Endogenous Markups and Trade Elasticities

Anton Cheremukhin and Paulina Restrepo-Echavarria

Working Paper 2632 **September 2026**

Research Department

https://doi.org/10.24149/wp2632

Endogenous Markups and Trade Elasticities*

Anton Cheremukhin† and Paulina Restrepo-Echavarria‡

August 2026

Abstract

We develop a search model of international trade where buyers allocate costly cognitive attention across suppliers. Suppliers exploit the cost of shifting attention to extract markups, giving dominant firms the cushion to absorb cost shocks while marginal suppliers pass them on. Validating this mechanism with US tariff data, we estimate a highly concentrated US domestic market. Consequently, while targeted tariffs spur offshore substitution, blanket tariffs force buyers toward captive domestic monopolies. Incumbents' growing market power raises markups and halves the aggregate macro elasticity relative to micro estimates. Integrating endogenous market-power-dependent elasticity along the path to autarky more than doubles the US gains from trade compared with standard constant-elasticity predictions.

Keywords: Trade Elasticity, Targeted Search, Rational Inattention, Gravity, Market Power, Pass-Through, Gains from Trade.

JEL Classification: F10, F11, F13, D83, L13.

* We thank Klaus Desmet, Christian Hellwig, and Fernando Leibovici, among others, for their insightful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Dallas or St. Louis, or the Federal Reserve System.

† Anton Cheremukhin, Federal Reserve Bank of Dallas, 2200 N Pearl St, Dallas TX 75201, chertosha@gmail.com.

‡ Paulina Restrepo-Echavarria, Federal Reserve Bank of St. Louis, One Federal Reserve Bank Plaza, St. Louis MO 63166, paulinares@me.com.

1 Introduction

Global trade policy has shifted over the last decade, moving from the targeted interventions of the late 2010s to the broad baseline tariffs of the mid-2020s. Empirical evaluations of these episodes reveal a complex, non-linear anatomy of trade wars. Targeted tariffs on specific origins were passed through almost fully to border prices, yet produced relatively muted effects on aggregate US consumer prices due to retail margin compression and rapid offshore substitution (Amiti et al., 2019b; Fajgelbaum et al., 2020; Cavallo et al., 2021; Fajgelbaum and Khandelwal, 2022). In contrast, recent blanket tariffs exerted much stronger, gradually accumulating upward pressure on domestic price levels (Cavallo et al., 2025; Hacıoğlu-Hoke et al., 2026; Minton et al., 2026), with border pass-through above 90 percent (Gopinath and Neiman, 2026) and a sizable indirect component from domestic producers raising prices as import competition weakened (Amiti et al., 2026). Furthermore, tariff pass-through is highly state-dependent: small tariffs are often absorbed by foreign suppliers, while larger interventions bite consumers heavily. Standard quantitative models struggle to explain these distinct macroeconomic outcomes because they typically rely on a constant elasticity of substitution (CES), which assumes that tariff pass-through is proportional and invariant to the scale or breadth of the shock.

The empirical flaw in this rigid assumption is also exposed by a pervasive puzzle in international macroeconomics. In the data, when buyers substitute across a long tail of small foreign varieties, they exhibit a high microeconomic trade elasticity, with a cross-country median of roughly 5.4. However, when an entire economy faces a uniform macroeconomic shock, the aggregate trade elasticity systematically collapses, to a median of roughly 3 and to about 1.5 for the United States (Imbs and Méjean, 2017). Standard models cannot seamlessly reconcile this attenuation, forcing researchers to artificially choose between micro and macro parameters depending on the application.

To resolve this, we develop a search model of international trade where buyers allocate scarce cognitive attention across potential suppliers. Instead of assuming that buyers inherently love variety, we model sourcing as a costly information problem using rational inattention. Exploring and maintaining global suppliers consumes real cognitive and physical resources. Because shifting this attention is costly, buyers naturally concentrate their sourcing, effectively becoming captive to their primary suppliers. Exporting firms recognize this cognitive friction. Rather than dropping prices to completely dominate a market, a supplier that has captured a large share of a buyer's attention optimally extracts a relationship-specific markup. Consequently, incomplete specialization arises naturally from this strategic pricing behavior rather than a mechanical preference for dispersion.

This attention-based theory provides a unified micro-foundation that spans the two dominant paradigms in quantitative trade. For a marginal supplier with a near-zero market share, demand is highly elastic. This supplier prices like a fragmented CES firm, fully passing cost shocks to the buyer. Conversely, a dominant supplier that has captured most of a buyer's attention faces heavily inelastic demand. It reprices like a large oligopolist in the variable-markup framework of Atkeson and Burstein (2008), absorbing shocks into its markup cushion. Whether a trading relationship behaves like a competitive CES variety or a powerful oligopolist is an endogenous state governed by its market share.

