \log \Delta\Pi^D_d/\partial p_L \, \partial \ln \tau_d$. The deflationary case ("shock absorber") therefore obtains if and only if $\Delta\Pi^D_d$ is log-submodular in (p_L, τ_d): a rise in the export wedge must amplify *the buyer's sensitivity to input prices. Intuitively, when the tariff squeezes the buyer's profit, preserving access to a cheap input link becomes disproportionately valuable, so the supplier finds it optimal to lower p^U_{du} to maintain the relationship—paralleling the curvature condition in Chipty and Snyder (1999).*

A useful simplification when the link-level first-order condition is evaluated holding all rival link prices fixed: τ_d scales the full unit cost $C_d^D = (\tau_d/z_d)P_d^M$ one-for-one (so $\partial \ln C_d^D / \partial \ln \tau_d = 1$), while $p_L = p_{du}^U$ enters $\ln C_d^D$ through the CES index with weight ω_{du}. Conditioning on the single-link FOC (which already holds rival prices fixed), the cross-derivative for the shock-absorber sign reduces to the second derivative of log-profit with respect to log-cost,

$$\frac{\partial^2 \log \Delta\Pi_d^D}{\partial p_L \, \partial \ln \tau_d} = \frac{\partial^2 \log \Delta\Pi_d^D}{\partial (\ln C_d^D)^2}. \tag{15}$$

The shock-absorber sign therefore holds if and only if log buyer-surplus is strictly concave *in log-cost. Lemma 1 below establishes this analytically for linear Cournot competition; for the CES model, the condition is verified numerically across the parameter grid in Table 9.*

Lemma 1 (Shock-absorber sign: linear Cournot)**.** *Suppose $D = 2$ and downstream Cournot competition takes place under linear inverse demand $p^D = a - b\sum_{d'} q_{d'}^D$, $a, b > 0$. If $\Pi_{d,-u}^D = 0$ (buyer d exits when link (d,u) is removed), then for all parameters consistent with positive equilibrium profits ($a > 2c_d - c_{d'}$),*

$$\frac{\partial^2 \log \Delta\Pi_d^D}{\partial (\ln C_d^D)^2} < 0,$$

and consequently $\partial \ln p_{du}^{U*} / \partial \ln \tau_d < 0$ (the shock-absorber direction holds strictly).

Proof. See Appendix G. □

Remark 4 (CES case)**.** *Under the CES demand structure of the main model, Lemma 7 in Appendix G establishes analytically that $\partial^2 \log \Delta\Pi_d^D / \partial (\ln C_d^D)^2 = -C_d^D \, d\epsilon_d / dC_d^D < 0$ for all $\sigma > \rho > 1$, under the full-incremental assumption $\Pi_{d,-u}^D = 0$. The mechanism is transparent: a cost increase lowers firm d's Cournot share s_d; since larger firms face more inelastic demand when $\sigma > \rho$ (Lemma 2), the fall in s_d raises ϵ_d, making $\log \Pi_d^D$ strictly more concave in log-cost. This result holds independent of η. In the calibrated multi-supplier model where $\Pi_{d,-u}^D > 0$, the sign is preserved numerically at all grid points with $\sigma \geq 6$ and $\eta \geq 2.5$ in*

Table 9. At $\eta = 1.5$, the sign condition is maintained but the channel produces a numerically negligible magnitude*: when input-bundle substitution is near-Leontief, expenditure shares ω_{du} barely shift with cost, so the log-Nash FOC moves by an amount indistinguishable from zero. This is a magnitude effect, not a sign reversal. Appendix G provides the full proof of Lemma 7 and reports the calibrated CES magnitudes.*

Corollary 1 (Cross-buyer spillover)**.** *Suppose buyers d and d' share at least one active upstream supplier ($a_{du} = a_{d'u} = 1$ for some u). Then a tighter export wedge on buyer d ($d \ln \tau_d > 0$) strictly raises the negotiated input price on every shared link: $\partial \ln p^{U*}_{d'u} / \partial \ln \tau_d > 0$. Consequently, buyer d''s materials price index strictly rises: $\partial \ln P^M_{d'} / \partial \ln \tau_d > 0$.*

Proof. By Proposition 2(a), $-J^{-1} \geq 0$ entrywise. Because buyers d and d' share supplier u, the pseudo-gradient $g_{L'}$ for $L' = (d', u)$ and the direct shock $\partial g_L / \partial \ln \tau_d$ for $L = (d, u)$ are coupled through the CES materials index $P^M_{d'}$ and the Cournot block, making the Jacobian J irreducible on the shared-link subgraph. Irreducibility of the M-matrix $-J$ implies $(-J)^{-1} > 0$ strictly on every entry corresponding to an active path, so $\partial \ln p^{U*}_{d'u} / \partial \ln \tau_d = [(-J)^{-1}]_{L'L} \cdot (-\partial g_L / \partial \ln \tau_d) > 0$ whenever $d \ln \tau_d > 0$. The claim for $\partial \ln P^M_{d'} / \partial \ln \tau_d$ then follows from the CES index formula. □

Corollary 1 is the object I confront with the event-study evidence in Section 2: where the data show opposite movements across regimes and buyers, the calibrated model generates elasticities of the same sign.

Proposition 2 establishes that, in the bilateral oligopoly, the trade wedge activates an additional channel through which the output price adjusts. Define the markup $\mu_d := p^D_d / C^D_d$; from Lemma 2, $\mu_d = \epsilon_d / (\epsilon_d - 1)$ in equilibrium. Decomposing the log-differential of the pricing rule $p^D_d = \mu_d (\tau_d / z_d) P^M_d$ yields the total pass-through identity:

$$d \ln p^D_d = \underbrace{d \ln \tau_d}_{\text{Direct shock}} + \underbrace{d \ln P^M_d}_{\substack{\text{Upstream incidence} \\ \text{(Input price channel)}}} + \underbrace{d \ln \mu_d}_{\substack{\text{Downstream incidence} \\ \text{(Markup channel)}}} . \tag{16}$$

This identity follows from $\ln p_d^D = \ln \mu_d + \ln \tau_d - \ln z_d + \ln P_d^M$; throughout the counterfactual exercises firm productivities z_d are held at their calibrated values, so $d \ln z_d = 0$ and the productivity term drops out. The frozen-input benchmark — which represents models that treat input prices as fixed at their calibrated baseline levels rather than allowing them to respond endogenously to the wedge shock (Appendix D) — sets $d \ln P_d^M = 0$, leaving only the markup response.[11] In contrast, this framework allows the input price index P_d^M to respond endogenously to the wedge. As shown in the quantitative results, this upstream incidence term is negative ($d \ln P_d^M / d \ln \tau_d < 0$), implying that suppliers lower their negotiated prices for the targeted buyer, partially offsetting the regulatory cost increase and dampening the total pass-through to final consumers.

3.6 Adding Countries and Multinational Production

This subsection embeds international trade and multinational production into the baseline model. The setup maps directly to the Korea–Taiwan corridor in Section 2.2: Korean HBM fabs ($n(j) =$ Korea) supply accelerator assembly plants ($n(i) =$ Taiwan), with iceberg trade costs $\kappa_{n'i}$ capturing cross-border logistics and compliance frictions. Because several inputs can act as bottlenecks simultaneously and to varying degrees, I also allow for multiple input types. Table 4 fixes notation for sites, firms, and input types.

	Production stage	Production site	Firm	Product type
Input factors	upstream	j	u	m
AI accelerators	downstream	i	d	$\cdot$
Output markets	final	n'	$\cdot$	$\cdot$

Table 4: Indices for sites, firms, and product types.

[11] This benchmark is distinct from a competitive-pricing counterfactual (where input prices equal marginal cost) or a Cournot upstream counterfactual. It is closest to the class of quantitative trade models that calibrate input prices from observed data and then hold them fixed in counterfactuals, treating input markets as integrated and unresponsive to buyer-specific shocks. The gap between the full-Nash-in-Nash and frozen-input bars therefore isolates precisely the endogenous bargaining channel that standard trade models abstract away from.

Countries, ownership, and input types Let $\mathcal{N}$ be countries, D downstream firms, U upstream firms, and $\mathcal{M}$ input types. Firm d operates plants $i \in \mathcal{I}_d$ with location map $n(i) \in \mathcal{N}$; similarly u operates plants $j \in \mathcal{J}_u$ with $n(j) \in \mathcal{N}$. Upstream firm u produces type $m(u) \in \mathcal{M}$.

Plant–specific costs and firm–level profits. For a buyer–plant i and a supplier–plant j, let $p^U_{ij,du}$ be the contract price. The materials price index faced by plant i is

$$P^M_{i,d} = \left(\sum_{u \in \mathsf{U}} \sum_{j \in \mathcal{J}_u} \left(p^U_{ij,du} \right)^{1-\eta} \right)^{\frac{1}{1-\eta}}.$$

Let $\kappa_{n'i} > 0$ be the iceberg cost to ship from plant i's country to destination n'. The unit cost (in plant-i currency) to serve n' is

$$C^{D,(n(i))}_{n',i,d} \;=\; \frac{\kappa_{n'i}}{z_d}\, P^M_{i,d}.$$

Profits are consolidated at the firm level. Firm d's profit is

$$\Pi^D_d = \sum_{n'} \sum_{i \in \mathcal{I}_d} \left(p^D_{n',i,d} - C^D_{n',i,d} \right) q^D_{n',i,d}.$$

Bargaining at the plant–pair level. Nash-in-Nash is conducted over the vector of plant–pair prices $\{p^U_{ij,du}\}_{i \in \mathcal{I}_d,\, j \in \mathcal{J}_u}$ for each firm pair (d, u), taking all other links as given. Buyer incremental surplus is computed from firm-level value with the *plant-specific* cost map above; seller incremental surplus aggregates the plant-pair sales.

The Cournot Jacobian's sign pattern and diagonal dominance are unchanged under the multi-country extension with multiple input types, and log–Nash product curvature is preserved as well. Hence, all existence/uniqueness and comparative-statics results in the baseline carry over in the multinational model; the three-step argument is given in Appendix H.

The model delivers a tightly parameterized map from primitives to outcomes in the

AI/HBM vertical. Given (i) the network of active links between HBM suppliers and accelerator designers, (ii) firm-level productivities $\boldsymbol{z}$, (iii) buyer-specific output wedges $\boldsymbol{\tau}$, and (iv) bargaining weights $\boldsymbol{\gamma}$, the Nash-in-Nash system pins down a unique vector of negotiated input prices p^{U*}, and hence unique downstream unit costs, quantities, and markups. The multi-country extension nests the empirical setting in Section 2: HBM and accelerator plants can be located in different countries, trade costs enter through downstream friction τ_d, and HBM scarcity is summarized in the materials price index P_d^M. In the next section, I feed calibrated AI/HBM primitives into this structure and study counterfactual "Chip War" shocks as changes in the buyer-specific wedges $\boldsymbol{\tau}$. The quantitative exercises trace how downstream export controls are transmitted upstream into HBM prices and across buyers, and how the resulting changes in P_d^M and markups decompose the observed policy incidence on prices and quantities.

4 Quantifying Upstream Incidence

I consider shocks that operate only through the three channels of the pass-through identity (16): the statutory wedge $\Delta \ln \tau_d$, the input-price response $\Delta \ln P_d^M$, and the markup adjustment $\Delta \ln \mu_d$. First, I allow output wedges τ_d to differ across buyers, holding the calibrated productivities $\boldsymbol{z}_d$ and $\boldsymbol{z}_u$ and the observed network of active links fixed. Second, I compare equilibria across "worlds" that differ only in vertical structure or bargaining weight vector $\boldsymbol{\gamma}$. One set of experiments toggles the Samsung↔NVIDIA link, keeping all other primitives fixed, to isolate how adding a single qualified HBM supplier for a dominant buyer reshapes negotiated prices and incidence.[12] Another set sweeps the bargaining weights $\boldsymbol{\gamma}$ away from the symmetric baseline, tracing how stronger buyer power reallocates the total ef-

[12]Industry reports indicate that, through 2023–mid-2024, NVIDIA sourced most of its HBM for AI accelerators from SK hynix (and, more recently, Micron), while Samsung lagged in both yields and qualification for NVIDIA's GPUs. Reuters notes that Samsung was still behind rivals in securing HBM supply deals with NVIDIA, and that only in 2024 did Samsung's 8-high HBM3E parts clear NVIDIA's tests (with 12-high still pending), after redesigns to mitigate heat and power-consumption problems.

fect of the policy shock between upstream input-price incidence, $\Delta \log P_d^M$, and downstream markups, $\Delta \log \mu_d$.

In the model, heterogeneous trade shocks across NVIDIA and AMD are captured by buyer–specific wedges τ_d. This maps naturally into the way recent U.S. export controls bite in practice. The rules are written at the product and destination level (performance thresholds, interconnect bandwidth, data-center location, end-user screening, licensing requirements), but from the perspective of a GPU vendor they show up as an increase in the effective cost of delivering accelerators to restricted buyers. Because firms differ in product mix, customer composition, and ability to re-engineer chips or reallocate supply, the same regulatory package can translate into very different effective wedges τ_d at the firm level.

It is therefore reasonable to think of scenarios in which NVIDIA is hit harder than AMD, modeled here as $\Delta\tau_{\text{NV}} > \Delta\tau_{\text{AMD}}$. This corresponds to environments where (i) NVIDIA's main SKUs for Chinese cloud and internet platforms lie squarely above the performance or interconnect thresholds, so they require licenses or substantial redesign, while AMD's relevant products fall just below; (ii) NVIDIA's China sales are more concentrated in controlled applications (large data-center training and inference) and less in "unrestricted" segments, so a larger share of its revenue is directly exposed; or (iii) AMD enters later and can design chips to the new thresholds from the outset, whereas NVIDIA faces an immediate disruption of an existing installed base. In all of these cases, the same legal shock is best represented as a larger increase in τ_d for NVIDIA than for AMD.

The opposite ordering, with $\Delta\tau_{\text{AMD}} > \Delta\tau_{\text{NV}}$, is also plausible ex ante. For example, if AMD's future accelerator portfolio is more heavily concentrated in very high-performance SKUs that exceed the control thresholds while NVIDIA manages to shift Chinese customers toward "de-featured" variants, the effective revenue loss or compliance cost per dollar of China demand could be larger for AMD. Likewise, if NVIDIA can diversify more rapidly across non-Chinese buyers or production sites, while AMD relies more heavily on China-facing design wins, then a common tightening of rules would again translate into a relatively

larger firm-level wedge for AMD.

The two heterogeneous-shock experiments in Section 4.3—a $(10\%, 5\%)$ increase in $(\tau_{\text{NV}}, \tau_{\text{AMD}})$ and the reverse $(5\%, 10\%)$—are best read as representations of these alternative policy environments: one in which NVIDIA is the more exposed firm and one in which AMD is. The decomposition and non-neutrality plots then show how, conditional on a given pair of wedges, the same legal shock propagates quite differently through upstream input prices and downstream markups, depending on which buyer is more tightly constrained. One additional clarification: the model treats τ_d as the firm-level effective output friction rather than a statutory rate applied uniformly to all of the firm's output. In practice, the appropriate calibration of τ_d is the China-revenue-share-weighted effective constraint — that is, the increase in the cost of serving the subset of the firm's customer base that is directly affected by destination-specific performance thresholds. For NVIDIA, which in 2022 derived a substantial share of data-center revenue from Chinese cloud customers, the effective τ_d is materially larger than it would be under a calculation that uses NVIDIA's global revenue. The 10% and 5% scenarios are therefore indicative of different assumptions about China-exposure shares rather than literal statutory rate increases; they bracket the plausible range given public estimates of China exposure. See Appendix D for details on shock protocols and the frozen-input benchmark.

I organize the results in three steps: (i) measuring input-price incidence by comparing the full Nash-in-Nash equilibrium against counterfactuals without endogenous bargaining; (ii) comparing incidence with and without the Samsung↔NVIDIA link (Figure 7); and (iii) sweeping a single buyer's wedge τ_d and bargaining weight γ_d to trace cross-buyer spillovers (Figure 8) and upstream repricing (Figure 9).

4.1 Calibration and Quantitative Setting

I calibrate the model in Section 3 to a stylized representation of the AI/HBM vertical documented in Section 2. I discipline the key primitives—technology parameters, bargaining

Table 5: Static model parameters and active upstream–downstream links (Mid–2024)

Panel A: Static model parameters and baseline calibration

Symbol	Name	Value	Dim.	Notes
σ	Downstream demand curvature	6.0	scalar	Cournot block curvature
ρ	Composite aggregator elasticity	4.0	scalar	Substitution across upstream bundles
η	Within–bundle substitution	2.5	scalar	Substitution across active links
D	# downstream firms	2	scalar	{NVIDIA, AMD}
U	# upstream firms	3	scalar	{Samsung, SK hynix, Micron}
$\boldsymbol{\gamma}^\top$	Downstream bargaining weights	$\begin{bmatrix}0.5 & 0.5\end{bmatrix}^\top$	$D \times 1$	Symmetric baseline; heterogeneous in sweeps

Panel B: Active upstream–downstream links (baseline adjacency $\text{active}(d, u) = 1$**)**

Downstream d	Upstream u	Note
NVIDIA	SK hynix	Mid–2024: qualified supplier
NVIDIA	Micron	Mid–2024: qualified supplier
AMD	Samsung	Mid–2024: qualified supplier
AMD	Micron	Mid–2024: qualified supplier

weights, and firm productivities—using a combination of external evidence and the concentration and trade facts in Section 2.1–2.2. On the structural side, I take the small-n network as given (a triopoly of HBM suppliers and a handful of accelerator buyers), impose demand and substitution elasticities from the literature, and treat Nash-in-Nash bargaining as the mechanism that pins down upstream prices. On the quantitative side, I choose firm-specific productivities to match indicative HBM and accelerator market shares, and I set baseline bargaining weights $\boldsymbol{\gamma}$ consistent with NVIDIA's observed dominance in the downstream segment. The goal is not to fit all reduced-form moments in Section 2, but to obtain a transparent, semi-structured environment in which the magnitudes of wedges and the strength of bargaining feedbacks are anchored to the industry context.

Panel A of Table 5 reports the primitives for the static calibration. The elasticities (σ, ρ, η) encode three design choices, guided by nested-CES estimates in the trade and production-network literature. First, a moderately high downstream curvature $\sigma = 6$ sustains Cournot markups in the accelerator market without making prices mechanically insensitive to cost movements; this leaves room for incidence to come from the vertical side rather than only

from demand.[13] Second, $\rho = 4.0$ governs the price elasticity of demand for the aggregate downstream accelerator bundle with respect to its price index.[14] Third, $\eta = 2.5$ keeps links within a bundle far from perfect substitutes but not Leontief: once several HBM stacks are qualified, buyers can shift some volume across suppliers, yet tight process and packaging windows prevent one vendor from completely replacing another.[15]

Baseline bargaining weights $\boldsymbol{\gamma}$ are set symmetrically at 0.5 for both buyers. This is deliberately conservative. In practice, NVIDIA likely has more bargaining leverage than AMD on qualified links, but imposing stark asymmetry in the baseline would mechanically exaggerate cross-buyer heterogeneity in incidence. The symmetric specification instead isolates the role of firm size and network position—driven by $\boldsymbol{z}_d$ and the active adjacency—from the role of heterogeneous bargaining leverage in shaping equilibrium incidence. I use this symmetric case as the reference point and then sweep $\boldsymbol{\gamma}$ in Figure 9 to trace how redistributing bargaining surplus between buyers and suppliers reallocates the burden between upstream repricing and downstream markup adjustment. In Nash-in-Nash, incidence is pinned down by relative outside options and scale; $\boldsymbol{\gamma}$ is therefore a transparent knob for tracing how buyer power reshapes the decomposition of price changes. Baseline frictions set $\boldsymbol{\tau}_d$ and $\boldsymbol{f}_{du}$ to one so that all movements in the counterfactuals are attributable to explicit shocks rather than hidden wedges.

Productivities $\boldsymbol{z}_d$ and $\boldsymbol{z}_u$ are jointly calibrated to match observed market shares via a two-stage Levenberg–Marquardt procedure. On the downstream side, I target the NVIDIA/AMD

[13] Empirical work on differentiated products typically estimates residual demand elasticities for individual firms that imply moderate markups rather than either near-perfect competition or extreme market power. For example, Hottman et al. (2016) quantifies firm heterogeneity in U.S. retail sectors using a structural demand system that delivers effective elasticities in the mid–single to low–double digits, while Feenstra and Weinstein (2017) back out similar ranges from price and expenditure data.

[14] In trade and macro models with nested CES preferences, analogous upper-tier demand elasticities are typically calibrated in the low single digits to match the sensitivity of sectoral quantities, and trade flows to relative prices; see, for example, Atkeson and Burstein (2008) and the discussion in Feenstra and Weinstein (2017).

[15] A growing literature emphasizes that substitution across intermediate inputs is limited relative to substitution across final outputs. Using the 2011 Tōhoku earthquake as a natural experiment, Boehm et al. (2019) shows that firms cannot easily reoptimize their input mix when particular suppliers are disrupted, so that shocks propagate powerfully along input–output links. At a more aggregate level, Oberfield and Raval (2021) estimates input substitution elasticities that are close to (and in many specifications below) one.

accelerator market-share ratio by choosing $\boldsymbol{z}_d$ (with AMD normalized to one). On the upstream side, I calibrate $\boldsymbol{z}_u$ to match the Counterpoint Research Q2 2025 HBM shares in Table 2: Samsung 17%, SK hynix 62%, and Micron 21%.[16] Micron's productivity serves as the normalization reference; the free parameters z_{Samsung} and $z_{\text{SK hynix}}$ are identified from the log revenue-share ratios relative to Micron. The two stages (updating $\boldsymbol{z}_d$ and $\boldsymbol{z}_u$ in turn) are iterated three times to account for cross-tier interdependence in equilibrium prices and shares.

Conditional on $\boldsymbol{z}_u$ from the upstream stage, I normalize AMD's productivity to one and choose the remaining downstream productivity in log space to match the accelerator share ratio. Let (s_d) denote the target downstream share vector and $\hat{s}_d(\boldsymbol{z}_d, \boldsymbol{z}_u)$ the model-implied shares. With $D = 2$, the calibration solves a one-dimensional nonlinear least-squares problem over

$$\theta = \log z_{\text{NVIDIA}}$$

to minimize the squared distance between target and model log share ratios relative to AMD,

$$\min_{\theta} \left[\log \frac{\hat{s}_{\text{NVIDIA}}(\theta)}{\hat{s}_{\text{AMD}}(\theta)} - \log \frac{s_{\text{NVIDIA}}}{s_{\text{AMD}}} \right]^2.$$

Each evaluation of the objective calls the equilibrium solver, and I update θ using a damped Gauss–Newton / Levenberg–Marquardt step with backtracking. In the quantitative exercises below, these productivity levels serve purely to pin down relative size and the associated bargaining shadows (the equilibrium sensitivities of upstream prices to buyer scale, captured by the off-diagonal elements of $-J^{-1}$) that condition upstream pricing responses; the main objects of interest are the comparative statics of Nash-in-Nash input prices and downstream markups when buyer-specific wedges change.

The aim of this calibration is not to fit levels with high precision but to make the model

[16] Q2 2025 is chosen because it corresponds to the period when Samsung had not yet qualified for NVIDIA's HBM3E program, directly validating the sparse baseline adjacency matrix (Samsung→AMD only). This framing also provides an empirical anchor for the non-neutrality exercise: Samsung's subsequent qualification in late 2025 is precisely the link-addition event studied in Section 4.

informative about the composition non-neutrality of trade shocks. A uniform output-side wedge—implemented as buyer-specific increases in τ_d—does more than raise prices one-for-one: it shifts the mixture between (i) upstream input bundle prices P_d^M and (ii) downstream markups $\mu_d = p_{D,d}/C_d^D$, and this mixture depends on who is linked to whom and on their bargaining weight configuration $\boldsymbol{\gamma}$. In the quantitative results below, I show that the same class of trade restrictions yields different decompositions for NVIDIA and AMD, and that activating the Samsung↔NVIDIA link reassigns percentage points of incidence from downstream markups to upstream prices (and vice versa). This is the mechanism I seek to highlight: identical policy wedges at the statutory margin can be compositionally non-neutral once vertical structure and extensive margins are taken seriously.

Finally, the chosen parameter values are conservative with respect to the main claim. Within reasonable ranges around $(\sigma, \rho, \eta) = (6, 4, 2.5)$ and alternative choices of $\boldsymbol{\gamma}$, the magnitude of the effects changes, but the qualitative pattern is robust: trade frictions reallocate burden across the chain and across buyers, and these reallocations are strongly mediated by vertical links and bargaining power.

4.2 Model Fit and Moment Discipline

The calibration targets and non-targeted moments are summarized in Tables 5 and 6. This subsection clarifies what the model is and is not asked to match, and confirms that the calibrated environment is internally consistent.

Targeted moments. The structural elasticities $(\sigma, \rho, \eta) = (6, 4, 2.5)$ are set from the external literature. The symmetric bargaining weights $\boldsymbol{\gamma} = (0.5, 0.5)^\top$ are imposed as a disciplined baseline, with heterogeneous sweeps explored in Figure 9b and Table 7. Upstream firm productivities $\boldsymbol{z}_u$ are jointly calibrated to match Counterpoint Research (Q2 2025) HBM market shares — SK hynix (62%), Samsung (17%), and Micron (21%) — as the primary upstream concentration moment, and NVIDIA's $\geq 80\%$ downstream share as the primary

downstream concentration moment (Table 6).

Targeted moments — upstream fit. As reported in Table 6, the jointly calibrated model matches all three Counterpoint Research Q2 2025 HBM shares to within numerical tolerance: Samsung 17%, SK hynix 62%, and Micron 21%. The Q2 2025 vintage is chosen precisely because it reflects the supply structure that the sparse adjacency matrix captures: Samsung was supplying AMD but had not yet cleared NVIDIA's HBM3E qualification program.[17] Using this snapshot aligns the calibration target with the institutional assumption, rather than asking the model to simultaneously explain Samsung's overall-market share (which would include Q3–Q4 qualification-driven recovery) with a baseline network structure that excludes the Samsung–NVIDIA link. As a supplementary quantity check, the model-implied output ratio of NVIDIA to AMD is approximately 4:1, consistent with publicly reported data-center GPU revenue splits ($\approx$ 80% NVIDIA vs. $\approx$ 20% AMD; see the note to Table 2).[18]

Internal consistency. The Nash-in-Nash fixed-point algorithm converges with a final residual below 10^{-8} (see Appendix I), confirming analytical uniqueness. All equilibrium solutions satisfy individual rationality: each firm's payoff weakly exceeds its disagreement payoff on every active link. The calibrated bargaining weight $\gamma = 0.5$ satisfies the ceiling condition $\gamma < \overline{\gamma} \approx 0.967$ in Assumption 2, so Proposition 1 applies analytically.

4.3 Counterfactual Analysis: The Chip War

Given this calibrated environment, I next use the model to study how buyer-specific export wedges propagate through a concentrated HBM–accelerator vertical. Throughout, I compare two heterogeneous shock patterns: **Scenario A** ($\%\Delta\tau_{\text{NV}} = +10\%$, $\%\Delta\tau_{\text{AMD}} = +5\%$),

[17] See the footnote on page 35 for the Reuters sourcing on Samsung's qualification timeline.

[18] Revenue split sources are reported in the note to Table 2. The model is calibrated to price-side moments (unit values and market shares), not absolute shipment volumes, which are commercially sensitive and subject to measurement error from the customs proxy.

Table 6: HBM vendor market shares: calibration targets vs. model fit

Upstream supplier u	Target s_u^{data} (%)	Model $\hat{s}_u^{\text{model}}$ (%)	Notes
Samsung	17.0	17.0	Samsung→AMD only (sparse adj.)
SK hynix	62.0	62.0	
Micron	21.0	21.0	Normalization reference

Notes: Target shares are Counterpoint Research Q2 2025 HBM revenue shares (also reported in Table 2). Model shares $\hat{s}_u^{\text{model}}$ are equilibrium HBM revenue shares from the jointly calibrated static bilateral oligopoly; upstream productivities $\boldsymbol{z}_u$ are identified via Levenberg–Marquardt to match all three share moments (Section 4.2). All three shares are targeted and matched to within numerical tolerance ($< 10^{-5}$). The sparse baseline adjacency (Samsung→AMD only, not NVIDIA) is maintained throughout; the feasibility of a 17% Samsung share with only 15% AMD downstream share is accommodated by Samsung's relatively high equilibrium productivity on the AMD link.

where NVIDIA bears the heavier exposure; and **Scenario B** ($\%\Delta\tau_{\text{NV}} = +5\%$, $\%\Delta\tau_{\text{AMD}} = +10\%$), where AMD bears the heavier exposure. The experiments are organized around three questions that speak directly to the empirical patterns in Section 2: (i) how heterogeneous wedges across buyers map into upstream repricing and downstream markups, and whether they can generate the opposite-signed responses across regimes seen in the event studies; (ii) how changes in vertical composition—most starkly, adding or removing a Samsung↔NVIDIA link—reallocate the incidence of a given policy shock between upstream input prices and downstream markups; and (iii) how shifts in bargaining leverage or in a single buyer's wedge compress or amplify cross-buyer differences in effective HBM costs. Throughout, I report outcomes at the level and in the decomposition between P_d^M and downstream markups. I interpret the numerical results as a quantitative counterpart to the Taiwan×MCP–IC evidence: they show when and how export controls on a dominant buyer can raise rival access to bottleneck inputs while tightening the bottleneck along the treated buyer's own links.

Table 6 confirms that the jointly calibrated model reproduces all three Q2 2025 HBM shares exactly. One feasibility note is worth recording: with Samsung exclusively supplying AMD in the baseline (AMD downstream share 15%), Samsung's maximum achievable HBM

revenue share is bounded by AMD's total HBM spending as a fraction of the market. That AMD's HBM intensity in the model is sufficient to accommodate a 17% Samsung share is a consequence of the CES input-bundle structure and the equilibrium price configuration; it is not guaranteed a priori. The calibration confirms feasibility. The non-neutrality experiments in Figures 7 and 16 study what happens when the Samsung–NVIDIA link is activated (the Q4 2025 qualification event), confirming that the qualitative sign pattern of incidence is robust to activating this margin.

Figure 6 summarizes how heterogeneous accelerator wedges map into downstream prices through the statutory component, input prices, and markups. In both configurations, the dominant contribution to $\%\Delta p_d$ comes mechanically from the imposed trade wedges: a 10 percent increase in NVIDIA's delivery wedge and a 5 percent increase for AMD (panel (a)), or the reverse pattern with NVIDIA at 5 percent and AMD at 10 percent (panel (b)). Beyond this mechanical component, two endogenous channels respond: markups adjust as firms reoptimize quantities, and the input price incidence channel shifts P_d^M as HBM suppliers reprice across different links. In the calibrated environment, the input-price channel moves in opposite directions for the two firms: $\%\Delta P^M_{\text{NV}} \approx -0.16\%$ (shock-absorber) and $\%\Delta P^M_{\text{AMD}} \approx +0.22\%$ (cross-buyer spillover). The markup adjustments have the same signs but are substantially larger in magnitude: $\%\Delta\mu_{\text{NV}} \approx -0.82\%$ and $\%\Delta\mu_{\text{AMD}} \approx +0.22\%$. The input-price channel is therefore roughly one-fifth the size of the markup response for NVIDIA at the baseline calibration ($|\%\Delta P^M_{\text{NV}}|/|\%\Delta\mu_{\text{NV}}| \approx 0.20$); across the active portion of the (σ, η) sensitivity grid ($\sigma \geq 6$, $\eta \geq 2.5$), this ratio ranges from 0.17 to 0.26, spanning roughly one-fifth to roughly one-quarter (Table 9). Under Scenario B, the analogous ratio for AMD is approximately 0.27, reflecting AMD's smaller baseline market share and more concentrated supplier exposure. While both channels are quantitatively nontrivial, the statutory wedges dominate the level of $\%\Delta p_d$. Whereas the secular AI boom ($d\ln A > 0$) drives equilibrium prices up — consistent with Figure 3b — the input-price incidence of the export control ($d\ln P^M/d\ln\tau < 0$) acts as a deflationary counterforce, dampening the total price spike.

To disentangle the upstream repricing channel from the mechanical pass-through of the statutory wedge, each panel of Figure 6 overlays a *frozen-input* ("downstream-only") benchmark: upstream link prices are held fixed at their pre-shock values while downstream Cournot quantities re-optimize in response to the changed wedge alone (see Appendix D for the algorithm). The frozen-input benchmark represents what a conventional competitive-input trade model would predict. The gap between the full-Nash-in-Nash bars (dark) and the frozen-input bars (light) therefore isolates the contribution of endogenous HBM repricing. In the calibrated scenario (a), NVIDIA's link-specific input price *falls* in full equilibrium — the shock-absorber mechanism of Proposition 2(a) — while AMD's materials price index *rises* via the cross-buyer spillover formalized in Corollary 1. Neither effect appears in the frozen benchmark, confirming that both the deflationary channel for the targeted buyer and the inflationary spillover to its rival are driven entirely by the endogenous upstream price renegotiation.

The zoomed panels on the right isolate these second-order channels.[19] When NVIDIA is more heavily exposed than AMD (panel (a)), NVIDIA's materials price index *falls* while AMD's *rises*: $\%\Delta P^M_{\text{NV}} < 0$ and $\%\Delta P^M_{\text{AMD}} > 0$. The frozen-input benchmark sets these terms to essentially zero by construction, so the gap between the dark (full NiN) and light (frozen) bars isolates the contribution of endogenous input prices. The full-Nash-in-Nash markup responses are $\%\Delta\mu_{\text{NV}} \approx -0.82\%$ and $\%\Delta\mu_{\text{AMD}} \approx +0.22\%$—firms internalize part of the shock through changes in μ_d—while the input-price channel ($\%\Delta P^M_{\text{NV}} \approx -0.16\%$, $\%\Delta P^M_{\text{AMD}} \approx +0.22\%$) reinforces the cross-buyer divergence in effective accelerator prices. These patterns highlight that focusing purely on markups while holding input costs fixed can miss an important part of the price effects: even modest movements in P^M_d are of the same order as markup adjustments and systematically push in the direction of compressing cross-buyer differences in effective accelerator prices.

For each heterogeneous shock pattern, Figure 7 compares the full equilibrium response of

[19] Each subfigure in Figure 6 shows a full decomposition bar chart on the left alongside a magnified view of $\%\Delta P^M_d$ and $\%\Delta\mu_d$ on the right.

$\%\Delta P_d^M$ and $\%\Delta\mu_d$ with and without a Samsung–NVIDIA link. It highlights that the composition of buyer wedges interacts with the extensive margin in a non-neutral way. When NVIDIA faces the larger wedge (panel (a)), the presence of an extra qualified supplier dampens the increase in NVIDIA's markup by about 46%. Its materials price index adjustment is almost gone with the new link with Samsung. The new link also attenuates AMD's markup decline by a similar order of magnitude, while AMD's materials price index shows only a negligible change. In other words, additional upstream competition partially insulates the exposed HBM suppliers from downstream shocks.[20]

When AMD is more heavily exposed than NVIDIA (panel (b)), the sign pattern reverses. The link now amplifies AMD's input price and markup effects while muting NVIDIA's. Comparing panels (a) and (b) shows that the same change in extensive margin—activating the Samsung–NVIDIA link—does not have a symmetric or linear effect. Its impact depends on which buyer is hit hardest by the policy shock. Hence, the decomposition of downstream price changes into wedges, input prices, and markups depends inherently on the cross-sectional distribution of τ_d.

Figure 8 documents how the magnitude of NVIDIA's wedge shapes the downstream incidence. NVIDIA's own accelerator price responds almost linearly to the shock, with a passthrough slightly below 1 due to the input-cost and markup adjustments, while the induced changes in the materials price index increase with the shock's magnitude across the range. AMD's input costs and prices increase gradually as NVIDIA's wedge widens, reflecting the spillover from NVIDIA's tighter access conditions onto AMD through the shared upstream suppliers. The figure reinforces that, while the statutory component dominates the level of $\%\Delta p_d$, movements in P_d^M accumulate monotonically as the wedge grows and are

[20]These non-neutrality patterns partly reflect that the Samsung–NVIDIA link lowers NVIDIA's baseline materials cost when productivity is held fixed across the "link" and "no-link" economies. With a lower pre-shock cost base, the same statutory wedge results in a smaller percentage change in NVIDIA's markup, even though its materials price index barely moves. To avoid this apples-to-oranges comparison, Appendix K recalibrates downstream productivities in the economy with the link so that pre-shock market shares match the benchmark and then repeats the non-neutrality counterfactuals, confirming that the qualitative patterns are robust to this recalibration.

an integral part of the total incidence.

Figure 9 turns to the role of bargaining leverage. Panel 9a plots the local derivative $\partial \log p^{U}_{\mathrm{NV},u} / \partial \log \tau_{\mathrm{NV}}$ for NVIDIA's contracts with SK hynix and Micron as γ_{NV} varies from 5 to 95 percent. For moderate values of γ_{NV} around the calibrated baseline, the derivative is negative for both suppliers: as NVIDIA becomes more exposed, its outside option value falls (its volume shrinks). However, because the supplier also loses a substantial volume, it has a strong incentive to retain the remaining business, which induces it to lower prices. When NVIDIA is assigned a very high bargaining weight, these derivatives approach zero or become slightly positive, indicating that the sign of the upstream price incidence depends on the downstream bargaining leverage.

Panel 9b translates these derivatives into changes in each buyer's materials price index under a fixed NVIDIA wedge shock of $\%\Delta\tau_{\mathrm{NV}} = 5\%$. Across the empirically plausible range of γ_{NV}, NVIDIA's P^{M}_{d} falls modestly, while AMD's rises, with the gap shrinking as γ_{NV} increases.

Remark 5 (Sign reversal at high buyer power)**.** *The shock-absorber mechanism has a boundary. Panel 9a shows the derivative* $\partial \log p^{U*}_{NV,u} / \partial \log \tau_{NV}$ *turning from negative (deflationary) to positive (inflationary) at approximately* $\gamma_{NV} \approx 0.85$ *at the baseline elasticities. The economic mechanism is as follows. At moderate buyer power, the supplier's volume motive dominates: when NVIDIA's constraint tightens and demand contracts, the supplier lowers input prices to retain the relationship, generating the shock-absorber effect. At very high buyer power, NVIDIA already extracts near-full surplus from each contract; the marginal loss in link surplus from a volume decline is small relative to the supplier's margin cost of a further price cut. In this regime, the supplier may instead raise prices to protect margin, amplifying rather than dampening the regulatory shock. The policy implication is direct: if NVIDIA's actual bargaining weight exceeds the* $\approx 85\%$ *threshold — plausible given its* $\geq 80\%$ *accelerator market share — the shock-absorber may be inoperative or sign-reversed. The sensitivity exercises in Tables 7 and 8 span* $\gamma_{NV} \in \{0.30, 0.50, 0.70\}$*, all strictly below*

the reversal threshold, and confirm the deflationary sign throughout this range. Disciplining γ_{NV} from direct contract or margin data is an important direction for future empirical work.

Together, the two panels show that the input-price incidence channel is robust over the range $\gamma_{\text{NV}} \leq 0.70$: the direction of incidence is stable, and the magnitude scales smoothly with γ_{NV} rather than hinging on knife-edge parameter choices. The reversal documented in Remark 5 defines the boundary of this robustness region.

Table 7 complements the figure with a numerical decomposition across a discrete grid of $\gamma_{\text{NV}} \in \{0.30, 0.50, 0.70\}$ and two asymmetric shock scenarios. Under Scenario A (NVIDIA bearing the heavier exposure), the upstream incidence ratio $\%\Delta P^M_{\text{NV}}/\%\Delta\mu_{\text{NV}}$ is positive and strictly below one for all three bargaining weights, confirming that the shock-absorber result is not a knife-edge artifact of the baseline γ. Under Scenario B (AMD bearing the heavier exposure), the roles reverse: AMD's input-price index rises as rival demand surges, while NVIDIA's ratio correspondingly increases, reflecting the cross-buyer spillover.

Table 7: Sensitivity of upstream incidence ratio to bargaining weight γ_{NV}

	Scenario A: NV +10%, AMD +5%		Scenario B: NV +5%, AMD +10%	
γ_{NV}	$\%\Delta P^M_{\text{NV}}$ / $\%\Delta\mu_{\text{NV}}$	$\%\Delta P^M_{\text{AMD}}$ / $\%\Delta\mu_{\text{AMD}}$	$\%\Delta P^M_{\text{NV}}$ / $\%\Delta\mu_{\text{NV}}$	$\%\Delta P^M_{\text{AMD}}$ / $\%\Delta\mu_{\text{AMD}}$
0.30	−0.22% / −0.85%	0.23% / 0.24%	0.20% / 0.24%	−0.22% / −0.81%
0.50	−0.16% / −0.82%	0.22% / 0.22%	0.15% / 0.21%	−0.21% / −0.78%
0.70	−0.11% / −0.79%	0.21% / 0.18%	0.10% / 0.18%	−0.19% / −0.75%

Notes: Each cell reports the percentage change in the materials price index P^M_d and the Cournot markup $\mu_d = p^D_d/C^D_d$ relative to the baseline at the corresponding γ_{NV}. The ratio $\%\Delta P^M_d/\%\Delta\mu_d$ measures the upstream incidence as a fraction of the downstream markup response. The baseline calibration uses $\gamma_{\text{NV}} = \gamma_{\text{AMD}} = 0.5$. Changes in γ_{AMD} are held at the baseline value of 0.5 throughout.

The symmetric baseline $\gamma_{\text{NV}} = \gamma_{\text{AMD}} = 0.5$ is disciplined but potentially implausible: given NVIDIA's dominant downstream share, one might expect it to command a stronger bargaining position vis-à-vis its HBM suppliers. Table 8 and Figure 10 assess the robustness of this baseline by comparing it to a heterogeneous calibration with $\gamma_{\text{NV}} = 0.7$ and $\gamma_{\text{AMD}} = 0.3$, with productivities re-calibrated to preserve the same pre-shock market shares. The qualitative sign pattern is unchanged: NVIDIA's materials price index still falls

in full equilibrium (shock absorber) while AMD's rises (cross-buyer spillover). Quantitatively, higher NVIDIA bargaining power modestly attenuates NVIDIA's own input-price drop ($\%\Delta P^M_{\text{NV}} = -0.107\%$ versus -0.163% in the symmetric case) and amplifies AMD's spillover ($\%\Delta P^M_{\text{AMD}} = +0.311\%$ versus $+0.221\%$). The markup responses are nearly identical across configurations, confirming that the main mechanism is robust to plausible heterogeneity in bilateral bargaining leverage.

Table 8: Price decomposition under symmetric vs. heterogeneous bargaining weights (Scenario A: $\tau_{\text{NV}} + 10\%$, $\tau_{\text{AMD}} + 5\%$)

Configuration	$\%\Delta P^M_{\text{NV}}$	$\%\Delta \mu_{\text{NV}}$	$\%\Delta P^M_{\text{AMD}}$	$\%\Delta \mu_{\text{AMD}}$
Symmetric ($\gamma_{\text{NV}} = \gamma_{\text{AMD}} = 0.5$)	-0.163%	-0.824%	$+0.221\%$	$+0.215\%$
Heterogeneous ($\gamma_{\text{NV}} = 0.7$, $\gamma_{\text{AMD}} = 0.3$)	-0.107%	-0.799%	$+0.311\%$	$+0.190\%$

Notes: Each row reports the percentage change in the materials price index P^M_d and the Cournot markup μ_d for NVIDIA and AMD under Scenario A ($\tau_{\text{NV}} + 10\%$, $\tau_{\text{AMD}} + 5\%$). Productivities are re-calibrated separately for each $\boldsymbol{\gamma}$ configuration to maintain the same pre-shock downstream market shares. Signs and qualitative patterns are unchanged across configurations.