We validate this mechanism using US customs data and the tariff variation of the 2018 to 2024 trade wars. Consistent with the model, marginal foreign suppliers with negligible market shares pass tariffs through to the US border almost entirely. However, dominant foreign suppliers with large pre-war sourcing shares absorb the vast majority of the tariff by compressing their own markups. This validates share-dependent absorption; the model further implies a non-linear response to tariff scale: small tariffs eat into the markup cushion of dominant suppliers, whereas large tariffs exhaust the cushion and force steep pass-through.

To quantify the pricing power of domestic firms, we rely on the observed cross-country gap between micro and macro trade elasticities. Because a dispersed, competitive market fails to reproduce this severe attenuation, the large empirical gap allows us to directly infer the degree of domestic supplier concentration required by the model. We find that the US domestic market is heavily dominated by established incumbents, operating much closer to the captive variable-markup regime.

This domestic asymmetry explains the diverging anatomy of recent trade wars. When policymakers levy a targeted tariff, buyers simply substitute toward un-tariffed offshore competitors. Domestic incumbents experience little captive surge in demand, keeping domestic markups nearly flat and consumer inflation muted. A blanket baseline tariff shuts down this foreign escape valve. US buyers are forced to redirect expenditure homeward into a captive domestic base. Rather than smoothly absorbing the new volume at constant prices, domestic incumbents aggressively hike their markups. This cross-price response halves the aggregate macro elasticity and inflicts a dual penalty on consumers: they pay the tariff on foreign goods and elevated monopoly rents on domestic goods.

Finally, because trade in our model is governed by slow-moving geographic awareness priors, the horizon of adjustment differs drastically depending on the policy. The price disruptions of targeted tariffs eventually wash out as supply chains successfully reorganize offshore. In contrast, blanket tariffs cause buyers to slowly forget offshore alternatives, allowing domestic markups to accumulate and lock in over time. By permanently entrenching domestic monopoly power, broad trade closures carry a staggering long-run cost. When we evaluate the ultimate limit of these blanket policies, the true welfare cost of autarky reaches 11.7 percent, more than double the 5.2 percent implied by the constant-elasticity formula of Arkolakis et al. (2012).

The paper connects to several strands of the literature. First, we build on the application of rational inattention and entropy-regularized choice to economic theory. Sims (2003) introduces rational inattention, and Matějka and McKay (2015) and Fosgerau et al. (2020) establish its equivalence with logit discrete choice. We apply these tools to international trade networks. Previous work studies information and search frictions in agricultural trade (Allen, 2014), supplier search in production networks (Bernard et al., 2019), and search and contracting in importing (Startz, 2021). Recent work emphasizes endogenous network formation: Arkolakis et al. (2025) model search across buyers and suppliers, finding that endogenous links amplify the aggregate trade elasticity, while Huneeus (2018) shows these firm-to-firm links adjust sluggishly to shocks. We contribute to this literature by formalizing how the cognitive cost of relationship maintenance generates a concentrated network structure and endogenous market power. Whereas the extensive margin of network search amplifies the aggregate trade elasticity, our intensive-margin domestic repricing mechanism severely attenuates it. The model adapts the information-based sequential-contracting framework developed by Cheremukhin et al. (2020) and Cheremukhin and Restrepo-Echavarria (2026) to the international sourcing environment.

Second, the paper provides a resolution to persistent empirical puzzles regarding trade elasticities. Imbs and Méjean (2015, 2017) document that elasticities estimated from aggregate data are severely attenuated relative to sector-level estimates, and Feenstra et al. (2018) find the home–import elasticity below the micro elasticity for a sizable share of US goods. Ruhl (2008) names a companion puzzle in the time dimension:

business-cycle price movements imply low elasticities, while trade liberalizations imply large ones. Drozd and Nosal (2012) and Kehoe and Ruhl (2013) show this short-run versus long-run gap arises from slow customer accumulation and new-goods margins. Our framework provides an informational micro-foundation for this sluggishness and integrates the resulting pricing-to-market dynamics. Furthermore, Simonovska and Waugh (2014) demonstrate that even the baseline level of this elasticity is uncertain in a welfare-critical way. Our model resolves the cross-sectional discrepancy as a natural byproduct of simultaneous incumbent repricing and domestic concentration, while Section 3.4 addresses the time dimension through slow awareness updating.