Table 9 reports $\%\Delta P^M$ and $\%\Delta \mu$ for NVIDIA and AMD across a 3×3 grid of (σ, η) combinations, holding $\rho = 4.0$ fixed and re-calibrating productivities at each grid point to maintain baseline market shares. Two findings stand out. First, the upstream repricing channel is essentially inactive at $\sigma = 4$: both the shock-absorber ($\%\Delta P^M_{\text{NV}} < 0$) and the cross-buyer spillover ($\%\Delta P^M_{\text{AMD}} > 0$) are near zero regardless of η, reflecting that low Cournot curvature gives suppliers little incentive to cut prices when demand contracts. The channel activates at $\sigma = 6$ and strengthens at $\sigma = 8$. Second, conditional on $\sigma \geq 6$, the within-bundle substitution elasticity η has a modest effect on magnitudes but leaves the qualitative sign pattern intact across all nine grid points. To provide an explicit uncertainty range: among the six active grid points ($\sigma \in \{6, 8\}$, $\eta \in \{2.5, 4.0\}$), the incidence ratio $|\%\Delta P^M_{\text{NV}}|/|\%\Delta \mu_{\text{NV}}|$ takes values $\{0.17, 0.20, 0.23, 0.26\}$, spanning a range of $[0.17, 0.26]$. The baseline value of 0.20 is therefore near the lower end of this range, implying that the "one-fifth" characterization is a conservative estimate of the shock-absorber's relative importance; under plausible higher values of σ or η, the upstream repricing channel accounts for up to roughly one-quarter of

the markup response.[21]

Table 9: Sensitivity of upstream incidence ratio to structural elasticities

	Panel A: NVIDIA ($\%\Delta P^M_{\text{NV}}$ / $\%\Delta\mu_{\text{NV}}$)			Panel B: AMD ($\%\Delta P^M_{\text{AMD}}$ / $\%\Delta\mu_{\text{AMD}}$)		
η σ	$\sigma = 4$	$\sigma = 6$	$\sigma = 8$	$\sigma = 4$	$\sigma = 6$	$\sigma = 8$
1.5	+0.000% / −0.469%	−0.132% / −0.749%	−0.081% / −0.855%	+0.000% / −0.121%	+0.340% / +0.140%	+0.531% / +0.222%
2.5*	+0.000% / −0.469%	−0.163% / −0.824%	−0.272% / −1.053%	+0.000% / −0.121%	+0.221% / +0.215%	+0.339% / +0.417%
4.0	−0.000% / −0.469%	−0.138% / −0.832%	−0.247% / −1.063%	+0.000% / −0.121%	+0.199% / +0.223%	+0.338% / +0.426%

Notes: Each cell reports the percentage change in the materials price index P^M_d and the Cournot markup μ_d for NVIDIA (Panel A) and AMD (Panel B) under Scenario A ($\tau_{\text{NV}}+10\%$, $\tau_{\text{AMD}}+5\%$). $\rho = 4.0$ is held fixed throughout. Downstream productivities are re-calibrated at each (σ, η) grid point to maintain pre-shock market shares $s_{\text{NV}} = 0.85$, $s_{\text{AMD}} = 0.15$. **Bold** entries mark the baseline calibration ($\sigma = 6$, $\eta = 2.5$, * row).

Together, the three sets of exercises — heterogeneous wedge decompositions, network non-neutrality counterfactuals, and bargaining-weight sensitivity sweeps — paint a consistent picture: the statutory incidence of export controls is systematically dampened by endogenous upstream repricing, and the magnitude and direction of this dampening depend critically on vertical structure and bargaining leverage. Section 5 draws the policy implications.

5 Conclusion

This paper studies how export controls propagate upstream through a highly concentrated vertical relationship between AI accelerator producers and HBM suppliers. Using Korean MCP–IC export unit values as a proxy for HBM and exploiting the staggered tightening of U.S. export controls on AI accelerators, I document that shocks targeted at a single dominant buyer generate complex reallocations of surplus between tiers. In particular, the event-study evidence shows that policy restrictions induce a sharp divergence in trade patterns across destinations, consistent with an input-market incidence channel operating through concentrated upstream bargaining.

[21] At the grid point $\sigma = 8$, $\rho = 4$, the regularity condition $\sigma \leq 2\rho$ holds with equality ($8 \leq 8$). The Gershgorin proof of the P-matrix property (Appendix E) uses the weak inequality $\sigma \leq 2\rho$, so the uniqueness result and the M-matrix sign structure apply at equality. Numerically, the Nash-in-Nash algorithm produces well-defined equilibria at this grid point, confirming that the analytical boundary case does not introduce practical instability.

To interpret these patterns, I develop a static model of bilateral oligopoly with Nash-in-Nash bargaining over link-specific HBM prices and Cournot competition in the accelerator market. The model delivers a tractable characterization of "Chip War incidence": the sign of the upstream price response depends on whether the supplier's volume motive dominates the local demand curvature. In the calibrated equilibrium — where the log-submodularity condition of Remark 3 is satisfied at the baseline calibration — upstream suppliers effectively act as shock absorbers: they lower input prices for the targeted buyer to preserve sales volume, thereby cushioning the regulatory blow. Comparative statics show that these link-level responses aggregate into non-neutral changes in the materials price index and markups — a composition non-neutrality — and that the extensive margin—which HBM–accelerator links are active—critically shapes the pattern of gains and losses across buyers.

Quantitatively, I calibrate the model to mid-2020s market shares in the HBM and AI accelerator markets and to a sparse baseline adjacency matrix that omits the Samsung–NVIDIA link. The calibrated environment reproduces the high upstream concentration and generates realistic degrees of pass-through. The decomposition exercises reveal that the endogenous input-price response is deflationary: the upstream price cut dampens the statutory shock, offsetting roughly one-fifth of the downstream markup response for the dominant buyer (NVIDIA) under Scenario A at the baseline calibration. Across the (σ, η) sensitivity grid, this ratio ranges from 0.17 to 0.26 among the active grid points, spanning roughly one-fifth to roughly one-quarter (Table 9). Table 7 confirms the ratio is stable across $\gamma_{\text{NV}} \in \{0.30, 0.50, 0.70\}$; the analogous ratio for AMD under Scenario B is approximately 0.27, reflecting its smaller initial share. Moreover, adding a single Samsung↔NVIDIA link reshapes incidence in a strongly non-neutral way: the extra upstream option further amplifies this dampening effect, insulating the targeted buyer from the full force of the export control (Figure 7).

These findings carry several policy implications. First, evaluating export controls solely through the lens of downstream prices leads to a significant overestimation of their efficacy.

The shock-absorber mechanism is most quantitatively important when the targeted buyer is large (high bargaining weight γ_d) and faces few qualified upstream alternatives — precisely the conditions that describe NVIDIA's relationship with SK hynix in the HBM market. Two conditions must hold jointly for this conclusion to apply. The log-submodularity condition (Remark 3) must be satisfied at the calibrated baseline — it is, but it need not hold universally. And buyer bargaining power must remain below the sign-reversal threshold (Remark 5, approximately $\gamma \approx 0.85$ at baseline elasticities): above this threshold, the shock-absorber inverts, and export controls instead raise input prices for the targeted buyer, amplifying rather than dampening the statutory incidence. In markets where both conditions hold, policymakers should expect a material wedge between statutory pass-through and effective incidence.

Second, the results imply that conventional monitoring of export controls — tracking whether the targeted firm's revenues or chip shipments fall — is insufficient. The upstream repricing channel operates through confidential contract prices between HBM suppliers and accelerator designers; these prices are neither publicly reported nor captured in aggregate customs data. An accurate assessment of control efficacy therefore requires monitoring whether input costs for the targeted buyer have declined, not just whether its downstream output has fallen.

Third, the findings create a tension with the political economy of export controls. If upstream suppliers partially absorb the statutory shock by lowering input prices for the targeted buyer, then controls that appear stringent to outside observers may be substantially less burdensome from the buyer's perspective. This endogenous moderation could weaken the political sustainability of controls — firms facing lower effective costs have weaker incentives to lobby for relief — while simultaneously reducing their strategic bite. Understanding this feedback loop is important for the design of durable technology-access restrictions in concentrated global value chains.

The analysis is deliberately static and focuses on a single upstream bottleneck. Extending

the framework to incorporate dynamic investment, capacity, and entry decisions is a natural next step. Such extensions would allow one to quantify how repeated export controls interact with longer-run supply-chain reorganization — for instance, whether the shock-absorber mechanism weakens as buyers diversify their supplier base and qualified alternatives proliferate. Nonetheless, the static results already underscore a simple lesson: in concentrated high-tech supply chains, the impact of buyer-specific trade policy is blunted if one ignores the input market. Any assessment of modern export controls must therefore take seriously the interplay between vertical structure, bargaining power, and the endogenous repricing of bottleneck inputs.

References

Adachi, T., Ebina, T., 2014. Double marginalization and cost pass-through: Weyl–Fabinger and Cowan meet Spengler and Bresnahan–Reiss. Economics Letters 122, 170–175. doi:10.1016/j.econlet.2013.11.010.

Alessandria, G., Kaboski, J., Midrigan, V., 2013. Trade wedges, inventories, and international business cycles. Journal of Monetary Economics 60, 1–20.

Alessandria, G., Kaboski, J.P., Midrigan, V., 2010. Inventories, lumpy trade, and large devaluations. American Economic Review 100, 2304–2339.

Alessandria, G., Kaboski, J.P., Midrigan, V., 2011. Us trade and inventory dynamics. American Economic Review 101, 303–307.

Alfaro, L., Antràs, P., Chor, D., Conconi, P., 2019. Internalizing global value chains: A firm-level analysis. Journal of Political Economy 127, 509–559. doi:10.1086/700935.

Alviarez, V.I., Fioretti, M., Kikkawa, A.K., Morlacco, M., 2025. Two-sided market power in firm-to-firm trade. American Economic Review Forthcoming; NBER Working Paper No. 31253; arXiv:2507.12848.

Amiti, M., Itskhoki, O., Konings, J., 2014. Importers, exporters, and exchange rate disconnect. American Economic Review 104, 1942–78.

Antràs, P., Chor, D., 2013. Organizing the global value chain. Econometrica 81, 2127–2204. doi:10.3982/ECTA10813.

Antras, P., Foley, C.F., 2015. Poultry in motion: a study of international trade finance practices. Journal of Political Economy 123, 853–901.

Antras, P., Helpman, E., 2004. Global sourcing. Journal of political Economy 112, 552–580.

Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., Yeaple, S., 2018. Innovation and production in the global economy. American Economic Review 108, 2128–2173.

Atkeson, A., Burstein, A., 2008. Pricing-to-market, trade costs, and international relative prices. American Economic Review 98, 1998–2031.

Bagwell, K., Staiger, R.W., Yurukoglu, A., 2020. "nash-in-nash" tariff bargaining. Journal of International Economics 122, 103263.

Baqaee, D.R., Farhi, E., 2024. Networks, barriers, and trade. Econometrica 92, 505–541. doi:10.3982/ECTA17513.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 119, 249–275.

Boehm, C.E., Flaaen, A., Pandalai-Nayar, N., 2019. Input linkages and the transmission of shocks: Firm-level evidence from the 2011 tōhoku earthquake. Review of Economics and Statistics 101, 60–75.

Brander, J.A., Spencer, B.J., 1985. Export subsidies and international market share rivalry. Journal of International Economics 18, 83–100. doi:10.1016/0022-1996(85)90006-6.

Branstetter, L., 2024. Export Controls and U.S.–China Technology Competition. Working Paper. Brookings Institution.

Caliendo, L., Parro, F., 2015. Estimates of the trade and welfare effects of nafta. The Review of Economic Studies 82, 1–44.

Chipty, T., Snyder, C.M., 1999. The role of firm size in bilateral bargaining: A study of the cable television industry. Review of Economics and Statistics 81, 326–340. doi:10.1162/003465399558201.

Collard-Wexler, A., Gowrisankaran, G., Lee, R.S., 2019. "nash-in-nash" bargaining: a microfoundation for applied work. Journal of Political Economy 127, 163–195.

Conley, T.G., Taber, C.R., 2011. Inference with "difference in differences" with a small number of policy changes. Review of Economics and Statistics 93, 113–125.

Crawford, G.S., Yurukoglu, A., 2012. The welfare effects of bundling in multichannel television markets. American Economic Review 102, 643–85.

Crosignani, M., Han, L., Macchiavelli, M., Silva, A.F., 2025. Securing technological leadership? The cost of export controls on firms. Journal of Financial Economics Forthcoming. NBER Staff Report 1096.

Dhyne, E., Kikkawa, A.K., Kong, X., Mogstad, M., Tintelnot, F., 2023. Endogenous production networks with fixed costs. Journal of International Economics 145, 103841.

Eaton, J., Jinkins, D., Tybout, J.R., Xu, D., 2022a. Two-sided search in international markets. Technical Report. National Bureau of Economic Research.

Eaton, J., Kortum, S.S., Kramarz, F., 2022b. Firm-to-Firm Trade: Imports, exports, and the labor market. Technical Report. National Bureau of Economic Research.

Feenstra, R.C., Weinstein, D.E., 2017. Globalization, markups, and us welfare. Journal of Political Economy 125, 1040–1074.

de Fontenay, C.C., Gans, J.S., 2014. Bilateral bargaining with externalities. Journal of Industrial Economics 62, 756–788. doi:10.1111/joie.12058.

Gaubert, C., Itskhoki, O., 2021. Granular comparative advantage. Journal of Political Economy 129, 871–939.

Gaudin, G., 2016. Pass-through, vertical contracts, and bargains. Economics Letters 139, 1–4. doi:10.1016/j.econlet.2015.12.003.

Goldberg, P.K., Juhász, R., Lane, N.J., Forte, G.L., Thurk, J., 2024. Industrial policy in the global semiconductor sector. Technical Report. National Bureau of Economic Research.

Helpman, E., Melitz, M.J., Yeaple, S.R., 2004. Export versus fdi with heterogeneous firms. American economic review 94, 300–316.

Horn, H., Wolinsky, A., 1988. Bilateral monopolies and incentives for merger. The RAND Journal of Economics , 408–419.

Hottman, C.J., Redding, S.J., Weinstein, D.E., 2016. Quantifying the sources of firm heterogeneity. The Quarterly Journal of Economics 131, 1291–1364. doi:10.1093/qje/qjw020.

Igami, M., 2017. Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. Journal of Political Economy 125, 798–847.

Inderst, R., Wey, C., 2003. Bargaining, mergers, and technology choice in bilaterally oligopolistic industries. RAND Journal of Economics 34, 1–19. doi:10.2307/3087446.

Juárez, L., 2025. Buyer market power and exchange rate pass-through. IDB Working Paper Series No. IDB-WP-01662 Doi:10.18235/0013557.

Kim, H.J., Park, S., Shin, K., Yang, J.w., 2024. The impact of export controls on international trade: Evidence from the Japan–Korea trade dispute in the semicon-

ductor industry. Journal of the Japanese and International Economies 74, 101240. doi:10.1016/j.jjie.2024.101240.

Klette, T.J., Kortum, S., 2004. Innovating firms and aggregate innovation. Journal of political economy 112, 986–1018.

Kreps, D.M., Scheinkman, J.A., 1983. Quantity precommitment and bertrand competition yield cournot outcomes. The Bell Journal of Economics , 326–337.

Lamadon, T., Mogstad, M., Setzler, B., 2022. Imperfect competition, compensating differentials, and rent sharing in the us labor market. American Economic Review 112, 169–212.

Lim, K., 2018. Endogenous production networks and the business cycle. Work. Pap .

Novshek, W., 1985. On the existence of Cournot equilibrium. The Review of Economic Studies 52, 85–98.

Oberfield, E., 2018. A theory of input–output architecture. Econometrica 86, 559–589.

Oberfield, E., Raval, D., 2021. Micro data and macro technology. Econometrica 89, 703–732.

Omdia, 2024. HBM supply agreements and pricing dynamics in the AI accelerator market. Industry Research Report. Omdia (formerly IHS Markit Technology).

Ramondo, N., Rodríguez-Clare, A., 2013. Trade, multinational production, and the gains from openness. Journal of Political Economy 121, 273–322.

Rey, P., Vergé, T., 2004. Bilateral control with vertical contracts. RAND Journal of Economics 35, 728–746.

Rosen, J.B., 1965. Existence and uniqueness of equilibrium points for concave n-person games. Econometrica: Journal of the Econometric Society , 520–534.

Stole, L.A., Zwiebel, J., 1996. Intra-firm bargaining under non-binding contracts. Review of Economic Studies 63, 375–410. doi:10.2307/2297888.

Weyl, E.G., Fabinger, M., 2013. Pass-through as an economic tool: Principles of incidence under imperfect competition. Journal of Political Economy 121, 528–583. doi:10.1086/670401.

A Data Appendix

I use monthly Korean customs export data from the Korea Trade Statistics Promotion Institute (TRASS), organized by HS10 product code and partner country, from January 2010 to the latest available month. For each destination-item pair $i = (c, g)$ and month t, the data report export values (USD) and export quantities (kg).

Variable construction. For each cell, I compute unit values $\text{UV}_{it} = \text{USD}_{it}/\text{kg}_{it}$, dropping cells with zero quantity. I construct 12-month trailing moving averages for all value, quantity, and unit-value series. Item shares (e.g., MCP–IC's share of Korea's total exports) use the world total row (CountryCode = WR) as the denominator. Bilateral decompositions compare Taiwan to the rest of the world, where "Others" equals WR minus Taiwan.

Index normalization. For figures that show indexed unit values, each item's UV is normalized to its January 2019–July 2022 mean (pre-policy = 100), so that levels are comparable across items with very different USD/kg magnitudes (e.g., MCP–IC peaks above 200,000 USD/kg while commodity DRAM/Flash are an order of magnitude lower).

Policy and technology overlays. Vertical markers and shaded windows indicate: (i) HBM3/H100 production ramp (October 2021), (ii) the October 2022 U.S. export-control regime (Entity List expansion and FDP Rule), and (iii) the October 2023 comprehensive update. These dates are taken from publicly available regulatory announcements and are consistent with the technology timeline in Appendix C.1.

B Event Study: Regression Estimates

Table 10 summarizes the event study estimates in a compact difference-in-differences format. Each column estimates the average treatment effect of Treat × Post, where Treat is the Taiwan × MCP–IC indicator and Post equals one for the twelve months following each policy event ($k \geq 1$). All specifications absorb unit, destination-by-month, and item-by-

month fixed effects, and include the AI-demand control ($\text{AI}_t \times W_i$). Standard errors are clustered by unit. Because there is a single treated unit (Taiwan×MCP–IC), these clustered standard errors and the associated significance stars are not asymptotically valid for the treatment coefficient; Fisher permutation p-values from Appendix J (Table 13) supersede them for inference.

Table 10: Average post-event treatment effect: Taiwan × MCP–IC exports

	Oct 2022 Controls		Oct 2023 Controls	
	(1) ln(value)	(2) ln(qty)	(3) ln(value)	(4) ln(qty)
Treat × Post	-0.808***	-1.621***	1.157**	1.045***
	(0.282)	(0.206)	(0.530)	(0.367)
Observations	1,250	1,250	1,237	1,237
Within R^2	0.007	0.028	0.023	0.017

Notes: Unit, destination-by-month, and item-by-month FEs absorbed.
Standard errors in parentheses, clustered by unit (not asymptotically valid with a single treated unit).
See Table 10 for Fisher permutation p-values, which supersede these stars for inference.
Treat: Taiwan × MCP-IC. Post: months $k \geq 1$ within ±12 window.
All outcomes residualized for unit trend and unit-by-month seasonality.
AI control: demeaned log NVIDIA data-center revenue × unit exposure weight.

The October 2022 controls (columns 1–2) produced a sharp, significant contraction in both export value and quantity. The October 2023 update (columns 3–4) produced an equally significant expansion in both outcomes. The sign reversal between the two regimes is consistent with the compliance-chill and capacity-reallocation mechanisms discussed in the main text. The full dynamic path for each event and outcome is shown in Figure 5.

Table 11: Major memory and packaging technology upgrades, 2021–2025

Date	Segment	Generation / Node	Key spec	Noted improvement
Dec 2022	DRAM (DDR5)	Samsung 12nm-class 16Gb	Up to 7.2 Gb/s	**23%** better power efficiency vs prior gen (vendor)[a]
Jan 2023	DRAM (LPDDR)	SK hynix LPDDR5T	9.6 Gb/s/pin	**+13%** speed vs LPDDR5X 8.5 Gb/s (vendor).
Oct 2022	HBM	HBM3 (mass prod. reported)	∼6.4 Gb/s/pin; up to ∼819 GB/s per stack	HBM3 vs HBM2E: higher pin rate/bandwidth (context).
Mar 2024	HBM	Samsung HBM3E (mass prod.)	9.8 Gb/s/pin; up to 1.25 TB/s per stack	**+53%** per-pin rate vs HBM3 6.4 Gb/s (calc.).
Mar 2024	HBM	Micron HBM3E (volume)	up to ≥1.2 TB/s per stack	Power efficiency and bandwidth uplift vs HBM3 (vendor).
Jul 2022	NAND	Micron 232-layer (V8)	232L 3D NAND	Industry-first 232L; perf./power gains vs prior gen (vendor).
Aug 2022	NAND	SK hynix 238-layer	238L 3D NAND	"World's highest" layers at the time (vendor).
May 2024	NAND	Samsung V9 (236-layer)	236L 3D NAND	Read speed up to 2.4 Gb/s; **>10%** energy efficiency vs V8 (vendor).
Jan 2025	NAND	SK hynix 321-layer (ann.)	321L 3D NAND	Next-gen density milestone (vendor).

Notes: "calc." indicates a simple calculation using vendor-stated numbers (e.g., $9.8/6.4 - 1 \approx 53\%$). Specs and improvements are as reported by vendors unless otherwise stated.

[a] Samsung's 12nm-class DDR5 PR highlights both the up-to-7.2 Gb/s data rate and 23% power-efficiency improvement.

C Event-Study Robustness: Alternative Designs and Sample Restrictions

C.1 Technology upgrade timeline

Table 11 lists major memory and packaging upgrades between 2021 and 2025—DDR5 and LPDDR5T on the DRAM side, successive HBM3/HBM3E ramps, and steep layer-count progressions in 3D NAND. I include vendor-stated or directly implied percentage improvements in speed, bandwidth, efficiency, or density to give a sense of magnitudes. The purpose of the table is twofold. First, it documents technology shocks that could move prices or quantities independently of policy, clarifying what "parallel trends" should (and should not) look like within the control groups. Second, it provides precise markers overlaid in the event-study figures to visually distinguish policy windows (Aug–Oct 2022 notice; Oct 7 2022–Oct 2023 regime; Oct 2023 update) from contemporaneous technology milestones (e.g., HBM3E commercialization). Because MCP–IC is the customs line that records HBM-attached packages, its behavior is expected to be more tightly coupled to HBM3/HBM3E timing than DRAM

(commodity DDR5/LPDDR) or NAND; the table makes those relative exposures transparent.

This context helps interpret the event-study results in Section 2. The design contrasts Taiwan×MCP–IC ("treated") against two sets of comparisons: (i) the same item shipped to other destinations, and (ii) other memory items (DRAM, NAND) shipped to Taiwan. The timeline shows why DRAM and NAND are informative controls: they experienced meaningful—but differently timed—technology progress (e.g., DDR5 efficiency, NAND layer counts) that should influence prices and volumes, yet they lack the same dependence on advanced-packaging bottlenecks and HBM qualification that bind MCP–IC to the accelerator chain. Consequently, when sharp, Taiwan-directed MCP–IC premia and volume shifts appear concentrated around the policy windows (with only muted echoes in DRAM/NAND), the pattern is consistent with downstream policy frictions reallocating toward the TSMC+CoWoS+HBM hub and bidding up upstream packages. At the same time, by (a) placing technology markers on the x-axis, (b) residualizing outcomes for seasonality, unit trends, and an interacted AI-market proxy, and (c) reporting local windows separately from stacked windows, I mitigate the risk that technology waves alone drive the estimated treatment effects. Remaining caveats include the vendor-reported nature of some specs, potential lags between disclosure and effective line ramp, and the coarse mapping from HS codes to narrowly defined HBM products; these considerations motivate the robustness variants reported alongside the main figures.

C.2 Stacked event studies

To complement the single-event specifications, I pool the Oct 2022 and Oct 2023 policy dates in a stacked event study. For each event, I create a separate relative-time window of ±12 months and treat the same destination–item cell appearing in two windows as two distinct panels. Estimation includes unit-by-event, destination-by-month-by-event, and item-by-month-by-event fixed effects, with the outcomes already residualized for unit

trends, unit-specific seasonality, and the interacted AI-market proxy. This design increases precision while preserving the short-run, local comparison around each policy change.

A concern with using a single long window is that medium-run technology cycles and unrelated shocks can distort trends and contaminate pre- and post-contrasts. I address this by (i) centering on each regime change separately, so slow-moving trends do not cumulate across years; (ii) re-indexing panels by event to allow unit-level composition and FE structure to differ across the two policy regimes; and (iii) retaining the same strong time fixed effects (destination×month and item×month within event), which absorb global seasonality and commodity-wide movements. Together, the stacked specification asks whether, on average across both policy changes, the treated Taiwan×MCP–IC cell rises relative to controls in the immediate months after each change, conditional on rich FEs and the AI-proxy residualization.

The stacked plots show muted pre-trends in the 12-month leads (though the joint F-test rejects the null formally, as discussed in Section 2.3), followed by a monotonic post-policy rise in both quantity and export value. The quantity response initially rises, with effects persisting through the subsequent year; export values follow a similar upward trend, consistent with scarce packaging capacity and tighter upstream input bargaining, which pass through into higher prices as volumes re-route to the qualified Taiwan hub. Economically, the pattern lines up with the mechanism emphasized in the paper: downstream restrictions induce reallocation toward the HBM bottleneck, increasing both volumes and the effective unit values received in that channel.

Stacking averages heterogeneous treatment intensity across the two regimes; if the Oct 2023 update is stronger, pooled coefficients blend that with the earlier shift. Confidence bands remain wide at longer horizons due to the limited number of months per window. As with any difference-in-differences design, residual differential shocks correlated with the treatment could bias the estimates; the residualization with the AI-market proxy and the strong time fixed effects mitigate—but cannot eliminate—this risk. For these reasons, I

keep stacked results in the appendix as a precision-oriented robustness check; the main text focuses on the two local windows separately.

C.3 Event studies with restricted samples

To complement the full-sample designs in the main text, I estimate two restricted-sample event studies that make the construction of the control groups more transparent. In both designs, I work with residualized outcomes. For each outcome $y \in \{\ln(\text{export value}), \ln(\text{quantity})\}$, I first remove unit fixed effects and HS–specific month–of–year seasonality using the residualization step described in Section 2, and then use the residuals as the dependent variable in the event–time regressions.

Design A keeps only the MCP–IC item and compares Taiwan to other destinations. The unit of observation is a destination–month; treated units are Taiwan destinations, and donors are other destinations that import MCP–IC from Korea. Design B keeps only Taiwan and compares MCP–IC to DRAM/NAND within Taiwan. Here, the unit of observation is an HS–code–month; treated units are MCP–IC observations, and donors are DRAM and NAND exports to Taiwan.

For each design and each policy date $T_0 \in \{2022\text{m}10, 2023\text{m}10\}$, I restrict the sample to a symmetric ± 12–month window around T_0 and estimate a saturated TWFE event study,

$$y_{it} = \sum_{k=-12,\, k\neq -1}^{12} \beta_k \, \mathbf{1}\{t - T_0 = k\} \times \text{Treat}_i \; + \; \alpha_i + \mu_t \; + \; \delta_1 \, \text{AIxW}_{it} \; + \; \delta_2 \, t \; + \; \delta_3 \, t \times \text{Treat}_i \; + \; \varepsilon_{it}, \tag{17}$$

where i indexes units (destination–item cells in Design A and HS–item cells in Design B), α_i are unit fixed effects, μ_t are month fixed effects, and AIxW_{it} is the AI–exposure control described in Section 2. I omit $k = -1$ as the reference month, so all β_k are interpretable as deviations from the month immediately before the policy change. Robust standard errors are used throughout. In Design A, I also reweight control destinations using overlap weights constructed from pre-period means and linear trends of the residualized outcomes; these

weights improve balance in the pre-window while keeping the average weight equal to one. Design B keeps uniform weights because the number of items within Taiwan is small and already quite balanced.

Figure 12 reports the restricted–sample event studies for Design A. Following the October 2022 policy change, MCP–IC exports to Taiwan exhibit a sharp and persistent decline compared to MCP–IC exports to other destinations. Both quantities and export values fall by roughly 1.5–2.5 log points in the months following October 2022, with no strong pre-trend in the preceding twelve months. Around the October 2023 tightening, the pattern reverses: relative MCP–IC exports to Taiwan jump on impact and trend upward over the following year, reaching gains of about 1–2 log points in quantities and 3 or more log points in values by late 2024. This pattern is consistent with a reallocation of high-bandwidth memory shipments toward Taiwan once downstream accelerator exports are more tightly constrained.

Figure 13 shows the corresponding results for Design B, which uses only Taiwanese trade and compares MCP–IC to DRAM/NAND. The 2022 event results in a large and sustained decline in MCP–IC exports relative to the DRAM/NAND control group, again with modest pre-trends. The 2023 event produces a mirror image: MCP–IC exports rise sharply relative to DRAM/NAND after October 2023, with effects that grow over time in both quantity and value. Across both designs, the restricted-sample event studies confirm the main-sample findings: the 2022 controls depress Korea-Taiwan MCP–IC trade relative to nearby margins, while the 2023 tightening is associated with a pronounced expansion of MCP–IC exports to Taiwan relative to appropriately chosen controls.

C.4 Narrow pre-period robustness

The full-window pre-period F-tests reported in Section 2.3 reject parallel trends for both export value and quantity around both policy events. This is expected: the early pre-period months ($k \leq -8$) coincide with the HBM3 development ramp, which differentially elevated MCP–IC relative to commodity memory before any policy intervention (Appendix C.1). To

confirm that the average treatment effect (ATT) is not sensitive to this early pre-period divergence, Table 12 reports the post-event ATT — defined as the mean of $k = 0, \ldots, 12$ coefficients from the event-study regression — under two pre-period definitions: the full window ($k = -12$ to -2) and a narrow window ($k = -7$ to -2) that excludes the HBM3-ramp months.

The ATT estimates are virtually identical across the two specifications for all four outcome–event combinations: the largest absolute difference is 0.056 log points (Oct 2023 quantity), which is smaller than the standard error for either specification. The narrow pre-period F-statistic for export value in October 2022 is $F(6, \cdot) = 1.31$ ($p = 0.268$), confirming a flat pre-policy path once the HBM3 ramp months are excluded. This stability supports reading the full-window event-study figures as robust descriptive profiles of the Taiwan–MCP–IC corridor through the policy window.

Figure 14 plots the narrow-window ($k = -7$ to $+12$) event studies alongside the full-window ($k = -12$ to $+12$) baseline. The post-event dynamics are indistinguishable across specifications; the narrow-window pre-period confidence bands are tighter and closer to zero, consistent with the flat narrow-window F-statistics.

D Quantitative Protocols

This appendix details the numerical algorithms and shock protocols used to generate the quantitative results in Section 4. The procedures are implemented in MATLAB and rely on the fixed-point solver described in Appendix I.

D.1 Constructing the Frozen-Input Benchmark

To isolate the input-price incidence channel shown in Figure 6, I compare the full Nash-in-Nash equilibrium to a "frozen-input" benchmark where upstream prices are prevented from adjusting. This counterfactual shuts down the bargaining stage while allowing downstream

Table 12: Narrow vs. full pre-period ATT comparison

	Oct 2022		Oct 2023	
Pre-period	ln(value)	ln(qty)	ln(value)	ln(qty)
Full (k=−12 to −2)	-0.814	-0.804	1.582	1.696
	(0.602)	(0.568)	(0.323)	(0.390)
Narrow (k=−7 to −2)	-0.813	-0.814	1.597	1.752
	(0.604)	(0.579)	(0.327)	(0.417)
Pre-trend F (full)	4.63	18.86	6.62	5.60
	[0.000]	[0.000]	[0.000]	[0.000]
Pre-trend F (narrow)	1.31	6.46	4.42	2.30
	[0.268]	[0.000]	[0.001]	[0.048]

Notes: ATT is the mean of post-event coefficients ($k = 0, \ldots, 12$).
Full pre-period: $k = -12$ to -2 (11 coef.); narrow: $k = -7$ to -2 (6 coef.).
Standard errors in parentheses; p-values in brackets. Specifications include destination×month, item×month FE, unit trend, seasonality, AI demand control; inverse-value weighted.

firms to re-optimize quantities in response to the trade wedge. The algorithm proceeds as follows:

1. **Retrieve Baseline Prices.** Let p_0^{U*} denote the vector of negotiated link prices in the baseline equilibrium (Section 4.1), solved under baseline wedges $\boldsymbol{\tau}^0$. Let $\mathcal{L}$ be the set of active links.

2. **Apply Statutory Shock.** Introduce the counterfactual buyer-specific output wedges $\boldsymbol{\tau}^1$. For example, in the NVIDIA-exposed scenario, $\tau^1_{\text{NVIDIA}} = 1.10 \times \tau^0_{\text{NVIDIA}}$.

3. **Compute Frozen Costs.** Calculate the new downstream unit costs $\mathsf{C}^D_{\text{frozen}}$ using the *shocked* wedges $\boldsymbol{\tau}^1$ but the *fixed* baseline link prices p_0^{U*}. For each buyer d:

$$C_d^D(\text{frozen}) \;=\; \frac{\tau_d^1}{z_d}\left(\sum_{u\in\mathcal{U}_d}(p_{0,du}^{U*})^{1-\eta}\right)^{\frac{1}{1-\eta}}. \tag{18}$$

Crucially, this step bypasses the Nash-in-Nash update loop; firms calculate costs as if

supplier contracts were rigid.

4. **Solve Partial Equilibrium.** Given the fixed cost vector $\mathsf{C}^D_{\text{frozen}}$, I solve for the new downstream Cournot quantities q^D and prices p^D using the algorithm in Definition 1.

The decomposition in Equation (16) is then computed by comparing the full Nash-in-Nash outcome (where p^U adjusts to p_1^{U*}) against this frozen benchmark. The difference between the total price change and the frozen-benchmark price change identifies the contribution of $\Delta \log P_d^M$.

D.2 Heterogeneous Shock Protocols

The results in Figures 6 and 7 utilize two standardized asymmetric shock vectors to represent different "Chip War" compliance scenarios. These are defined in the calibration as:

- **Scenario A (NVIDIA-exposed):** $\%\Delta\tau_{\text{NVIDIA}} = +10\%$, $\%\Delta\tau_{\text{AMD}} = +5\%$. This corresponds to a regime where export controls bind significantly tighter on the dominant firm's high-performance SKUs (e.g., H100/H200 restriction).
- **Scenario B (AMD-exposed):** $\%\Delta\tau_{\text{NVIDIA}} = +5\%$, $\%\Delta\tau_{\text{AMD}} = +10\%$. This represents a regime where the dominant firm successfully mitigates compliance costs (e.g., via specific licenses or China-tailored chips) while the rival faces steeper effective barriers.

For the continuous sweep in Figure 8, I vary $\%\Delta\tau_{\text{NVIDIA}}$ from 5% to 50% in 5% increments while holding $\Delta\tau_{\text{AMD}} = 0$, solving the full Nash-in-Nash system at each step to trace the non-linear response of input prices.

E Proof of Proposition 1

The first part of the proof establishes the demand curvature objects that drive both markup formulas and diagonal dominance.

Lemma 2 (Demand curvature and perceived elasticity)**.** *Under* (2)*,*

$$p_d^D = A\Big(\tfrac{q_d^D}{Q^D}\Big)^{-1/\sigma}\big(Q^D\big)^{-1/\rho}, \qquad Q^D = \Big(\sum_k \big(q_k^D\big)^{(\sigma-1)/\sigma}\Big)^{\sigma/(\sigma-1)}, \quad \sigma > 1,\ \rho \in (1,\sigma],$$

let

$$s_d = \frac{\big(q_d^D\big)^{(\sigma-1)/\sigma}}{\sum_k (q_k^D)^{(\sigma-1)/\sigma}} \in (0,1), \qquad \epsilon_d = \Big(\tfrac{s_d}{\rho} + \tfrac{1-s_d}{\sigma}\Big)^{-1}.$$

Then, holding q_{-d} *fixed:*

(a) ***Perceived elasticity.*** $\dfrac{\partial \log p_d^D}{\partial \log q_d^D} = -\dfrac{1}{\epsilon_d} = -\Big[\tfrac{1-s_d}{\sigma} + \tfrac{s_d}{\rho}\Big].$

(b) ***Marginal revenue.*** $MR_d = p_d^D\big(1 - \tfrac{1}{\epsilon_d}\big)$ *and* $\partial MR_d/\partial q_d^D < 0.$

(c) ***Markup FOC.*** *With constant unit cost* $C_d^D > 0$*,*

$$p_d^D = \frac{\epsilon_d}{\epsilon_d - 1}\, C_d^D.$$

Proof. Write $\log p_d^D = \log A - \frac{1}{\sigma}\log q_d^D + \Big(\frac{1}{\sigma} - \frac{1}{\rho}\Big)\log Q^D$. Let $M = \sum_k \big(q_k^D\big)^{\frac{\sigma-1}{\sigma}}$, so $\log Q^D = \frac{\sigma}{\sigma-1}\log M$ and

$$\frac{\partial \log M}{\partial \log q_d^D} = \frac{q_d^D\, \partial M/\partial q_d^D}{M} = \frac{\sigma-1}{\sigma}\frac{\big(q_d^D\big)^{\frac{\sigma-1}{\sigma}}}{M} = \frac{\sigma-1}{\sigma}\, s_d.$$

Hence $\frac{\partial \log Q^D}{\partial \log q_d^D} = s_d$ and

$$\frac{\partial \log p_d^D}{\partial \log q_d^D} = -\frac{1}{\sigma} + \Big(\frac{1}{\sigma} - \frac{1}{\rho}\Big)s_d = -\Big[\frac{1-s_d}{\sigma} + \frac{s_d}{\rho}\Big] = -\frac{1}{\epsilon_d},$$

which proves (i).

For (ii), note that $\partial p_d^D/\partial q_d^D = (p_d^D/q_d^D)\,\partial \log p_d^D/\partial \log q_d^D = -(p_d^D/q_d^D)(1/\epsilon_d)$. Thus $MR_d = p_d^D + q_d^D \partial p_d^D/\partial q_d^D = p_d^D(1 - 1/\epsilon_d)$. Differentiate w.r.t. q_d^D (holding q_{-d} fixed):

$$\frac{\partial MR_d}{\partial q_d^D} = \underbrace{\frac{\partial p_d^D}{\partial q_d^D}}_{<0}\Big(1 - \frac{1}{\epsilon_d}\Big) - p_d^D\,\frac{\partial}{\partial q_d^D}\Big(\frac{1}{\epsilon_d}\Big).$$

Since s_d increases with q_d^D and $\frac{\partial}{\partial s_d}(1/\epsilon_d) = (1/\rho - 1/\sigma) \geq 0$ (by $\rho \leq \sigma$), $\frac{\partial}{\partial q_d^D}(1/\epsilon_d) > 0$. The first term is negative (downward-sloping inverse demand with $\epsilon_d > 1$), and the second term is also negative, so $\partial MR_d/\partial q_d^D < 0$.

For (iii), with constant unit cost C_d^D, the Cournot FOC sets $MR_d = C_d^D$. Using the expression in (ii), $p_d^D(1 - 1/\epsilon_d) = C_d^D$, which rearranges to the stated markup formula. □

Building on Lemma 2, the next result pins down existence and a simple condition for uniqueness of the Cournot equilibrium.

Lemma 3 (Cournot equilibrium: existence and uniqueness)**.** *Let $\sigma > 1$ and $\rho \in (1, \sigma]$. With constant unit costs $C_d^D > 0$, the Cournot game admits an equilibrium. Moreover, the equilibrium is unique if*

$$\sigma \leq 2\rho.$$

Proof. Existence. The profit function π_d is continuous and strictly concave in q_d (since $MR_d' < 0$ by Lemma 2). Strategies are bounded; existence follows from Novshek (1985).

Uniqueness. Let J be the Jacobian of the marginal profit functions $g_d(\mathsf{q}) = MR_d - C_d^D$. For uniqueness, it suffices to show that J is strictly diagonally dominant (Rosen 1965), i.e., $|J_{dd}| > \sum_{k \neq d} |J_{dk}|$. Note first that under substitute goods, cross-effects are negative ($\partial MR_d/\partial q_k < 0$).

Using the elasticity formulas from Lemma 2, we compute the elasticity-scaled Jacobian terms:

$$-\frac{q_d}{p_d} J_{dd} = \left[\frac{1 - s_d}{\sigma} + \frac{s_d}{\rho}\right] + \Delta s_d(1 - s_d)$$

$$-\frac{q_k}{p_d} J_{dk} = \Delta s_k(1 - s_d)$$

where $\Delta \equiv \frac{1}{\rho} - \frac{1}{\sigma}$. Summing the cross-terms over $k \neq d$ (and noting $\sum_{k \neq d} s_k = 1 - s_d$):

$$\sum_{k \neq d} \left|\frac{q_k}{p_d} J_{dk}\right| = \Delta(1 - s_d) \sum_{k \neq d} s_k = \Delta(1 - s_d)^2.$$

The condition for diagonal dominance (after normalizing by p_d) is effectively:

$$\frac{q_d}{p_d}|J_{dd}| > \sum_{k \neq d} \frac{q_k}{p_d}|J_{dk}| \cdot \frac{q_d}{q_k}$$

Define $f(s_d)$ as the LHS minus the RHS:

$$f(s_d) \;=\; \frac{1-s_d}{\sigma} + \frac{s_d}{\rho} + \Delta s_d(1-s_d) - \Delta(1-s_d)^2 \;=\; \frac{1-s_d}{\sigma} + \frac{s_d}{\rho} + \Delta(1-s_d)(2s_d-1),$$

where $\Delta = 1/\rho - 1/\sigma > 0$. Differentiating with respect to s_d:

$$f'(s_d) = -\frac{1}{\sigma} + \frac{1}{\rho} + \Delta\Big[-(2s_d-1) + (1-s_d)\cdot 2\Big] = \Delta + \Delta(3-4s_d) = 4\Delta(1-s_d) \;\geq\; 0.$$

Hence f is nondecreasing on $[0,1]$ and its minimum is attained at $s_d = 0$:

$$f(0) = \frac{1}{\sigma} + \Delta\cdot(-1) = \frac{1}{\sigma} - \frac{1}{\rho} + \frac{1}{\sigma} = \frac{2}{\sigma} - \frac{1}{\rho}.$$

This is nonnegative if and only if $\sigma \leq 2\rho$. Thus J is a P-matrix globally, and by Rosen (1965), the equilibrium is unique. □

Uniqueness delivers clean comparative statics with respect to unit costs and their structural shifters.

Corollary 2 (Comparative statics)**.** *Under Lemma 3, for each firm d, $\partial q_d^*/\partial C_d < 0$ and $\partial s_d^*/\partial C_d < 0$. With $C_d^D = (\tau_d/z_d)P_d^M$ the same signs apply to (z_d, τ_d, P_d^M) in the obvious way.*

Proof. Let $F_d(\mathsf{q};\mathsf{C}) = MR_d(\mathsf{q}) - C_d$ denote firm d's first-order condition. By Lemma 2, $\partial F_d/\partial q_d < 0$. Let G be the Jacobian of the system. From Lemma 3, G is a stable matrix with negative diagonals. Standard comparative statics for stable games of strategic substitutes imply that the inverse Jacobian G^{-1} has negative diagonal entries ($(G^{-1})_{dd} < 0$) and positive off-diagonal entries ($(G^{-1})_{kd} > 0$).