Third, we offer an alternative foundation for variable markups that fundamentally alters how pro-competitive effects enter the gains from trade. Following Krugman (1987), models relying on demand curvature or oligopoly (Atkeson and Burstein, 2008; Melitz and Ottaviano, 2008; Arkolakis et al., 2019) generate markups that vary with market share or competitive toughness. A quantitative literature evaluates the resulting pro-competitive gains: for instance, Edmond et al. (2015) and Hsu et al. (2020) find positive pro-competitive additions, Holmes et al. (2014) show that the sign is ambiguous, and Arkolakis et al. (2019) find it slightly negative. However, these models typically apply a static misallocation or markup-level correction to a baseline calculation where the trade elasticity is held fixed, and where the elasticity does vary, as with endogenous firm selection (Melitz and Redding, 2015), it varies through selection and the level of trade costs, a distinct channel from our scope-of-shock repricing mechanism. In our framework, pricing is derived directly from the resource cost of sourcing. Consequently, the object that changes is the aggregate trade elasticity itself, which becomes state-dependent and is integrated along the path to autarky. Our mechanism also provides a theoretical foundation for the heterogeneous, market-share-dependent pass-through of shocks (Amiti et al., 2019a; Cavallo et al., 2021; Heise, 2024) and bilateral dependence in firm-to-firm pricing (Dhyne et al., 2022; Alviarez et al., 2023) observed in recent micro-data.

Finally, we contribute to the quantitative gravity and welfare literature (Eaton and Kortum, 2002; Anderson and van Wincoop, 2003). While Arkolakis et al. (2012) demonstrated that gains can be evaluated using a sufficient statistic based on a constant trade elasticity, subsequent work shows this arithmetic is fragile. Different microstructures map the same micro price information into different trade-elasticity estimates, with large welfare consequences (Simonovska and Waugh, 2026), capital accumulation adds dynamic gains (Ravikumar et al., 2019), slow exporter-capacity accumulation causes the elasticity to rise with the horizon, so that static statistics misstate the gains of reform (Alessandria et al., 2021), and anticipated policy dynamics bias reduced-form elasticity estimates (Alessandria et al., 2025). We add a distinct, static channel: because market power is tied to the attention portfolio, the relevant macro elasticity is state-dependent, and integrating this schedule alters the sufficient-statistic calculation even at a point in time.

The paper is organized as follows. Section 2 presents the model and the pricing equilibrium. Section 3 studies the limiting cases, pass-through, and aggregation. Section 4 presents the evidence and identification strategy, and evaluates the welfare calculation and policy implications. Section 5 concludes.

2 The Model

We develop a model of international trade where buyers search for sellers across borders.[1] The economy features sequential contracting and information costs. Buyers allocate costly attention across potential suppliers. Sellers post prices and accommodate incoming inquiries.

2.1 The economic environment

The model describes the market for a single good in a world of N countries. Each country $x \in \{1, \ldots, N\}$ has a measure μ_x of identical buyers, each demanding one unit of the good, so we use x for both the buyer type and location. Sellers are indexed by $j \in J$, each located in a country $i(j) \in \{1, \ldots, N\}$. A country may contain several sellers. Each seller j is a single firm with μ_j units of the good to sell. Buyer x sources from a fixed set of sellers $J_x^+ \subseteq J$ that includes sellers at home and abroad.

A match between buyer x and seller j generates a valuation v_{xj} per unit. The seller produces the good at unit cost c_j, and delivering it to buyer x incurs an iceberg cost $\tau_{xj} \geq 1$ that depends only on the pair of countries, with $\tau_{xj} = 1$ when $i(j) = x$. The joint surplus of a realized match is thus $v_{xj} - \tau_{xj} c_j$.

[1] We adapt the theoretical setup and notation from the sequential matching framework of Cheremukhin and Restrepo-Echavarria (2026).

Every seller posts a buyer-type-specific unit price P_{xj} (there can be price discrimination). When a match forms, the seller delivers at that price, so the buyer receives the net surplus $v_{xj} - P_{xj}$, and the seller earns the profit $P_{xj} - \tau_{xj} c_j$.

Assumption 1 (Voluntary Trade)**.** Once a match forms, either party may decline to trade. An agent that declines, or that remains unmatched, receives an outside option normalized to zero.

Assumption 1 pins down the range of prices that can arise. A buyer refuses any price above its valuation, so a seller posting $P_{xj} > v_{xj}$ attracts attention but sells nothing, and a seller refuses to deliver below its delivered cost. On every active pair, therefore, $\tau_{xj} c_j \leq P_{xj} \leq v_{xj}$, and both parties' payoffs are non-negative; we take this interval as the seller's action space without loss of generality.

The game has three stages. First, every seller posts its prices. Second, buyers observe all posted prices and allocate their sourcing attention across the sellers in J_x^+. Third, buyers contact sellers, and matches are realized according to an aggregate matching technology.

2.2 Information costs and rational inattention

Evaluating foreign suppliers requires cognitive attention, technical due diligence, and relationship management. We model this friction using entropy-regularized targeted search.