(a) Quantity. By the Implicit Function Theorem:

$$\frac{\partial \mathsf{q}^*}{\partial C_d} = G^{-1} e_d.$$

Thus $\frac{\partial q_d^*}{\partial C_d} = (G^{-1})_{dd} < 0$. (Own output falls when own cost rises).

Market share. Differentiating the share definition $s_d \propto q_d^{(\sigma-1)/\sigma} / \sum q_k^{(\sigma-1)/\sigma}$ yields:

$$\frac{\partial s_d}{\partial C_d} \propto (1 - s_d) \frac{\partial q_d^*}{\partial C_d} - \sum_{k \neq d} s_k \frac{q_d^*}{q_k^*} \frac{\partial q_k^*}{\partial C_d}.$$

(Note: we use elasticities of substitution to normalize). Since $\partial q_d^*/\partial C_d < 0$ and $\partial q_k^*/\partial C_d > 0$ (rivals expand when my cost rises), both terms contribute negatively. Thus $\partial s_d/\partial C_d < 0$.

(b) Markup. The markup is $\mu_d = \epsilon_d/(\epsilon_d - 1)$. This is strictly decreasing in ϵ_d. From Lemma 2, if $\sigma > \rho$, perceived elasticity ϵ_d is strictly decreasing in market share s_d (larger firms face more inelastic demand). Chain rule: $C_d \uparrow \implies s_d \downarrow \implies \epsilon_d \uparrow \implies \mu_d \downarrow$. Thus, pass-through is incomplete and markups fall.

(c) Structural shifters. Follows immediately from $C_d^D = (\tau_d/z_d) P_d^M$. □

Turning to the vertical stage, I characterize the shape of incremental surpluses on a single link.

Lemma 4 (Link–surplus shape)**.** *Fix a link (d, u), and let $\eta > 1$ be the elasticity in P_d^M. There exists a nonempty feasible price interval $\mathcal{P} = (\underline{p}, \overline{p})$ such that on $\mathcal{P}$:*

(a) ***Buyer side.*** *The incremental surplus $\Delta\Pi_d^D(p)$ is positive and strictly decreasing in p. Moreover, when the buyer curvature dominance inequality in Assumption 2 holds, $\Delta\Pi_d^D$ is (strictly) log–concave in p on $\mathcal{P}$.*

(b) ***Seller side.*** *Let C_{du}^U denote the supplier's unit cost on (d, u) and set*

$$p^\dagger := \frac{\eta}{\eta - 1} C_{du}^U.$$

On any interval contained in $(C_{du}^U,\, p^\dagger]$*, the seller's incremental surplus* $\Delta\Pi_u^U(p)$ *is positive, strictly increasing, and strictly log–concave in* p*.*

Proof. Assumption 2(a) implies a buyer–curvature dominance inequality

$$-\frac{\partial \ln q_d^{D*}}{\partial \ln C_d^D} \;\geq\; \eta\,\frac{1-\omega_{du}}{\omega_{du}} \qquad \text{for all } p \in \mathcal{P}_{du}. \tag{19}$$

This follows from Assumption 2(a) because $\omega_{du} \in [\varepsilon, 1-\varepsilon]$ implies $\frac{1-\omega_{du}}{\omega_{du}} \leq \frac{1-\varepsilon}{\varepsilon}$, so $-\partial \ln q_d^{D*}/\partial \ln C_d^D \geq \eta(1-\varepsilon)/\varepsilon$ yields condition (19).

(a) Buyer side. By Cournot uniqueness (Lemma 3), the profit function $\Pi_d^D(C_d^D)$ is C^2 in the unit cost C_d^D and satisfies

$$\frac{\partial \Pi_d^D}{\partial C_d^D} = -\,q_d^D < 0, \qquad \frac{\partial^2 \Pi_d^D}{\partial (C_d^D)^2} = -\,\frac{\partial q_d^D}{\partial C_d^D} \;>\; 0,$$

the second inequality by the own–cost comparative statics (Corollary 2). With the CES index $P_d^M(p) = \left(\sum_{u'} (p_{du'}^U)^{1-\eta}\right)^{1/(1-\eta)}$ and $C_d^D = (\tau_d/z_d) P_d^M$, holding other upstream prices fixed gives

$$\frac{\partial C_d^D}{\partial p} = \frac{\omega_{du}\, C_d^D}{p} > 0, \qquad \frac{\partial^2 C_d^D}{\partial p^2} = -\,\frac{\eta\,\omega_{du}(1-\omega_{du})\, C_d^D}{p^2} \leq 0,$$

where $\omega_{du} = \left(p/P_d^M\right)^{1-\eta} \in (0,1)$ is the cost share of input u. By the envelope rule and chain rule,

$$\frac{d}{dp}\,\Pi_d^D\big(C_d^D(p)\big) = \Pi_C^D\,\frac{\partial C_d^D}{\partial p} = -\,q_d^D\,\frac{\partial C_d^D}{\partial p} \;<\; 0,$$

so the buyer's incremental surplus $\Delta\Pi_d^D(p) := \Pi_d^D(C_d^D(p)) - \Pi_d^D(C_d^D(\infty))$ is strictly decreasing wherever it is positive. A direct differentiation (as in the log–concavity footnote given earlier) yields

$$(\log \Delta\Pi_d^D)'' \leq 0 \quad\Longleftrightarrow\quad -\frac{\partial \ln q_d^D}{\partial \ln C_d^D} \;\geq\; \eta\,\frac{1-\omega_{du}}{\omega_{du}},$$

which is exactly the buyer–curvature dominance condition (19). Hence, under (19), $\Delta\Pi_d^D$ is (strictly) log–concave on any interval where it is positive.

(b) Seller side. When the link is severed, the seller supplies zero on (d, u), so for a given p, the incremental surplus is

$$\Delta\Pi_u^U(p) = \Big(p - C_{du}^U\Big)\, q_{du}^U(p).$$

With rivals' prices held fixed, the CES structure implies within–bundle Hicksian demand

$$q_{du}^U(p) \;=\; X_d \left(\frac{p}{P_d^M}\right)^{-\eta},$$

where $X_d > 0$ collects the composite quantity scale.[22] Hence

$$\Delta\Pi_u^U(p) = X_d\,(p - C_{du}^U)\,p^{-\eta}\,(P_d^M)^{\eta}.$$

Differentiating,

$$\frac{d}{dp}\Delta\Pi_u^U(p) = X_d\,p^{-\eta-1}(P_d^M)^{\eta}\,\Big(\eta\,C_{du}^U - (\eta-1)p\Big),$$

so $\Delta\Pi_u^U$ is strictly increasing on $(C_{du}^U,\, p^\dagger)$ and strictly decreasing on $(p^\dagger, \infty)$, with a unique interior maximizer at $p^\dagger = \frac{\eta}{\eta-1}C_{du}^U$. For log–concavity,

$$\frac{d^2}{dp^2}\Big(\log\Delta\Pi_u^U(p)\Big) = \frac{d^2}{dp^2}\Big(\log(p - C_{du}^U) - \eta\log p\Big) = -\frac{1}{(p - C_{du}^U)^2} + \frac{\eta}{p^2} < 0$$

whenever $p \le p^{\mathrm{lc}} := \frac{\sqrt{\eta}}{\sqrt{\eta}-1}\,C_{du}^U$. Since $p^\dagger \le p^{\mathrm{lc}}$ for all $\eta > 1$, $\Delta\Pi_u^U$ is strictly log–concave on $(C_{du}^U,\, p^\dagger]$.

Feasible interval. Let $\overline{p}^{\,D}$ be the unique price at which the buyer becomes indifferent, i.e. $\Delta\Pi_d^D(\overline{p}^{\,D}) = 0$ (existence/uniqueness by strict monotonicity). Set $\overline{p} := \min\{p^\dagger, \overline{p}^{\,D}\}$ and choose any $\underline{p} \in (C_{du}^U,\, \overline{p})$ small enough that both surpluses are positive. Then $\mathcal{P} = (\underline{p}, \overline{p})$

[22] This keeps the standard "within–bundle" (Hicksian) response separate from the induced Cournot feedback through q_d^D. The latter is controlled globally by uniqueness and compactness; it can only shrink the interval on which the monotonicity/log–concavity claims hold. The stated interval $(C_{du}^U, p^\dagger]$ is valid under the within–bundle calculus and remains valid (possibly after a harmless inward adjustment) once the Cournot feedback is reintroduced; see the small-gain arguments in the main text.

satisfies the claims in (a)–(b), completing the proof. □

The buyer-side log–concavity depends on the curvature–dominance inequality (19). Economically, it requires that the (absolute) cost elasticity of the downstream quantity be large enough relative to the input's cost share ω_{du} (which itself is small unless a single input dominates the bundle). On the seller side, no additional assumption is needed beyond $\eta > 1$. The surplus is strictly quasi–concave with a unique interior peak at $p^\dagger$, and it is strictly log–concave on $(C_{du}^U, p^\dagger]$.

The preceding shape guarantees a unique bargained price when rivals' prices are held fixed.

Lemma 5 (Nash-in-Nash on one link)**.** *Fix rivals' prices. Let $\gamma \in (0,1)$ and suppose the buyer-side curvature–dominance inequality (19) holds so that $\log \Delta\Pi_d^D$ is concave in own link price on the feasible set. Then the one-link Nash-in-Nash problem has a unique interior maximizer $p^* \in (C_{du}^U, \overline{p})$, where $\overline{p} = \min\{p^\dagger, \overline{p}^D\}$ with $p^\dagger = \frac{\eta}{\eta-1} C_{du}^U$ and $\overline{p}^D$ the buyer-indifference price from Lemma 4.*

Proof. Let $p \equiv p_{du}^U$ and write the log–Nash objective

$$h(p) \;:=\; \gamma \log \Delta\Pi_d^D(p) + (1-\gamma) \log \Delta\Pi_u^U(p).$$

By Lemma 4, on $(C_{du}^U, \overline{p})$ I have $\Delta\Pi_d^D(p) > 0$ and $\Delta\Pi_u^U(p) > 0$, and $\log \Delta\Pi_u^U$ is strictly concave; under (19), $\log \Delta\Pi_d^D$ is concave. Therefore, h is *strictly concave* on $(C_{du}^U, \overline{p})$ because it is a positive-weighted sum of a strictly concave and a concave function.

Compute the derivative:

$$h'(p) = \gamma \, \frac{\Delta\Pi_d^{D\,\prime}(p)}{\Delta\Pi_d^D(p)} + (1-\gamma) \, \frac{\Delta\Pi_u^{U\,\prime}(p)}{\Delta\Pi_u^U(p)}.$$

As $p \uparrow \overline{p}$, there are two cases. If $\overline{p} = \overline{p}^D$, then $\Delta\Pi_d^D(p) \downarrow 0$ while $\Delta\Pi_d^{D\,\prime}(p)$ stays finite and negative, so $h'(p) \to -\infty$. If $\overline{p} = p^\dagger$, then $\Delta\Pi_u^{U\,\prime}(p^\dagger) = 0$ and $\Delta\Pi_d^{D\,\prime}(p^\dagger) < 0$, hence $h'(\overline{p}^-) < 0$.

As $p \downarrow C_{du}^U$, I have $\Delta\Pi_u^U(p) = (p - C_{du}^U)\,\kappa_{du}\,p^{-\eta} \downarrow 0$ and

$$\frac{\Delta\Pi_u^{U\,\prime}(p)}{\Delta\Pi_u^U(p)} = \frac{1}{p - C_{du}^U} - \frac{\eta}{p} \;\longrightarrow\; +\infty,$$

while $\Delta\Pi_d^D(p)$ remains bounded away from zero and $\Delta\Pi_d^{D\,\prime}(p)$ remains finite, so the buyer term contributes a finite number. Hence $h'(p) \to +\infty$ as $p \downarrow C_{du}^U$.

By continuity of h' on $(C_{du}^U, \overline{p})$ and the opposite boundary limits, there exists $p^* \in (C_{du}^U, \overline{p})$ with $h'(p^*) = 0$. Strict concavity of h implies this stationary point is unique and globally maximizing on $(C_{du}^U, \overline{p})$. (Note logs enforce interiority: both surpluses are positive on the open interval.) This proves the claim. □

Lemma 5 solves the *single-link* problem: the log–Nash objective is strictly concave in its own price, so the link admits a unique interior optimizer. By contrast, the Nash-in-Nash outcome is the joint root of the stacked first-order conditions $\Phi(\mathsf{p}^U) = 0$, one per link. Links interact only through P_d^M and the induced Cournot quantities; under the Cournot curvature and uniqueness (Lemmas 2–3 and Corollary 2), these cross-effects are dominated by own-link curvature (Lemma 4). Hence the Jacobian $J = \partial\Phi/\partial\mathsf{p}^U$ is a P-matrix (diagonal strict concavity in the sense of Rosen), delivering a unique fixed point p^{U*} and the comparative statics in Proposition 1. (Equivalently, cyclic best responses that replace each link by its Lemma 5 optimizer converge to p^{U*}.)

Preparing the comparative statics part of Proposition 1, I show that the Jacobian matrix J is an M-matrix given the uniqueness condition for the downstream Cournot competition.

Lemma 6 (NiN Jacobian is an M–matrix)**.** *Fix the active–link set $\mathcal{L}$. Let $g(\mathsf{p}^U)$ be the log–Nash pseudo–gradient stacked over links $L = (d, u) \in \mathcal{L}$, and let $J := \partial g/\partial\mathsf{p}^U$ be its Jacobian. Under Assumptions 1 and 2:*

(a) J has strictly negative diagonals and weakly non–positive off–diagonals (i.e., $J_{LL} < 0$ and $J_{LL'} \le 0$ for $L' \neq L$).

(b) There exists a positive diagonal scaling $R := \operatorname{diag}(p_L)_{L\in\mathcal{L}}$ *such that the symmetric part*

$$S := \tfrac{1}{2}\left(RJ + J^\top R\right)$$

is negative definite on the feasible box $\mathcal{P}$*; in particular, each row of* RJ *is strictly diagonally dominant:*

$$(RJ)_{LL} < 0 \quad \textit{and} \quad \sum_{L'\neq L} \left|(RJ)_{LL'}\right| < -\,(RJ)_{LL} \qquad \forall\, L.$$

(c) Setting $A := -J$*, I have that* A *is a nonsingular* M*–matrix. Consequently* $(-J)^{-1} = A^{-1} \geq 0$ *(entrywise), and* J *is a* P*–matrix.*

Proof. (a) By Lemma 4 and Assumption 2(a), each link's own log–surplus is (strictly) concave in its own price, hence $J_{LL} < 0$. Cross–link effects operate only via the CES materials price index within a buyer and via the Cournot block across buyers; both are strategic substitutes under Lemmas 2–3, so $J_{LL'} \leq 0$ for $L' \neq L$.

(b) With $R = \operatorname{diag}(p_L)$ and the bounds developed in Step 3, Assumption 1 gives uniform share and margin bounds, while Assumption 2(a)–(b) bounds own–curvature below and cross-feedback effects above. Hence, for each row L,

$$(RJ)_{LL} \leq -\,\underline{\kappa}_d/p_L < 0, \qquad \sum_{L'\neq L} \left|(RJ)_{LL'}\right| \leq \chi_d/p_L$$

with $\chi_d < \underline{\kappa}_d$ buyer–by–buyer. Gershgorin implies $S = \frac{1}{2}(RJ + J^\top R) \prec 0$ on $\mathcal{P}$.

(c) Write $A := -J$. Then A is a Z–matrix with nonnegative off–diagonals, and by (b) I can take $x = (p_L)_{L\in\mathcal{L}} > 0$ to get

$$Ax \;=\; -Jx \;=\; -R^{-1}(RJ)\,x \;=\; -R^{-1}(RJ)\mathbf{1},$$

whose L–th component equals $(\widetilde{A}\mathbf{1})_L$ with $\widetilde{A} := -RJ$, strictly positive by the strict diagonal

dominance established in (b). Thus $Ax > 0$. By the positive–vector characterization of nonsingular M–matrices (e.g., Berman–Plemmons), A is a nonsingular M–matrix; hence $A^{-1} \geq 0$ and J is a P–matrix. □

Proof of Proposition 1 (NiN system: uniqueness and comparative statics).

Step 1 (Curvature of the Cournot block). By Lemma 2 and Lemma 3, the Cournot stage has a unique equilibrium for any fixed vector of unit costs; own best-reply slopes dominate cross effects (diagonal dominance).

Step 2 (Own-link strict concavity). By Lemma 4 and Lemma 5, holding rivals' link prices fixed, the log Nash product for each link is strictly concave in its own price.

Step 3 (Joint system and weighted diagonal dominance). Let $\mathcal{L}$ be the set of active links and write $\mathsf{p}^U = (p_L)_{L\in\mathcal{L}}$ with $L = (d, u)$. Define for each link

$$h_L(p_L; \mathsf{p}^U_{-L}) = \gamma_d \log \Delta\Pi^D_d(p_L; \mathsf{p}^U_{-L}) + (1 - \gamma_d) \log \Delta\Pi^U_u(p_L; \mathsf{p}^U_{-L}),$$

the pseudo–gradient $g = (g_L)_L$ with $g_L := \partial h_L/\partial p_L$, and its Jacobian $J := \partial g/\partial \mathsf{p}^U$. Work on the compact box $\mathcal{P} = \prod_L [\underline{p}_L, \overline{p}_L]$ given by Assumption 1.

(a) *Own curvature (negative diagonals with a primitive bound).* By Lemma 4 and Assumption 2(a), $\log \Delta\Pi^D_d$ is concave in p_L, and $\log \Delta\Pi^U_u$ is strictly concave. Hence $J_{LL} < 0$ and, using $\omega_{du} \in [\varepsilon, 1 - \varepsilon]$,

$$J_{LL} = \frac{\partial^2 h_L}{\partial p_L^2} \;\leq\; -\frac{1}{p_L^2}\Big[\gamma_d\,\eta\,\frac{1-\varepsilon}{\varepsilon} + (1-\gamma_d)\,\eta\Big].$$

Let $R := \mathrm{diag}(p_L)_{L\in\mathcal{L}}$ and $S := \frac{1}{2}(RJ + J^\top R)$; then

$$S_{LL} \;=\; p_L J_{LL} \;\leq\; -\frac{1}{p_L}\Big[\gamma_d\,\eta\,\frac{1-\varepsilon}{\varepsilon} + (1-\gamma_d)\,\eta\Big].$$

(b) *Cross terms (uniform primitive bounds).* Links interact only (i) within a buyer via P^M_d

and (ii) across buyers via the Cournot block. Using the CES identities (each cross term contributes a factor $\propto \omega/p$) and the standard nested-CES demand bounds (Lemmas 2–3),

$$\sum_{\substack{L'=(d,u')\\ L'\neq L}} |S_{LL'}| \;\le\; \frac{1}{p_L}\,\frac{1-\varepsilon}{\varepsilon} \underbrace{\Big[\gamma_d\,\eta\Big]}_{\text{within-buyer bound}}, \qquad \sum_{\substack{L'=(d',u')\\ d'\neq d}} |S_{LL'}| \;\le\; \frac{1}{p_L} \underbrace{\Big(\frac{1}{\rho}-\frac{1}{\sigma}\Big)}_{\text{Cournot cross cap}}.$$

By Assumption 2(b) I have $\frac{1}{\rho}-\frac{1}{\sigma} < (1-\gamma_d)\,\eta$, hence

$$\sum_{L'\neq L} |S_{LL'}| \;\le\; \frac{1}{p_L}\Big[\gamma_d\,\eta\,\frac{1-\varepsilon}{\varepsilon} + (1-\gamma_d)\,\eta\Big] \quad \text{with strict inequality from the across-buyer part.}$$

(c) *Gershgorin dominance and a P–matrix.* Combining (a)–(b),

$$\sum_{L'\neq L} |S_{LL'}| \;<\; -\,S_{LL} \qquad \text{for every row } L,$$

so all Gershgorin disks of S lie strictly on the negative real axis and $S \prec 0$ on $\mathcal{P}$. Therefore, the game is diagonally strictly concave in the sense of Rosen (1965); J is a P–matrix and the NiN system $g(\mathsf{p}^U) = 0$ admits a unique solution p^{U*}.

F Proof of Proposition 2

Step 1 (Regularity and derivative formula). Fix the active set of links and work on the compact box $\mathcal{P}$ from Assumption 1. By Lemmas 2–5 and Assumption 2: (i) the Cournot equilibrium map $C^D \mapsto q^D(C^D)$ is C^1 in all arguments; (ii) the CES index $P_d^M(\cdot)$ and the transforms $p \mapsto \log(p - C_{du}^U)$ are C^∞ on $\{p \ge C_{du}^U + m\}$; hence (iii) each link objective h_L and the pseudo–gradient

$$g(\mathbf{p}^U, \Theta) = \Big(g_L(\mathbf{p}^U, \Theta)\Big)_{L\in\mathcal{L}} = \Big(\frac{\partial h_L}{\partial p_L}(\mathbf{p}^U, \Theta)\Big)_{L\in\mathcal{L}}$$

are C^1 in $(\mathbf{p}^U, \Theta)$ for any collection of primitives Θ (e.g. $\boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{z}, \ldots$).