Buyers choose a discrete probability distribution of attention over sellers, λ_{xj}, and we measure the buyer's information acquisition cost as the relative entropy (Kullback–Leibler divergence) between its chosen probability distribution (attention strategy) λ_{xj} and a default awareness prior $\tilde{a}_{xj}$. The awareness prior represents the baseline probability that a buyer encounters a seller in the absence of active search. It captures the influence of physical distance, shared borders, historical ties, and general brand visibility. The buyer's total information processing effort is:

$$\kappa_x \;=\; \sum_j \lambda_{xj} \ln \frac{\lambda_{xj}}{\tilde{a}_{xj}}. \tag{2.1}$$

Assumption 2 (Linear Entropic Information Cost)**.** The total cost of information processing is linear in relative entropy. Buyers pay $C_x(\lambda_x) = \theta_x \, \kappa_x$. The parameter $\theta_x > 0$ represents the unit price of information processing.

The parameter θ_x is the shadow price of attention capacity. Because payoffs enter the attention index in levels, θ_x takes the same payoff units as v, P, and c, per nat. Within a product and currency, changes in θ_x measure direct changes in the resource price of reallocating sourcing attention. If a buyer exerts zero search effort, its sourcing collapses to the geographical/awareness prior ($\lambda_{xj} = \tilde{a}_{xj}$) and its information cost is zero.

2.3 Matching technology and congestion

Buyers of type x and units offered by seller j meet through a matching function, which determines how many matches form and hence how many units are sold. The measure of buyers of type x directing attention to seller j is $c_{xj} = \mu_x \lambda_{xj}$. The measure of units seller j offers to those buyers is a fixed share of its stock, $d_{xj} = \tilde{s}_{xj}\mu_j$, reflecting where its inventory, agents, or distributors are located. Matching frictions are strong enough that matches never exhaust the units on offer, so total sales never exceed the seller's capacity μ_j.

Assumption 3 (Regularity of Matching)**.** The matching function $M(c, d)$ is twice continuously differentiable, increasing, concave in each argument, and exhibits constant returns to scale.

We specify the matching technology as a Cobb–Douglas function $M_{xj} = \psi_{xj}\, c_{xj}^{\alpha}\, d_{xj}^{1-\alpha}$ with $\alpha \in (0, 1)$. In submarket $\{x, j\}$ a measure c_{xj} of identical buyers meets a measure d_{xj} of identical units, and each match sells one unit. Buyers therefore congest one another, and the meeting rate per unit of attention is

$$m_{xj} \;\equiv\; \frac{M(c_{xj}, d_{xj})}{\mu_x\, \lambda_{xj}} \;=\; A_{xj}\, \lambda_{xj}^{\alpha-1}, \tag{2.2}$$

where A_{xj} absorbs population, capacity, and matching-efficiency parameters.

Buyers are atomistic and take meeting rates as given (they don't internalize the fact that their choice of λ_x affects m_{xj}). However, a price-posting seller can internalize (to some degree or fully) that changing its price alters local congestion, m_{xj} (through the effect it has on λ_{xj}). We summarize the intensity of this internalization by $\rho \in [0, 1]$. Hence, the seller's perceived elasticity of realized matches with respect to buyer attention is: $\eta \;\equiv\; 1 \;-\; \rho\big(1-\alpha\big)$. When sellers fully internalize physical congestion

($\rho = 1$), the elasticity collapses to the standard technological weight ($\eta = \alpha$). In the myopic benchmark ($\rho = 0$), sellers treat congestion as fixed and $\eta = 1$.

The meeting rate m_{xj} measures the probability that a match is produced given a unit of buyer attention. Two markets can host the same number of firms but exhibit vastly different meeting rates. Better certification standards, platform design, legal enforceability, and logistics all improve m_{xj} without altering the raw number of listed suppliers.

2.4 The buyer attention problem

In the second stage of the game, buyers observe all posted prices. Taking matching rates as given, the buyer allocates attention to maximize expected net surplus.

Let $\Delta(J_x^+)$ denote the unit simplex of attention allocations. The buyer solves:

$$\max_{\lambda_x \in \Delta(J_x^+)} \quad \sum_j \lambda_{xj}\, m_{xj}\left(v_{xj} - P_{xj}\right) \;-\; \theta_x \sum_j \lambda_{xj} \ln \frac{\lambda_{xj}}{\tilde{a}_{xj}}. \tag{2.3}$$

Proposition 2.1 (Optimal Sourcing Strategy)**.** *For positive attention prices and prior weights, the buyer's optimization problem has a unique interior solution characterized by the prior-tilted multinomial logit sourcing rule:*

$$\lambda_{xj} \;=\; \frac{\tilde{a}_{xj}\,\exp\!\left(\frac{m_{xj}\,(v_{xj}-P_{xj})}{\theta_x}\right)}{Z_x}, \qquad \textit{where} \quad Z_x \;\equiv\; \sum_k \tilde{a}_{xk}\,\exp\!\left(\frac{m_{xk}\,(v_{xk}-P_{xk})}{\theta_x}\right). \tag{2.4}$$