By Proposition 1 and Lemma 6, the Jacobian

$$J(\mathbf{p}^U, \Theta) \;=\; \frac{\partial g(\mathbf{p}^U, \Theta)}{\partial \mathbf{p}^U}$$

is a P– and M–matrix at $\mathbf{p}^{U*}$, hence nonsingular. Therefore, by the Implicit Function Theorem, there exist neighborhoods $\mathcal{V}$ of Θ and $\mathcal{U}$ of $\mathbf{p}^{U*}$ and a C^1 map $\phi : \mathcal{V} \to \mathcal{U}$ such that

$$g\Big(\phi(\Theta), \Theta\Big) = 0, \qquad \phi(\Theta_0) = \mathbf{p}^{U*}(\Theta_0).$$

Differentiating $g(\phi(\Theta), \Theta) \equiv 0$ with respect to any scalar primitive Θ_k yields

$$\frac{\partial \mathbf{p}^{U*}}{\partial \Theta_k} \;=\; -\,J\Big(\mathbf{p}^{U*}(\Theta), \Theta\Big)^{-1} \left.\frac{\partial g(\mathbf{p}^U, \Theta)}{\partial \Theta_k}\right|_{\mathbf{p}^U = \mathbf{p}^{U*}(\Theta)}.$$

Under Assumptions 1–2, Lemma 6 further shows that $-J(\mathbf{p}^{U*}(\Theta), \Theta)^{-1}$ is entrywise nonnegative. This proves part (a) of Proposition 2.

Step 2 (Buyer–specific wedges and productivity). Fix a buyer d and an active link $L = (d, u)$ with $\omega_{du} \in (\varepsilon, 1-\varepsilon)$ and $p^{U*}_{du} - C^U_{du} > 0$ at the baseline. On the box $\mathcal{P}$ and in a neighbourhood of Θ, the local derivative formulas in Step 1 apply with $\Theta_k \in \{\ln \tau_d, \ln z_d\}$.

By construction, the unit cost for the buyer d satisfies $C^D_d \propto \tau_d / z_d$. Holding the active set fixed, log changes in τ_d or z_d therefore operate through the same cost channel, with opposite signs. Differentiating the one-link first-order condition $g_L(\mathbf{p}^U, \Theta) = 0$ with respect to $\ln \tau_d$ or $\ln z_d$, keeping rival link prices temporarily fixed, yields expressions of the form

$$\frac{\partial g_L}{\partial \ln \tau_d} = \Xi^{(\tau)}_{d,u}(\Theta), \qquad \frac{\partial g_L}{\partial \ln z_d} = \Xi^{(z)}_{d,u}(\Theta),$$

where $\Xi^{(\tau)}_{d,u}$ and $\Xi^{(z)}_{d,u}$ are continuous functions of primitives and equilibrium objects (shares, markups, demand elasticities). No additional global sign restrictions on these objects follow from the curvature inequalities alone: buyer surplus and seller surplus both respond to C^D_d

and p^U_{du}, and the net effect depends on their relative magnitudes at the calibrated equilibrium.

Stacking across all links and applying the derivative formula in Step 1, the elasticities for the focal link can be written as

$$\frac{\partial \ln p^{U*}_{du}}{\partial \ln \tau_d} \;=\; \Psi^{(\tau)}_{d,u}(\Theta) := \frac{1}{p^{U*}_{du}}\Big[-\; J(\mathbf{p}^{U*},\Theta)^{-1}\frac{\partial g(\mathbf{p}^{U*},\Theta)}{\partial \ln \tau_d}\Big]_L,$$

$$\frac{\partial \ln p^{U*}_{du}}{\partial \ln z_d} \;=\; \Psi^{(z)}_{d,u}(\Theta) := \frac{1}{p^{U*}_{du}}\Big[-\; J(\mathbf{p}^{U*},\Theta)^{-1}\frac{\partial g(\mathbf{p}^{U*},\Theta)}{\partial \ln z_d}\Big]_L,$$

where $[\cdot]_L$ denotes the Lth component. Because $-J^{-1}$ is entrywise nonnegative and the shock vectors $\partial g/\partial \ln \tau_d$ and $\partial g/\partial \ln z_d$ are continuous in $(\mathbf{p}^{U*},\Theta)$, the functions $\Psi^{(\tau)}_{d,u}$ and $\Psi^{(z)}_{d,u}$ are continuous in Θ on the region where the active set is unchanged. The same representation applies to any rival link $L' = (d',u')$ with $d' \neq d$, yielding well–defined elasticities $\partial \ln p^{U*}_{d'u'}/\partial \ln \tau_d$ and $\partial \ln p^{U*}_{d'u'}/\partial \ln z_d$.

The proposition therefore, does not impose a global sign on $\Psi^{(\tau)}_{d,u}$ or $\Psi^{(z)}_{d,u}$ from primitives alone: their signs are determined locally at the calibrated equilibrium through the shock vectors and the nonnegative matrix $-J^{-1}$. This proves part (b).

Step 3 (Bargaining weights). Consider now a change in the bargaining weight γ_d for buyer d. For any link $L = (d,u)$ attached to d,

$$g_L(\mathbf{p}^U,\Theta) = \gamma_d\,\frac{\partial \log \Delta\Pi^D_d}{\partial p^U_{du}} + (1-\gamma_d)\,\frac{\partial \log \Delta\Pi^U_u}{\partial p^U_{du}},$$

where $\Delta\Pi^D_d$ and $\Delta\Pi^U_u$ are buyer and seller incremental surpluses on link L. Differentiating with respect to γ_d gives

$$\frac{\partial g_L}{\partial \gamma_d} = \frac{\partial}{\partial p^U_{du}}\Big(\log \Delta\Pi^D_d - \log \Delta\Pi^U_u\Big).$$

By Lemma 4, along an active link, the buyer's incremental surplus $\Delta\Pi^D_d$ is strictly decreasing and log–concave in p^U_{du}, while the seller's incremental surplus $\Delta\Pi^U_u$ is strictly increasing and

log–concave in p^U_{du}. Hence

$$\frac{\partial}{\partial p^U_{du}} \log \Delta\Pi^D_d \;<\; 0, \qquad \frac{\partial}{\partial p^U_{du}} \log \Delta\Pi^U_u \;>\; 0,$$

so that

$$\frac{\partial g_L}{\partial \gamma_d} \;=\; \frac{\partial}{\partial p^U_{du}} \log \Delta\Pi^D_d - \frac{\partial}{\partial p^U_{du}} \log \Delta\Pi^U_u \;<\; 0.$$

For any link $L' = (d', u')$ with $d' \neq d$, the FOC $g_{L'}$ does not depend directly on γ_d, so $\partial g_{L'}/\partial \gamma_d = 0$.

Stacking across links and using the derivative formula from Step 1,

$$\frac{\partial \mathbf{p}^{U*}}{\partial \gamma_d} \;=\; -\, J(\mathbf{p}^{U*}, \Theta)^{-1} \frac{\partial g(\mathbf{p}^{U*}, \Theta)}{\partial \gamma_d}.$$

The shock vector $\partial g/\partial \gamma_d$ has strictly negative entries on the d–rows (links attached to d) and zeros elsewhere. Since $-J^{-1}$ is entrywise nonnegative and J is invertible, every link in the d–block satisfies

$$\frac{\partial p^{U*}_{du}}{\partial \gamma_d} \;<\; 0.$$

For any rival buyer $d' \neq d$, the corresponding links $L' = (d', u')$ receive shocks only through the matrix propagation $-J^{-1} \geq 0$ from the negative entries on the d–rows. Therefore

$$\frac{\partial p^{U*}_{d'u'}}{\partial \gamma_d} \;\leq\; 0,$$

with strict inequality whenever there is a path in the network from some $L = (d, u)$ to $L' = (d', u')$ (so that the corresponding entries of $-J^{-1}$ are strictly positive).

This establishes part (c) and completes the proof of Proposition 2. □

G Proof of Lemma 1 and Numerical Illustration

Proof of Lemma 1. With unit costs $c_d = (\tau_d/z_d)p_L$ and $c_{d'}$ fixed, the two-firm linear Cournot equilibrium gives $\Pi_d^D = (a - 2c_d + c_{d'})^2/(9b)$. Since $\Pi_{d,-u}^D = 0$, $\Delta\Pi_d^D = \Pi_d^D$. Setting $\pi \equiv a - 2c_d + c_{d'} > 0$:

$$\log \Delta\Pi_d^D = \text{const} + 2\log\pi.$$

Differentiating twice with respect to $\ln c_d$ (using $\partial\pi/\partial \ln c_d = -2c_d$):

$$\frac{\partial \log \Delta\Pi_d^D}{\partial \ln c_d} = \frac{-4c_d}{\pi}, \qquad \frac{\partial^2 \log \Delta\Pi_d^D}{\partial (\ln c_d)^2} = \frac{-4c_d}{\pi}\left(1 + \frac{2c_d}{\pi}\right) < 0.$$

The inequality holds since $c_d, \pi > 0$. Combined with equation (15) and $\partial \ln c_d/\partial \ln \tau_d = \partial \ln c_d/\partial \ln p_L = 1$, the claim follows. □

Numerical illustration. To make the mechanism concrete, consider the linear-Cournot benchmark with parameters $a = 10$, $b = 1$, and unit costs $c_{\text{NV}} = 2$, $c_{\text{AMD}} = 3$. The pre-shock Cournot margin for NVIDIA is $\pi \equiv a - 2c_{\text{NV}} + c_{\text{AMD}} = 9$ and profit $\Pi_{\text{NV}}^D = \pi^2/9 = 9$. The log-derivative of profit with respect to log-cost is $\partial \ln \Pi_{\text{NV}}^D/\partial \ln c_{\text{NV}} = -4c_{\text{NV}}/\pi = -8/9 \approx -0.889$. Now suppose the export wedge rises so that c_{NV} increases by 1% (holding p_L fixed momentarily). The Cournot margin compresses to $\pi' = 8.96$ and profit falls to $(8.96)^2/9 \approx 8.924$. The log-derivative becomes $-4 \cdot 2.02/8.96 \approx -0.902$: the buyer's marginal sensitivity to input costs has *increased in magnitude*. This is the log-submodularity condition in action — the wedge amplifies how much NVIDIA cares about each additional dollar of HBM cost. The second-order curvature $\partial^2 \ln \Pi^D/\partial(\ln c)^2 \approx -1.28$ (at baseline) is unambiguously negative, so the supplier lowers p_L to restore the Nash-product balance — the shock-absorber activates automatically.

Lemma 7 (Shock-absorber direction: nested-CES Cournot)**.** *Under $\sigma > \rho > 1$ and the*

full-incremental assumption $\Pi^D_{d,-u} = 0$,

$$\frac{\partial^2 \log \Delta\Pi^D_d}{\partial(\ln C^D_d)^2} \;=\; -C^D_d \,\frac{d\epsilon_d}{dC^D_d} \;<\; 0.$$

The shock-absorber direction therefore holds for all $\sigma > \rho > 1$*, irrespective of* η*.*

Proof. Step 1 ($d\epsilon_d/dC^D_d > 0$*).* From Lemma 2, $\epsilon_d = \sigma\rho/\big(\rho + s_d(\sigma-\rho)\big)$, which is strictly decreasing in s_d since $\sigma > \rho > 0$. By Corollary 2, s_d is strictly decreasing in C^D_d. Hence $d\epsilon_d/dC^D_d > 0$.

Step 2 (log-curvature formula). With $\Pi^D_{d,-u} = 0$, the Cournot markup gives $\Pi^D_d = C^D_d q^D_d/(\epsilon_d - 1)$; denote it π for short, with $C = C^D_d$ and $x = \ln C$. Since $d\pi/dx = C(d\pi/dC) = -Cq$ (envelope theorem, $d\pi/dC = -q$):

$$\frac{d^2 \log \pi}{dx^2} \;=\; \frac{\pi\,\frac{d^2\pi}{dx^2} - \left(\frac{d\pi}{dx}\right)^2}{\pi^2} \;=\; \frac{-C[(q + Cq_C)\pi + q^2 C]}{\pi^2}.$$

Differentiating the identity $\pi(\epsilon_d - 1) = Cq$ with respect to C yields $\pi'(\epsilon_d - 1) + \pi\epsilon'_d = q + Cq_C$, i.e. $-q(\epsilon_d - 1) + \pi\epsilon'_d = q + Cq_C$, so

$$(q + Cq_C)\pi + q^2 C \;=\; \Big[\pi\epsilon'_d - q(\epsilon_d - 1)\Big]\pi + q^2 C \;=\; \pi^2\epsilon'_d - \underbrace{q\pi(\epsilon_d - 1)}_{=Cq^2} + Cq^2 \;=\; \pi^2\,\frac{d\epsilon_d}{dC}.$$

Substituting gives $d^2 \log \pi/dx^2 = -C\, d\epsilon_d/dC < 0$ by Step 1. □

Remark 6. *The full-incremental assumption* $\Pi^D_{d,-u} = 0$ *(buyer has no outside option) is the same as in Lemma 1. In the calibrated multi-supplier model where* $\Pi^D_{d,-u} > 0$*, the second log-derivative generalises to* $-C\, d\epsilon_d/dC - \bar\pi\, q(1 - \epsilon_q)/\pi^2$ *(where* $\bar\pi = \Pi^D_{d,-u}$ *and* $\epsilon_q = -Cq_C/q$*), which is negative whenever* $\epsilon_q \geq 1$ *or whenever* $\bar\pi$ *is small relative to* π*. Numerically this sign is maintained at all six* $\sigma \geq 6$*,* $\eta \geq 2.5$ *grid points in Table 9.*

In the calibrated CES model the same logic applies, but the magnitude is tempered by the multi-supplier CES index and the Cournot block. Under Scenario A ($\tau_{\text{NV}} + 10\%$, $\tau_{\text{AMD}} + 5\%$)

at the baseline $(\sigma, \rho, \eta) = (6, 4, 2.5)$: SK hynix and Micron together lower NVIDIA's input-price index by $\%\Delta P^M_{\text{NV}} = -0.163\%$. Although small relative to the 10% statutory shock, this deflationary response is systematic and present at all six $\sigma \geq 6$, $\eta \geq 2.5$ grid points in Table 9. The ratio $|\%\Delta P^M_{\text{NV}}|/|\%\Delta \mu_{\text{NV}}| \approx 0.20$ reflects the dilution of the log-concavity signal through the multi-input CES nest; as η rises, the signal strengthens and the shock-absorber contribution grows.

H Extension to the Multi-Country Model

The following proof sketch verifies that the existence, uniqueness, and comparative-statics results of Propositions 1–2 extend to the multinational setting of Section 3.6.

(i) Cournot block. Firm d's profit aggregates plant-level revenues $\sum_{n'} \sum_{i \in \mathcal{I}_d} (p^D_{n',i,d} - C^D_{n',i,d}) q^D_{n',i,d}$, where each plant-specific unit cost $C^D_{n',i,d} = \kappa_{n'i} P^M_{i,d}/z_d$ inherits its CES materials index from the plant-pair links. Because $P^M_{i,d}$ is a constant-elasticity aggregate over upstream prices, the nested-CES curvature bounds in Lemma 2 apply plant-by-plant, and firm-level profit remains strictly concave in own output. The Cournot equilibrium exists, is unique, and has a diagonally dominant best-reply Jacobian for any fixed vector of plant-pair prices (Lemmas 2–3).

(ii) One plant-pair link. Fix all plant-pair prices except (i, j, d, u). Buyer d's incremental surplus $\Delta\Pi^D_d$ is the change in d's Cournot profit when link (i, j) is active vs. removed; seller u's incremental surplus $\Delta\Pi^U_u$ is the revenue on that link net of cost. The structure of each surplus is identical to the single-country baseline: $\Delta\Pi^D_d$ is strictly decreasing and (under Assumption 2(a)) log-concave in $p^U_{ij,du}$, and $\Delta\Pi^U_u$ is strictly log-concave. Lemmas 4–5 therefore deliver a unique interior optimizer for each plant-pair link in isolation.

(iii) Joint system. Stack all plant-pair FOCs into the pseudo-gradient g^{MC}, with Jacobian J^{MC}. Own-plant-pair entries J^{MC}_{LL} satisfy the same lower bound as in the single-country case. Cross-link entries couple only through $P^M_{i,d}$ within a buyer and through the Cournot

block across buyers; both remain bounded by the same constants (σ, ρ, η). The Gershgorin argument in Appendix E therefore applies without modification, yielding $J^{\mathrm{MC}} \in P$-matrix, and Rosen's theorem delivers a unique plant-pair price vector. The comparative-statics sign pattern follows from the same implicit-function argument, since $-[J^{\mathrm{MC}}]^{-1} \geq 0$ element-wise by the M-matrix property.

I Computational Algorithm

This appendix summarizes the numerical algorithm used to compute the bilateral oligopoly equilibrium described in Section 3. The goal is to solve for the Nash-in-Nash upstream price vector $\mathbf{p}^{U*}$ that jointly satisfies the link-level bargaining first-order conditions and the downstream Cournot best responses. The MATLAB implementation follows a nested structure: an inner Cournot solver that maps unit costs into quantities and profits, and an outer NiN solver that updates link prices until the bargaining conditions are satisfied.

Inner loop: Cournot block

Given a candidate vector of upstream prices $\mathbf{p}^U$, the code first computes downstream unit costs and the associated Cournot equilibrium.

(i) materials price index and unit costs. For each downstream firm d, the program computes the CES materials price index

$$P_d^M(\mathbf{p}^U) = \left(\sum_u \left(p_{du}^U \right)^{1-\eta} \right)^{\frac{1}{1-\eta}}$$

and the unit delivery cost

$$C_d^D(\mathbf{p}^U) = \frac{\tau_d}{z_d} P_d^M(\mathbf{p}^U).$$

These operations are vectorized over all active links (d, u) in the code.

(ii) Cournot quantities and profits. Taking $\mathbf{C}^D = (C_d^D)_d$ as given, the algorithm solves for the Cournot equilibrium quantities $\mathbf{q}^{D*}(\mathbf{C}^D)$ defined in Definition 1. In the MATLAB implementation, this is done by iterating best responses in log-quantities (a standard Gauss-Seidel or Jacobi update) until the maximum change in $\{q_d^D\}$ falls below a tolerance. At each iteration, firm d's best response is computed from its first-order condition under the nested CES demand system (1), holding rivals' quantities fixed.

Once convergence is reached, the code records, for each firm d, the equilibrium quantity q_d^{D*}, price p_d^D, and profit

$$\tilde{\Pi}_d^D(\mathbf{C}^D) = \left(p_d^D(\mathbf{q}^{D*}) - C_d^D\right) q_d^{D*},$$

as well as the link-level input demands q_{du}^U and upstream profits

$$\Pi_u^U(\mathbf{p}^U) = \sum_d \left(p_{du}^U - C_{du}^U\right) q_{du}^U(\mathbf{p}^U)$$

using the CES demand system (5). These objects are then passed to the outer NiN loop.

Outer loop: Nash-in-Nash over link prices

The outer loop treats the Cournot block as a black box that maps $\mathbf{p}^U$ into downstream and upstream surplus. It iteratively updates each link price to maximize the one-link log-Nash objective, holding all other prices fixed.

(i) Incremental surplus on a link. Fix a link $L = (d, u)$ and a current price vector $\mathbf{p}^U$. The code evaluates:

- the "with-link" outcome at $\mathbf{p}^U$ (all active links present);
- the "without-link" outcome at the same prices but with the (d, u) link severed, implemented by either (i) dropping the link from the CES aggregator, or equivalently (ii) setting p_{du}^U to a very large value so that $\omega_{du} \approx 0$ and $q_{du}^U \approx 0$.

Each evaluation calls the Cournot solver described above and returns the firm-level values Π_d^D and Π_u^U in the two scenarios. The incremental surplus terms are then

$$\Delta\Pi_d^D = \Pi_d^D - \Pi_{d,-u}^D, \qquad \Delta\Pi_u^U = \Pi_u^U - \Pi_{-d,u}^U.$$

These are functions of the single scalar p_{du}^U when all other link prices are held fixed.

(ii) One-dimensional link update. For link $L = (d, u)$, the algorithm considers the log-Nash objective

$$h_L(p_{du}^U) = \gamma_d \log \Delta\Pi_d^D + \Big(1 - \gamma_d\Big) \log \Delta\Pi_u^U,$$

with other link prices fixed at their current values. Under the curvature conditions in Assumption 2, this is a strictly concave function of p_{du}^U on the feasible interval where the link is active. The link update consists of finding the unique maximizer of h_L on this interval.

In the MATLAB implementation, this univariate problem is solved numerically by a safeguarded Newton or bisection routine. The code brackets the optimum between a low price (slightly above C_{du}^U) and a high price that drives the link share ω_{du} close to zero, and then iterates on the first-order condition $\partial h_L / \partial p_{du}^U = 0$ until the change in p_{du}^U is smaller than a tolerance. Because h_L is concave, this one-dimensional routine is stable and fast.

(iii) Cyclic best-response over links. The full NiN solver cycles through all active links $L \in \mathcal{L}$ in a predetermined order. Starting from an initial guess $\mathbf{p}^{U,(0)}$ (often a cost-plus markup), the algorithm:

- fixes all link prices except p_{du}^U;
- runs the inner Cournot solver to compute the incremental surplus for (d, u);
- updates p_{du}^U to the maximizer of h_L as in step (ii);
- moves to the next link and repeats.

One full pass over all links produces an updated price vector $\mathbf{p}^{U,(1)}$. The procedure continues for $k = 0, 1, 2, \ldots$ until the sup norm of the price change satisfies

$$\left\| \mathbf{p}^{U,(k+1)} - \mathbf{p}^{U,(k)} \right\|_{\infty} < \varepsilon_{\text{NiN}},$$

for a small tolerance ε_{NiN} (e.g. 10^{-8} in log-prices). Under the conditions in Proposition 1, these cyclic link-by-link updates implement a global fixed-point iteration on the log-Nash pseudo-gradient $g(\mathbf{p}^U)$ and converge to the unique NiN solution $\mathbf{p}^{U*}$.

Outputs and diagnostics

Once the price vector $\mathbf{p}^{U*}$ has converged, the algorithm performs a final Cournot solve to obtain the equilibrium quantities $\mathbf{q}^{D*}$, downstream prices $\mathbf{p}^{D*}$, and profits $\{\Pi_d^{D*}, \Pi_u^{U*}\}$. For the quantitative exercises in Section 4, the code also records:

- P_d^M and implied input-price component of downstream unit costs $\log P_d^M$;
- downstream markups and output prices;
- link-level expenditure shares ω_{du} and markups $p_{du}^{U*} - C_{du}^U$.