Holding meeting rates fixed, the semi-elasticity of the attention that a buyer of type x devotes to a seller of type j with respect to price is given by:

$$z_{xj} \;\equiv\; -\frac{\partial \ln \lambda_{xj}}{\partial P_{xj}} \;=\; \frac{m_{xj}}{\theta_x}\left(1 - \lambda_{xj}\right). \tag{2.5}$$

The allocation of attention, λ_{xj} (see Equation 2.4) depends on three distinct forces.[2] The prior distribution of attention $\ln \tilde{a}_{xj}$, that establishes the default geography of trade given by distance, language, and historical ties which dictate where a buyer looks before exerting active effort. The directed return $m_{xj}(v_{xj} - P_{xj})/\theta_x$ that captures the economic fundamentals and combines the offered surplus with the meeting rate, scaled by the

[2] Taking the natural logarithm of λ_{xj} yields a gravity-type share equation (Head and Mayer, 2014).

price of attention. And the partition term $\ln Z_x$ that summarizes the complete set of competing global alternatives.

A suitably parameterized random-utility logit or constant elasticity of substitution (CES) demand system can rationalize the same cross-sectional choice probabilities, so a single cross-section of trade shares cannot distinguish between them; the models differ in how fundamentals and prices map into those shares. However, the underlying economic intuition differs entirely. In a CES model, dispersion represents consumed utility. Buyers spread their purchases because they inherently love variety. In our framework, buyers spread their purchases because finding and verifying the single best global supplier requires scarce cognitive capacity. Sourcing dispersion therefore consumes real physical and cognitive resources. Furthermore, the sensitivity index m_{xj}/θ_x in the semi-elasticity responds directly to changes in matching conditions and verification technologies, rather than remaining an invariant preference parameter.

An attention share does not automatically translate into an observed expenditure share. Realized trade expenditure between a pair (x, j) is proportional to the product of the attention share, the meeting rate, and the price $(\lambda_{xj} m_{xj} P_{xj})$. The observed expenditure share is therefore:

$$\lambda^{\$}_{xj} = \frac{\lambda_{xj} m_{xj} P_{xj}}{\sum_k \lambda_{xk} m_{xk} P_{xk}}. \tag{2.6}$$

Attention shares and expenditure shares coincide only when match conversion rates and prices are uniform across all options. When they vary, observed expenditure shares act as proxies for the underlying attention shares. Origin–product fixed effects absorb the permanent components of these differences, but they do not identify attention shares from expenditure shares, so the empirical mapping treats the observed shares as proxies.

Love of variety versus cognitive cost. The entropy formulation supports a literal procurement interpretation. A buyer begins with a default prior distribution of familiar suppliers. The buyer then expends real resources to move probability mass toward relationships offering higher expected surplus. Supplier qualification, due diligence, and relationship maintenance represent concrete, costly features of international trade.

This mechanism redefines the standard assumption of a "love of variety." A single cross-section of sourcing shares cannot distinguish our attention model from a standard constant elasticity of substitution (CES) or random-utility model. Both frameworks can fit the same choice probabilities. The theoretical distinction emerges when prices, search technology, or meeting rates change.

In a standard preferences interpretation, the dispersion parameter reflects utility curvature. Buyers spread their purchases because they inherently enjoy variety. In our framework, dispersion is a costly response to scarce attention. The term m_{xj}/θ_x acts as a technological price-sensitivity index. Exploring and onboarding alternative international suppliers requires costly cognitive reallocations. Cheaper verification technology lowers θ_x and makes buyers more responsive. Thicker markets raise m_{xj} and accomplish the same outcome. Our model assigns a resource-based meaning to an otherwise familiar gravity coefficient.

The share of attention that the buyer assigns to the seller (2.4) defines the buyer's captivity. We introduce no additional market-power parameter. If a supplier already captures nearly all of a buyer's attention ($\lambda_{xj} \to 1$), offering higher payoffs cannot attract much more probability mass. The buyer has exhausted its cognitive attention budget on that relationship. This mechanical normalization grants an incumbent pricing power once the seller internalizes the buyer's constrained response. Furthermore, λ_{xj} describes a continuous within-buyer sourcing portfolio. This contrasts with models where each buyer draws only one supplier and λ represents a population frequency. The relationship markup requires a buyer to concentrate its own procurement attention.

2.5 The seller pricing problem

In the first stage, sellers post prices.[3] Sellers possess a first-mover advantage. They fully internalize how their posted prices shape the buyers' optimal attention allocations.