These objects are used to construct the decompositions and incidence measures reported in the counterfactual section. In practice, for the calibrated AI/HBM environment, the nested Cournot–NiN algorithm converges rapidly from a wide range of initial guesses, and the same equilibrium is selected across all experiments that hold the active set of links fixed.

J Permutation / Placebo Inference

With a single treated unit (Taiwan×MCP-IC), standard asymptotic inference based on clustered standard errors is unreliable. I implement a Fisher permutation test following Bertrand et al. (2004) and Conley and Taber (2011). Placebo units are all destination×product cells

with at least 24 non-missing monthly observations over the estimation window; the 53–54 units consist of 17–18 non-Taiwan destinations × 3 product codes (MCP-IC, DRAM, NAND). For each of these control units (country×HS-code cells), I re-assign treatment to that unit and compute the same test statistic: the average treatment effect (ATT), defined as the difference between the mean residualized outcome in the post-event window ($k = 0, \ldots, 12$) and the pre-event window ($k = -12, \ldots, -2$), after partialling out destination×month and item×month fixed effects and the AI demand control. The empirical p-value is the fraction of placebo ATTs with $|\widehat{\text{ATT}}_{\text{placebo}}| \geq |\widehat{\text{ATT}}_{\text{actual}}|$.

Table 13 reports the results. The October 2022 quantity contraction ($\widehat{\text{ATT}} = -1.09$ log points) is significant at the 5% level ($p = 0.019$): only 1 of 54 placebo units exhibits a post-event decline as large in magnitude. The export-value estimates for both events and the quantity estimate for October 2023 have p-values in the range $[0.13, 0.23]$, reflecting the wide confidence bands visible in Figure 5. These results are consistent with what is achievable with $N_{\text{placebo}} \approx 54$: the minimum attainable p-value is $1/54 \approx 0.019$. The October 2022 quantity contraction is at this minimum, indicating that the actual treated unit's response is the most extreme in the entire distribution. The remaining outcomes' larger p-values reflect their relatively wider cross-unit variation rather than an absence of policy effects; the corresponding event-study figures show visually sharp responses even where the permutation test is underpowered.

Table 13: Permutation / placebo inference for the single treated unit

Event	Outcome	ATT†	N_{placebo}	p-value
Oct 2022	ln(export value)	−0.574	54	0.222
Oct 2022	ln(quantity)	−1.088	54	0.019
Oct 2023	ln(export value)	+0.624	53	0.132
Oct 2023	ln(quantity)	+0.574	53	0.226

Notes: Fisher permutation test for a single treated unit (Taiwan×MCP-IC). Each of the remaining (country×HS-code) cells serves as a placebo treated unit ($N_{\text{placebo}} \in \{53, 54\}$). ATT†: mean residualized outcome in the post-event window ($k = 0, \ldots, 12$) minus the pre-event window ($k = -12, \ldots, -2$), after partialling out destination×month and item×month fixed effects and the AI demand control. The p-value is the fraction of placebo ATTs with $|\widehat{\text{ATT}}_{\text{placebo}}| \geq |\widehat{\text{ATT}}_{\text{actual}}|$.

Figure 15 plots the full placebo distribution for each event–outcome combination. Each

histogram shows the ATT computed for the 53–54 control units; the dashed vertical line marks the actual Taiwan×MCP-IC estimate. The October 2022 quantity contraction (bottom-left panel) lies at the far left tail of the distribution — no control unit exhibits a decline as large — confirming that this response is uniquely extreme. For the remaining three panels the actual estimate is visible but sits closer to the center of the placebo distribution, consistent with the p-values in Table 13.

K Robustness: Extensive Margin Recalibration

As discussed in Section 4, adding a Samsung–NVIDIA link to the baseline model mechanically lowers NVIDIA's input costs and raises its market share purely through an efficiency effect (love-of-variety in the CES input bundle). This creates a potential confounding factor when comparing the incidence of trade shocks across the "link" and "no-link" economies: NVIDIA is effectively a larger, lower-cost firm in the "with-link" counterfactual.

To ensure the non-neutrality results in Figure 7 are driven by network structure and bargaining options rather than baseline level differences, I implement a recalibration procedure. This procedure standardizes the pre-shock market structure across the extensive-margin experiments.

K.1 Recalibration Algorithm

Let $\mathcal{E}_{\text{base}}$ be the calibrated baseline economy (no Samsung–NVIDIA link) and $\mathcal{E}_{\text{link}}$ be the counterfactual economy with the link active. The recalibration proceeds as follows:

1. **Target Baseline Shares.** I treat the equilibrium market shares from $\mathcal{E}_{\text{base}}$ as the target vector $\mathbf{s}^D_{\text{target}}$. In the main calibration, these are approximately $s_{\text{NVIDIA}} \approx 85\%$ and $s_{\text{AMD}} \approx 15\%$.

2. **Activate Link.** I update the adjacency matrix in $\mathcal{E}_{\text{link}}$ to include the ($d = \text{NVIDIA}, u = \text{Samsung}$) pair.

3. **Invert for Productivities.** I numerically invert the equilibrium mapping to find a new downstream productivity vector $\boldsymbol{z}'_d$ such that the equilibrium shares in $\mathcal{E}_{\text{link}}$ match $\mathsf{s}^D_{\text{target}}$. The algorithm solves the nonlinear least-squares problem:

$$\min_{\theta} \left\| \log\left(\frac{\hat{s}_{\text{NVIDIA}}(\boldsymbol{z}_d(\theta))}{\hat{s}_{\text{AMD}}(\boldsymbol{z}_d(\theta))} \right) - \log\left(\frac{s^{\text{target}}_{\text{NVIDIA}}}{s^{\text{target}}_{\text{AMD}}} \right) \right\|^2, \tag{20}$$

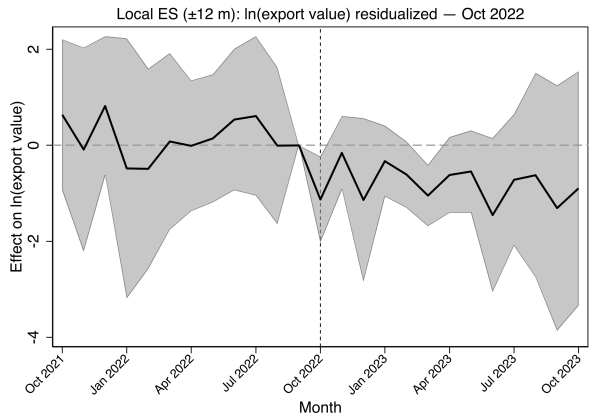
where θ parameterizes NVIDIA's productivity (holding AMD's normalized to 1) and $\boldsymbol{z}_u$ is held fixed. This is solved via a Levenberg-Marquardt routine.

4. **Simulate Shocks.** The trade shocks defined in Appendix D are then applied to this recalibrated economy $\mathcal{E}'_{\text{link}}$.

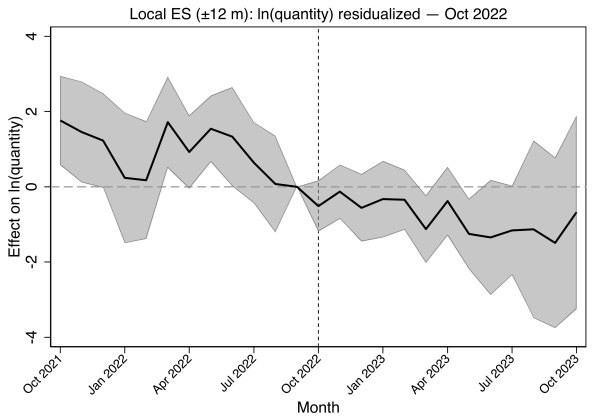
K.2 Robustness of Incidence Results

Intuitively, this procedure "soaks up" the efficiency gain of the new link into a lower calibrated productivity for NVIDIA ($z'_{\text{NVIDIA}} < z_{\text{NVIDIA}}$), such that the observable market structure looks identical to the baseline before the trade shock hits.

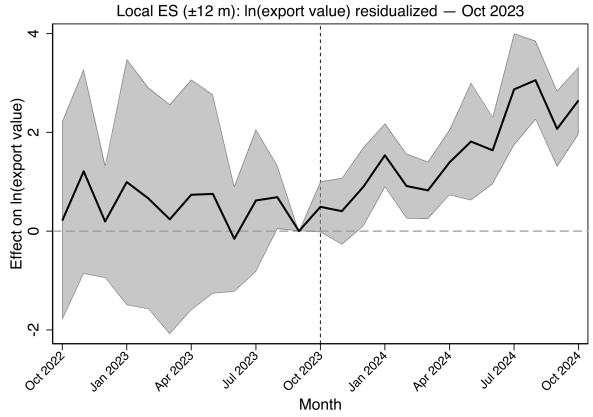
The results in figure 16 confirm the non-neutrality result: even when starting from identical market shares, the presence of the additional upstream link dampens the pass-through of NVIDIA-specific shocks into NVIDIA's markups and input prices. This confirms that the attenuation mechanism reported in Section 4 is driven by the change in bargaining outside options and network topology, not merely by the lower baseline cost of the bundle.



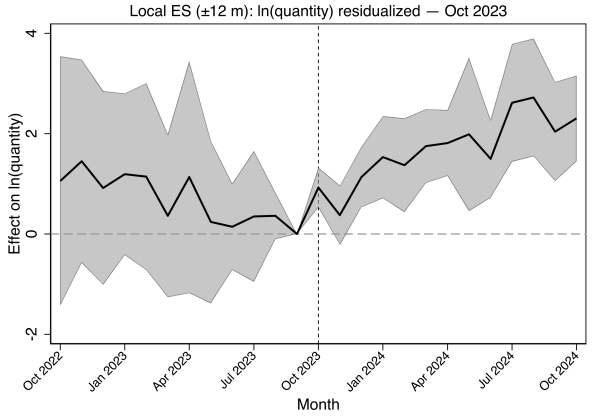
(a) Local event study (Oct 2022): ln(export value).



(b) Local event study (Oct 2022): ln(quantity).

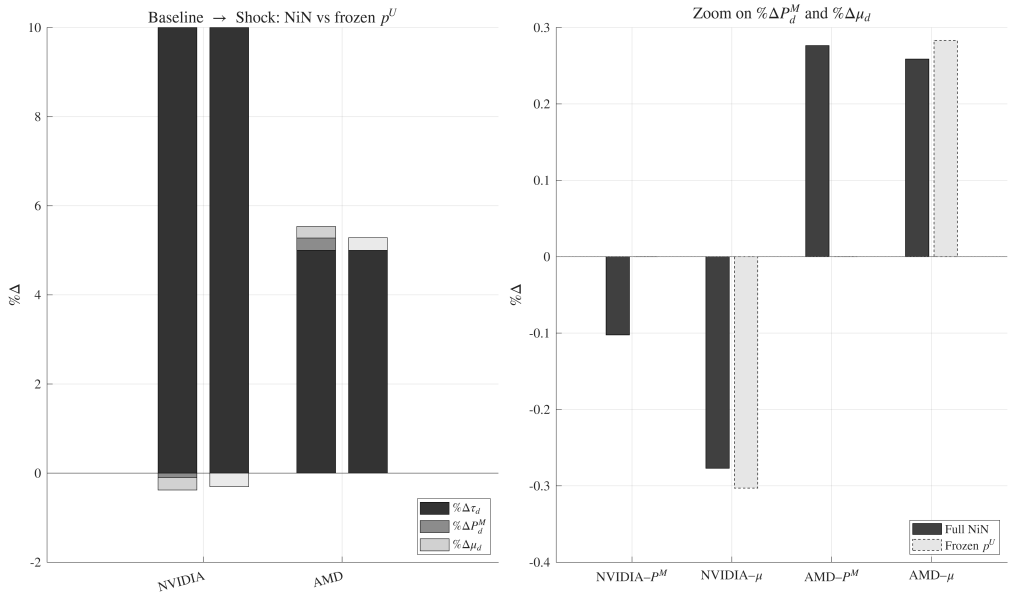


(c) Local event study (Oct 2023): ln(export value).

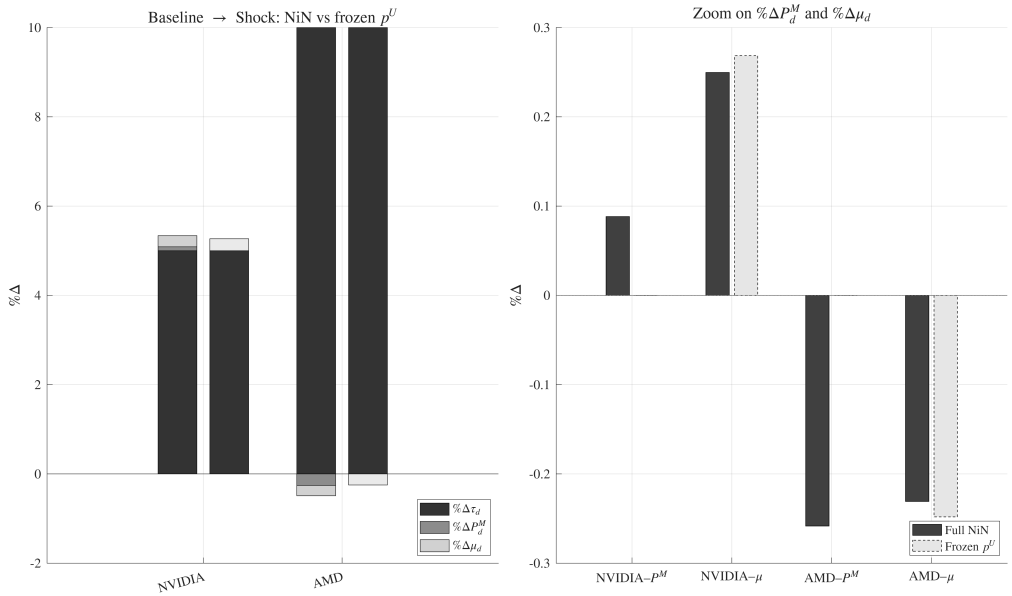


(d) Local event study (Oct 2023): ln(quantity).

Figure 5: Local ±12 month event studies for Taiwan×MCP–IC. Outcomes are residualized for unit fixed effects, year-month fixed effects, unit trends, and unit-specific seasonality. Regressions include destination-by-month and item-by-month fixed effects; standard errors are clustered by unit, and observations are inverse-value weighted (to prevent the Taiwan×MCP–IC cell — the highest-value cell in the panel — from mechanically driving the comparison; Appendix **??** reports unweighted estimates as a robustness check). Pre-period confidence bands are wide because there is a single treated unit (Taiwan×MCP–IC). Pre-period coefficients are non-zero for several outcomes, reflecting anticipatory behavior and technology-cycle dynamics documented in Appendix C.1; the figures are best read as descriptive dynamic profiles of the Taiwan–MCP–IC cell through each policy window rather than as parallel-trends tests. Stacked windows and restricted-sample designs are reported in Appendix C.2 and C.3.

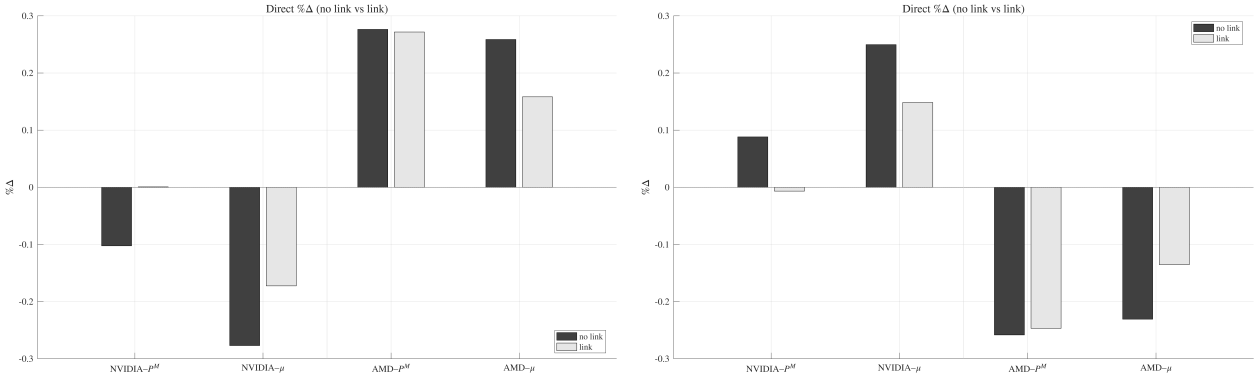


(a) NVIDIA +10%, AMD +5%



(b) NVIDIA +5%, AMD +10%

Figure 6: Price decomposition for heterogeneous accelerator wedges. Each panel reports, for NVIDIA and AMD, the decomposition of the change in accelerator prices into (i) the statutory trade wedge $\%\Delta\tau_d$, (ii) the induced change in the materials price index $\%\Delta P_d^M$, and (iii) the markup component $\%\Delta\mu_d$, comparing the full Nash-in-Nash equilibrium to a benchmark with upstream prices frozen at baseline.



(a) NVIDIA +10%, AMD +5%

(b) NVIDIA +5%, AMD +10%

Figure 7: Decomposition non-neutrality: full Nash-in-Nash equilibria with and without a Samsung–NVIDIA link. Each bar reports the change in the materials price index or markup for NVIDIA and AMD when moving from the pre-shock to post-shock equilibrium, once without the Samsung–NVIDIA link (dark bar) and once with the link active (light bar).

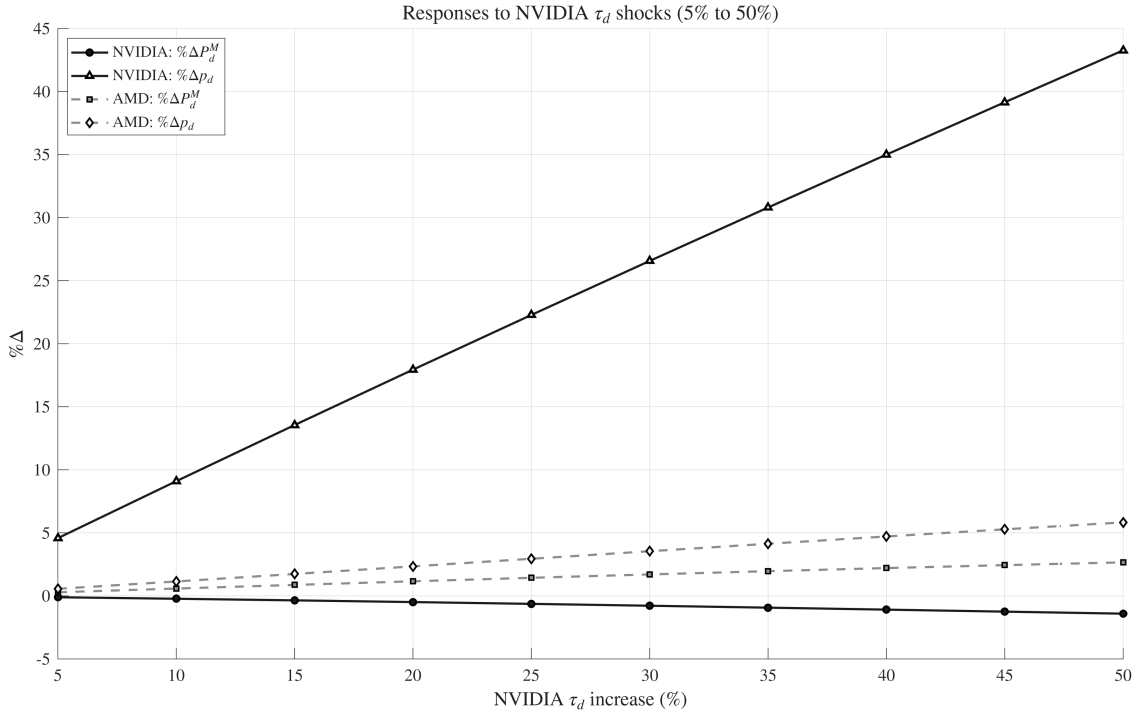
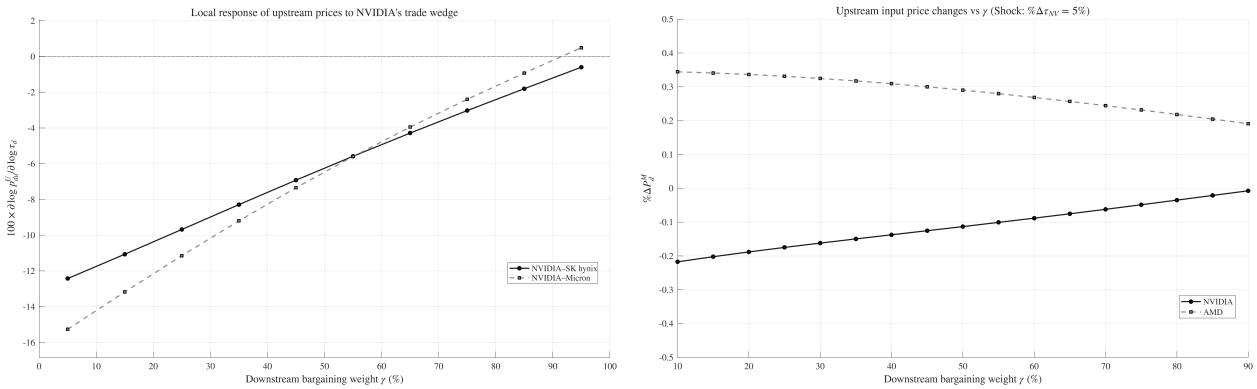


Figure 8: Responses of accelerator input costs and prices to NVIDIA's trade wedge. The figure plots the percent change in the materials price index P_d^M and the downstream price p_d for NVIDIA and AMD as NVIDIA's wedge τ_{NV} increases from 5 to 50 percent, holding AMD's wedge fixed. The y-axis unit is percent change from the pre-shock equilibrium.