The seller chooses the price P_{xj} to maximize expected operating profit:

$$\max_{P_{xj}} \quad \lambda_{xj}(P_{xj})\, m_{xj}\left(\lambda_{xj}(P_{xj})\right)\, \big(P_{xj} - \tau_{xj} c_j\big). \tag{2.7}$$

A lower price attracts attention. When $\rho > 0$, this additional attention congests the submarket and lowers the meeting rate. This feedback dampens the buyer's attention response. Differentiating the buyer's logit condition while carrying this feedback yields

[3] We have considered an extension that models the seller screening allocation either exogenously or endogenously based on the seller's price of attention. In both cases the screening decision separates from the pricing decision conditional on the meeting rate and leaves the pricing equation unchanged (Appendix B).

the seller-perceived demand slope $D_{xj} \equiv -\partial \ln \lambda_{xj} / \partial P_{xj}$:

$$D_{xj} \ = \ \frac{z_{xj}}{1 + \big(1-\eta\big)\, z_{xj}\, \big(v_{xj} - P_{xj}\big)}. \tag{2.8}$$

Proposition 2.2 (Equilibrium Posted Price)**.** *At any interior solution, the equilibrium price admits the closed-form expression:*

$$P^*_{xj} \ = \ \tau_{xj}\, c_j \ + \ \big(1-\eta\big)\big(v_{xj} - \tau_{xj} c_j\big) \ + \ \frac{1}{z_{xj}}, \tag{2.9}$$

where $1/z_{xj} = \frac{\theta_x}{m_{xj}(1-\lambda_{xj})}$ *is the endogenous relationship markup.*

The anatomy of the price equation. Equation (2.9) reveals the economic forces governing international markups. The price decomposes into three additive components. The first component is the marginal cost ($\tau_{xj} c_j$). In a frictionless, perfectly competitive trade model, the price equals this cost floor.

The second component is the surplus share generated by the seller's perceived contribution to matching, $(1-\eta)(v_{xj} - \tau_{xj} c_j)$. When sellers treat congestion as fixed ($\eta = 1$), this term vanishes. When sellers fully internalize the physical congestion response ($\eta = \alpha$), the retained share matches the technological share and mirrors the classical Hosios surplus-sharing condition (Hosios, 1990).

The third component is the relationship markup. It does not arise from curvature in final demand. It arises from the cognitive and resource cost of replacing a concentrated sourcing relationship. A higher attention price θ_x raises the cost of reallocation and increases the markup. A higher meeting rate lowers the markup because a unit of attention becomes more productive. Most importantly, a higher relationship share λ_{xj} raises the markup because the buyer has fewer attention margins left to reallocate. These forces vary across destinations for the same exporter. They generate pricing-to-market without altering the primitive valuation technology.

A two-supplier example. A simple example illustrates how the sourcing share influences multiple outcomes simultaneously. Suppose a buyer allocates $\lambda_1 = 70\%$ of its attention to Supplier 1 and $\lambda_2 = 30\%$ to Supplier 2. The buyer's own-payoff response at Supplier 1 is proportional to $1 - \lambda_1 = 1 - 0.7 = 0.3$. The response at Supplier 2 is proportional to $1 - \lambda_2 = 1 - 0.3 = 0.7$. The buyer is therefore less responsive at the relationship that already commands most of its attention.

The seller markup is proportional to the inverse of the attention left outside the relationship. Supplier 1 has a markup component proportional to 1/0.30. Supplier 2 has a markup component proportional to 1/0.70. Assuming common meeting rates and attention costs, Supplier 1's markup component is more than twice as large. This dynamic requires only that the buyer relies more heavily on Supplier 1. Supplier 1 does not need to possess a large market share in the global economy.

This difference in buyer dependence also dictates how each supplier reacts to a cost shock. Supplier 1 absorbs more of the shock into its markup and changes its border price more aggressively. Supplier 2 passes through a larger fraction of the cost shock. Sourcing sensitivity, the relationship markup, and pass-through move together because they are governed by the same buyer portfolio.

Symmetric subdivision and the attention price. The model also yields a useful invariance result for measurement. Suppose a sourcing option with total attention share Λ and prior mass $\tilde{a}$ is partitioned into K symmetric sub-options. Each sub-option receives share Λ/K and prior $\tilde{a}/K$. The Kullback-Leibler contribution of the entire block simplifies to $\Lambda \ln(\Lambda/\tilde{a})$. Changing a symmetric statistical partition does not mechanically alter the cost θ_x of attention. It does, however, alter the individual supplier share that determines the relationship markup from $1-\Lambda$ to $1-\frac{\Lambda}{K}$. A country-level elasticity therefore depends on the joint pricing response of the suppliers within the origin.

2.6 Matching equilibrium and global properties

Definition 2.3 (Matching Equilibrium)**.** A matching equilibrium is a collection of buyer attention strategies $\{\lambda_x^*\}$, seller prices $\{P_{xj}^*\}$, and submarket matching rates $\{m_{xj}^*\}$ such that:

1. *Buyer Optimality:* Given prices and matching rates, attention allocation λ_x^* solves the buyer's problem (2.3).
2. *Seller Optimality:* Given the buyer's best-response function and perceived congestion elasticities, the price P_{xj}^* solves the seller's profit maximization problem (2.7).
3. *Technological Consistency:* The submarket matching rates satisfy the aggregate

matching function $m^*_{xj} = A_{xj}(\lambda^*_{xj})^{\alpha-1}$ for all active pairs.

Substituting the price best response into the buyer conditions defines the equilibrium potential function:

$$\Phi_\rho(\lambda) \equiv \sum_x \sum_j \big(v_{xj} - \tau_{xj} c_j\big)\, M_{xj}(\lambda) + \frac{\alpha}{\eta}\Big[\sum_x \sum_j \mu_x\, \theta_x \ln\big(1 - \lambda_{xj}\big) - \sum_x \mu_x\, \theta_x \sum_j \lambda_{xj} \ln \frac{\lambda_{xj}}{\tilde{a}_{xj}}\Big]. \tag{2.10}$$

Proposition 2.4 (Core Equilibrium Potential)**.** *The first-order conditions characterizing the matching equilibrium coincide with the gradient conditions of the weighted potential $\Phi_\rho(\lambda)$. The potential is strictly concave on the relative interior of the product of the strategy simplices. The interior matching equilibrium in λ is therefore unique, and prices are uniquely recovered from (2.9).*

The uniqueness statement covers the full interior-price equilibrium on the fixed active set, not merely the first-order-condition system; price-ceiling cases are handled by the Kuhn–Tucker branch discussed below: strict concavity delivers a unique global maximizer of the potential, and global optimality of each seller's price choice follows because, given the buyer's logit best response, the derivative of the seller's log profit is strictly decreasing in its price, so the first-order condition identifies the unique maximum.

The potential provides an economic representation of the pricing feedback generated by costly attention. In the reduced allocation problem, seller pricing acts as a concave penalty on concentration. The physical-surplus term rewards matches in high-surplus pairs. The entropic term penalizes deviation from the prior of attention. The logarithmic term penalizes allocations that render an overly captive relationship.

This framework explains why smooth information costs regularize assignment models. As the attention price θ_x shrinks, the entropy penalty weakens. The allocation approaches deterministic sorting. However, for any positive θ_x, the marginal entropy term diverges near a zero share. Similarly, the captivity penalty diverges as a share approaches one. Within the active set, the equilibrium remains interior in shares. When the available surplus cannot cover the relationship markup, the price simply reaches the buyer's valuation ceiling. At this boundary, the seller acts as a fixed-price option: interior formulas apply up to the ceiling, and beyond it the price stops moving in the constrained direction, so comparative statics use the corresponding price-ceiling branch.

2.7 Efficiency interpretation and the Harberger wedge

We evaluate the efficiency of the matching equilibrium by comparing the core potential to the allocation chosen by a utilitarian world planner. We assume the full-congestion benchmark ($\rho = 1$) and the absence of fiscal tariffs. The planner values match surplus minus information costs:

$$\Omega^{\text{plan}}(\lambda) \;=\; \sum_{xj} \big(v_{xj} - \tau_{xj} c_j\big) M_{xj}(\lambda) \;-\; \sum_{x} \mu_x \theta_x \sum_{j} \lambda_{xj} \ln \frac{\lambda_{xj}}{\tilde{a}_{xj}}. \tag{2.11}$$

Because $\eta = \alpha$ at this benchmark, the market potential decomposes as:

$$\Phi_\rho(\lambda) \;=\; \Omega^{\text{plan}}(\lambda) \;+\; \sum_{xj} \mu_x \theta_x \ln(1 - \lambda_{xj}). \tag{2.12}$$

The second term captures the reduced-form footprint of relationship pricing. It represents the allocation distortion introduced by seller market power. A second-order Taylor expansion around the planner allocation yields the deadweight welfare loss:

$$\mathcal{L}^{\text{DWL}} \;\simeq\; \frac{1}{2} \nabla \mathcal{P}' \, \mathcal{H}^{-1} \, \nabla \mathcal{P}, \qquad \mathcal{P}(\lambda) \;=\; \sum_{xj} \mu_x \theta_x \ln(1 - \lambda_{xj}). \tag{2.13}$$

Here, $\mathcal{H} = -\nabla^2 \Omega^{\text{plan}}$ is the planner's Hessian projected onto the simplex tangent space. The second-order loss is therefore a quadratic form in the pricing wedge, weighted by the curvature of the planner's objective. A markup that is uniform across the options in a buyer's choice set cancels in the logit normalization, so dispersion in relationship markups is the source of the allocative loss. Opening trade improves this allocation by compressing the relative relationship wedges faced by domestic buyers.

3 From Gravity to Endogenous Trade Elasticities

This section explores the theoretical implications of the framework. We first collapse the buyer's sourcing rule and the seller's optimal price into a single scalar condition to understand how fundamental advantages translate into market dominance. We use this representation to establish the limiting cases of the model. We then ask how cost shocks transmit to prices, deriving the pass-through rate. Finally, we establish the endogenous trade elasticity schedules for the buyer-seller relationship, a country of origin, and the macro economy, showing how aggregation mutes the measured responses. We finish by discussing the welfare gains from trade.