(a) Local upstream price incidence **(b)** Input price changes by buyer

Figure 9: Sensitivity of upstream incidence to downstream bargaining weight γ_{NV}. Panel (a) shows the local derivative $100 \times \partial \log p^U_{\text{NV},u} / \partial \log \tau_{\text{NV}}$ for NVIDIA's links to SK hynix and Micron; the derivative changes sign at very high γ_{NV} (above approximately 85%), marking the limit of the shock-absorber mechanism. Panel (b) shows the resulting percent change in the materials price index P^M_d for NVIDIA and AMD under a 5 percent NVIDIA wedge shock.

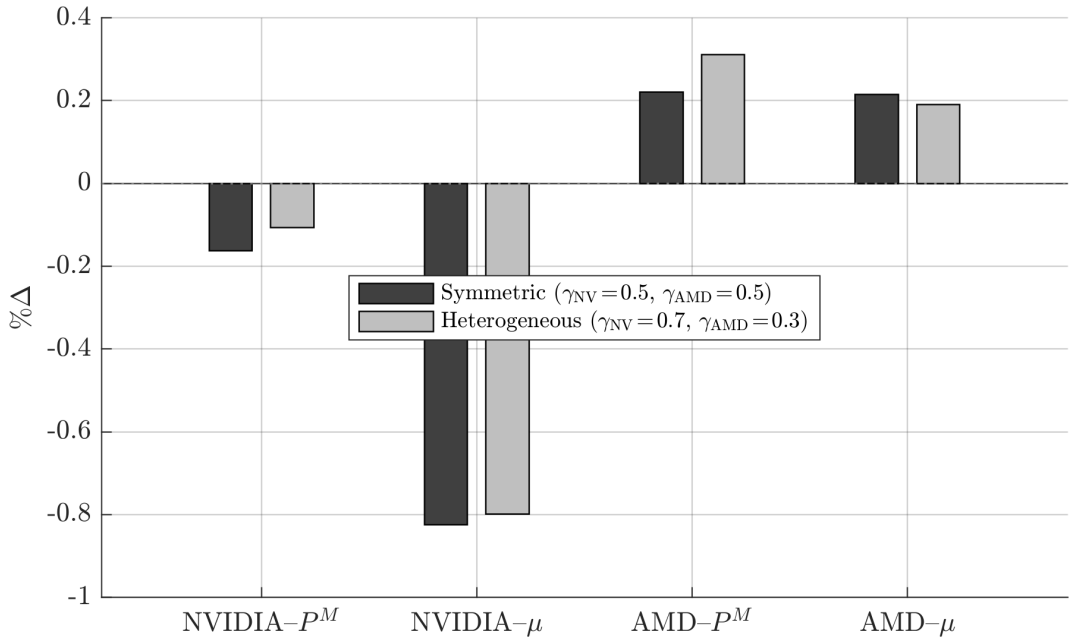
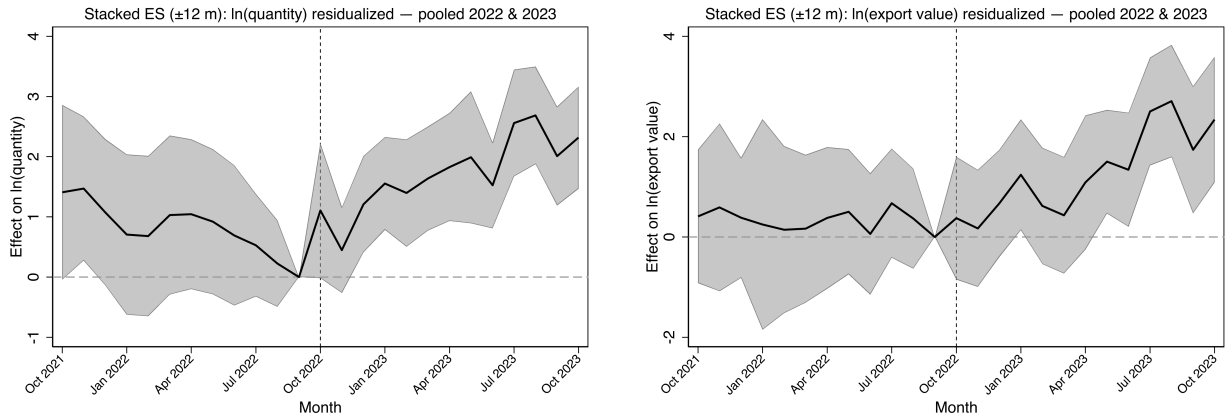


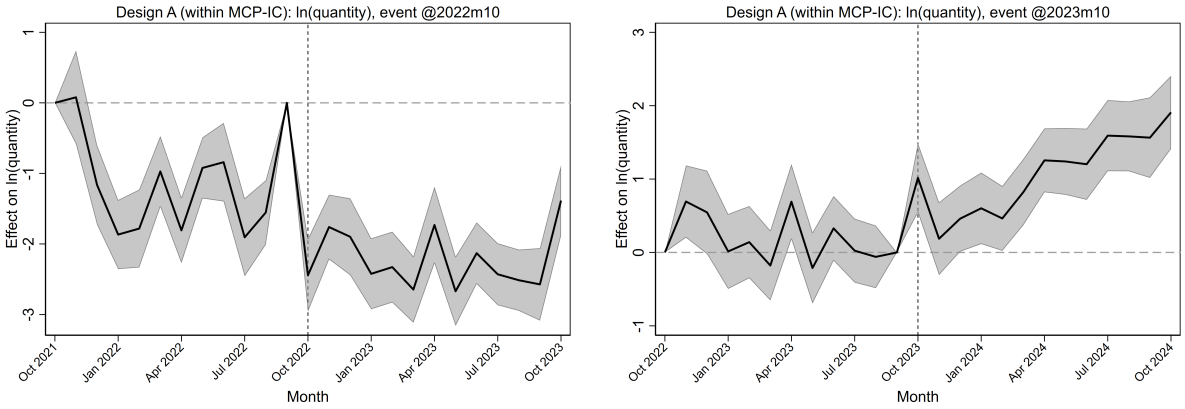
Figure 10: Price decomposition under symmetric vs. heterogeneous bargaining weights (τ_{NV}+10%, τ_{AMD} + 5%). Dark bars: symmetric $\gamma_{\text{NV}} = \gamma_{\text{AMD}} = 0.5$. Light bars: heterogeneous $\gamma_{\text{NV}} = 0.7$, $\gamma_{\text{AMD}} = 0.3$. Productivities are re-calibrated separately for each $\boldsymbol{\gamma}$ configuration. The qualitative sign pattern (NVIDIA shock absorption, AMD cross-buyer spillover) is preserved across configurations.



(a) Stacked ES: ln(quantity) residualized—pooled 2022 & 2023.

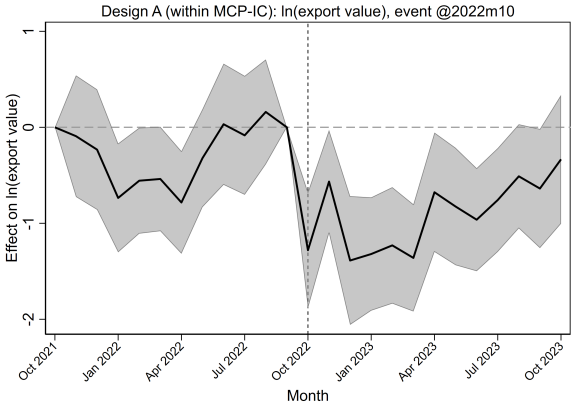
(b) Stacked ES: ln(export value) residualized—pooled 2022 & 2023.

Figure 11: Stacked local event studies (±12 months) around Oct 2022 and Oct 2023. For each event, I create a separate panel and estimate with unit-by-event fixed effects and destination-by-month-by-event and item-by-month-by-event fixed effects. Outcomes are residualized for unit trends, unit-specific seasonality, and an interacted AI-market proxy. Shaded bands denote 95% CIs.

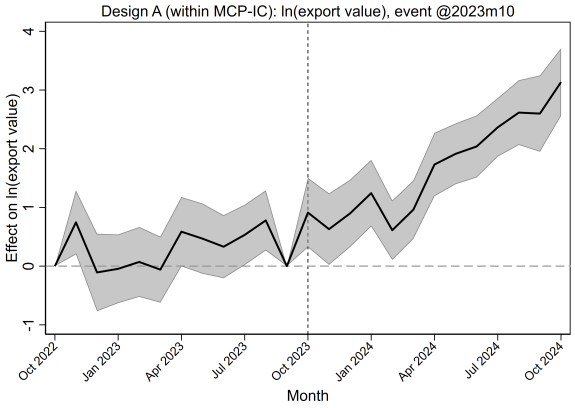


(a) ln(quantity), event at 2022m10

(b) ln(quantity), event at 2023m10

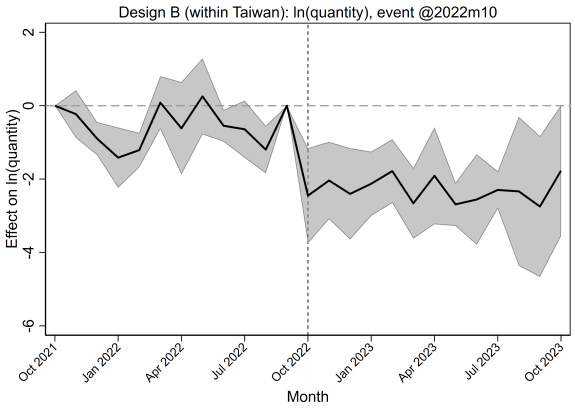


(c) ln(export value), event at 2022m10

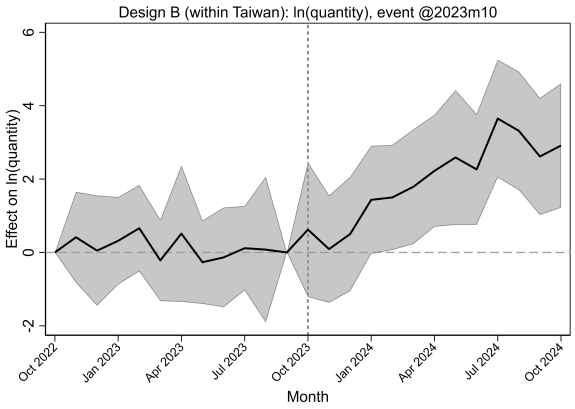


(d) ln(export value), event at 2023m10

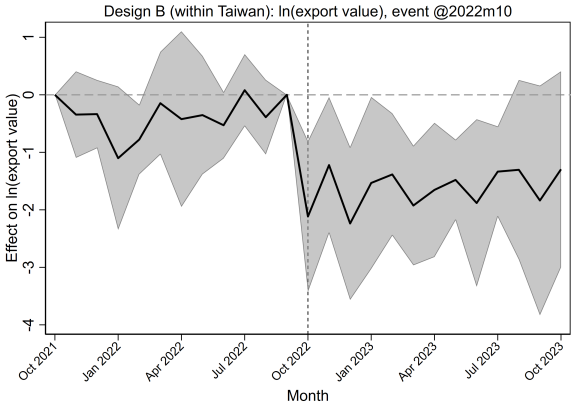
Figure 12: Event studies with restricted controls, Design A (within MCP–IC across destinations). The treated series is MCP–IC exports to Taiwan; donors are MCP–IC exports to other destinations. Outcomes are residualized as described in the text. Shaded bands show 95% confidence intervals.



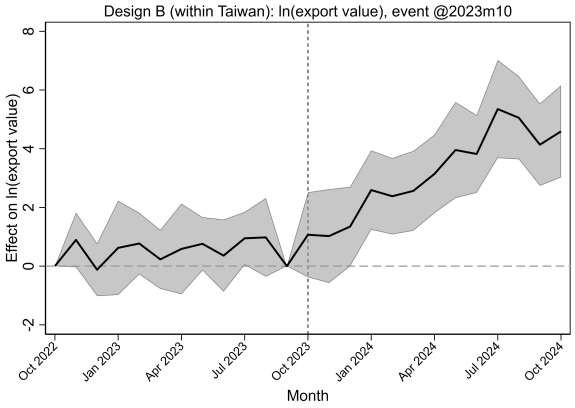
(a) ln(quantity), event at 2022m10



(b) ln(quantity), event at 2023m10

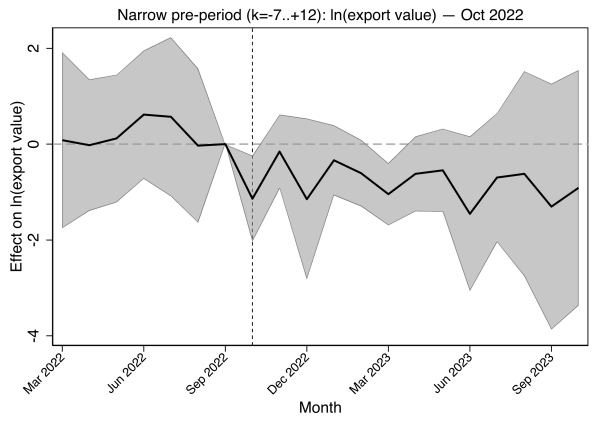


(c) ln(export value), event at 2022m10

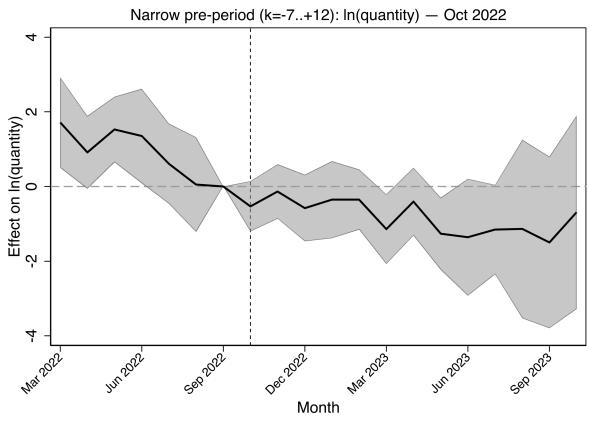


(d) ln(export value), event at 2023m10

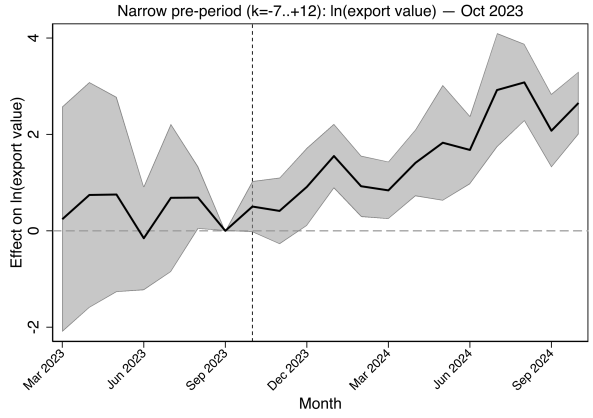
Figure 13: Event studies with restricted controls, Design B (within Taiwan across items). The treated series is MCP–IC exports to Taiwan; donors are DRAM and NAND exports to Taiwan. Outcomes are residualized as described in the text. Shaded bands show 95% confidence intervals.



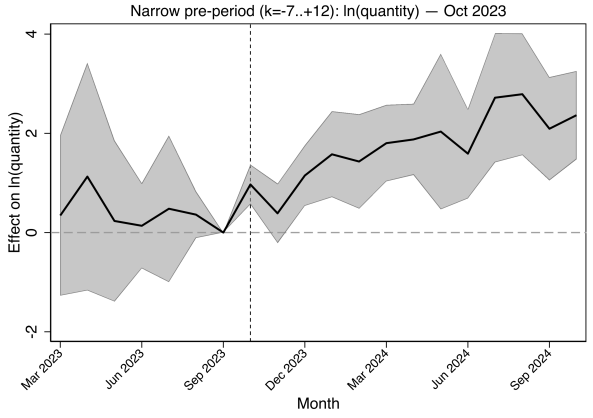
(a) Narrow window, Oct 2022: ln(export value).



(b) Narrow window, Oct 2022: ln(quantity).

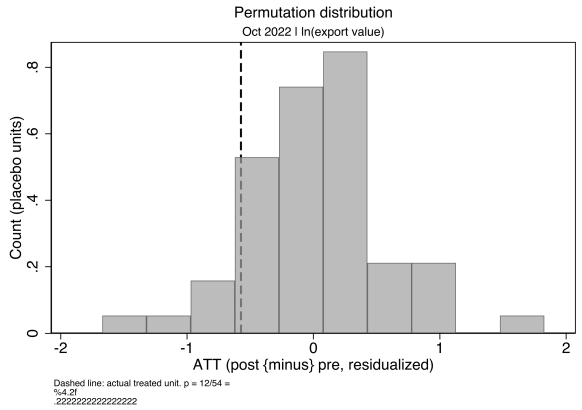


(c) Narrow window, Oct 2023: ln(export value).

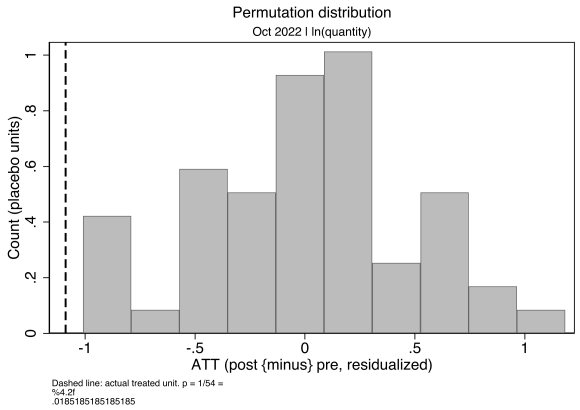


(d) Narrow window, Oct 2023: ln(quantity).

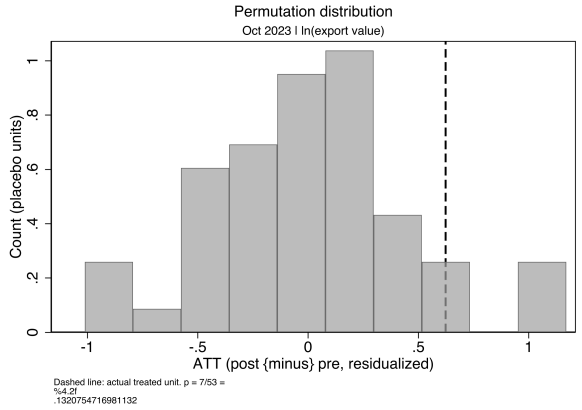
Figure 14: Event studies with narrow pre-period ($k = -7$ to $+12$). Specification identical to Figure 5 except the window extends only 7 months before each event, excluding the HBM3 development ramp months ($k \leq -8$). Post-event dynamics are indistinguishable from the full-window baseline.



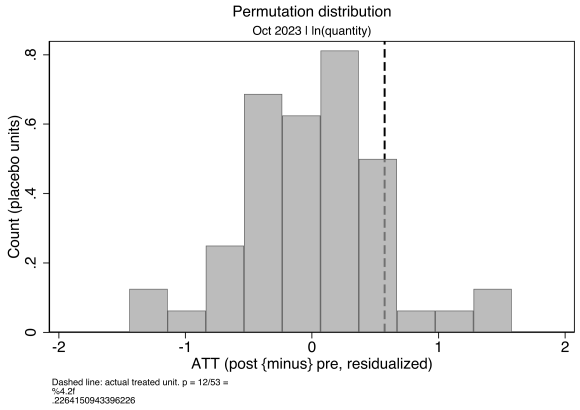
(a) Oct 2022 – Export value



(b) Oct 2022 – Quantity

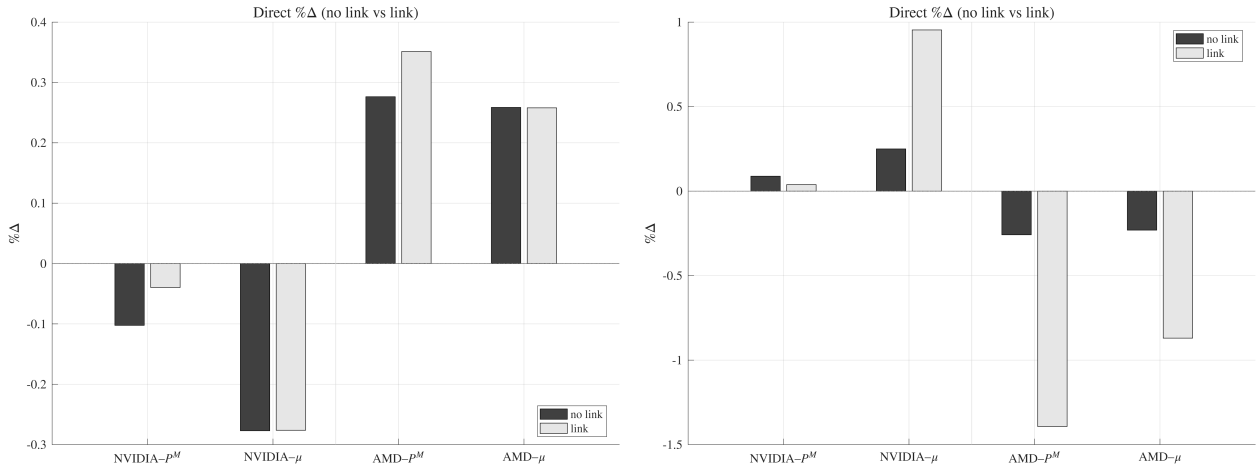


(c) Oct 2023 – Export value



(d) Oct 2023 – Quantity

Figure 15: Placebo distributions for Fisher permutation test. Each panel shows a histogram of the ATT computed for the 53–54 control (country×HS-code) units. The dashed vertical line marks the actual Taiwan×MCP-IC ATT. The p-value is the fraction of placebo units with $|\widehat{\mathrm{ATT}}_{\mathrm{placebo}}| \geq |\widehat{\mathrm{ATT}}_{\mathrm{actual}}|$.



(a) NVIDIA +10%, AMD +5%

(b) NVIDIA +5%, AMD +10%

Figure 16: This figure repeats the counterfactual from Figure 7, but recalibrates baseline productivity so that pre-shock market shares are identical in the link and no-link economies.