3.1 Dominance and the limiting cases

The buyer's optimal sourcing rule (2.4) and the seller's optimal posted price (2.9) form a coupled system of equations. To study how a pair's fundamental cost and valuation translate into market share, it is analytically useful to disentangle this relationship from the rest of the world. We do this by isolating seller j. Let $Z_{x,-j} \equiv \sum_{k \neq j} \tilde{a}_{xk} \exp(m_{xk}(v_{xk} - P_{xk})/\theta_x)$ denote the partition function excluding seller j. This rest-of-world term is invariant to seller j's own contract.

By substituting the equilibrium price (2.9) into the exponent of the logit sourcing rule (2.4), we eliminate the price and obtain an implicit scalar equation for the equilibrium share.

Proposition 3.1 (Dominance representation)**.** *In any interior matching equilibrium, the sourcing share of pair (x, j) satisfies the scalar condition:*

$$\underbrace{\ln \frac{\lambda_{xj}}{1-\lambda_{xj}} + \frac{1}{1-\lambda_{xj}}}_{\textit{dominance function } g(\lambda_{xj})} = \ln \tilde{a}_{xj} + \underbrace{\frac{m_{xj}}{\theta_x} \eta \left(v_{xj} - \tau_{xj} c_j\right)}_{\textit{fundamentals}} - \ln Z_{x,-j}. \tag{3.1}$$

The dominance function $g(\lambda)$ is strictly increasing on $(0, 1)$ and maps onto $(-\infty, +\infty)$. Given fundamentals and the rest-of-world term, the pair's share is unique and increasing in the joint surplus $v_{xj} - \tau_{xj} c_j$, the attention prior $\tilde{a}_{xj}$, and the meeting rate m_{xj}.

Conditional on the matching state, the right-hand side of Equation (3.1) is entirely free of seller j's posted price. Productivity and costs enter through the surplus $v_{xj} - \tau_{xj} c_j$. Geography enters through the prior $\tilde{a}_{xj}$. The competitive environment enters through $Z_{x,-j}$. The left-hand side absorbs the seller's endogenous pricing response. The logistic term $\ln[\lambda/(1-\lambda)]$ is the standard discrete-choice component. The barrier term $(1-\lambda)^{-1}$ is the endogenous markup penalty.

This representation reveals why incomplete specialization naturally arises without assuming a mechanical "love of variety." The barrier term diverges as $\lambda_{xj} \to 1$. Driving a pair share to unity therefore requires unboundedly large fundamentals. As an exporter's cost advantage grows, its buyer becomes progressively more captive. The exporter finds it optimal to convert further improvements in fundamentals into higher markups rather than into more market share. Taking the remainder of the market would require dropping prices and conceding the rents that were created by concentration.

The information price parameter θ_x spans the classical international trade paradigms as limiting cases of this single framework.

The awareness gravity limit. When the price of attention diverges ($\theta_x \to \infty$), the fundamentals term vanishes from Equation (3.1). Sourcing collapses to the attention prior ($\lambda_{xj} \to \tilde{a}_{xj}$). Price sensitivity drops to zero. Attention is then governed exclusively by geographic distance, historical ties, and cultural familiarity, and realized trade inherits this pattern when conversion rates and prices add no further heterogeneity.

The Ricardian comparative advantage limit. When attention is free ($\theta_x \to 0$), buyers allocate all attention exclusively to the single supplier offering the highest effective surplus. The allocation reproduces the sharp specialization patterns of deterministic Ricardian comparative advantage (Dornbusch et al., 1977). However, the departure from the classical limit concerns pricing.

Corollary 3.2 (Ricardian rents)**.** *Let j^* be the dominant supplier for buyer x, and let j_2 denote the runner-up. As $\theta_x \to 0$, holding meeting rates constant, the winning seller's markup converges to a positive limit, provided the interior price remains below the buyer's valuation (otherwise the valuation ceiling binds):*

$$\frac{1}{z_{xj^*}} \longrightarrow \eta\left(v_{xj^*} - \tau_{xj^*}c_{j^*}\right) - \frac{m_{xj_2}}{m_{xj^*}}\eta\left(v_{xj_2} - \tau_{xj_2}c_{j_2}\right). \tag{3.2}$$

Frictionless trade is Ricardian in quantities but not in prices. The winning seller charges a permanent rent equal to its comparative-advantage gap over the second-best alternative. Relationship rents persist in the deterministic limit because market power is an endogenous, permanent consequence of sourcing concentration. At intermediate values of information costs, the model generates smooth gravity flows: trade responds to economic fundamentals but remains anchored by informational geography.

Connections to standard demand systems. Beyond spanning the Ricardian and gravity limits, the information-cost framework connects to the most prominent demand systems in the quantitative trade literature through two modifications. If we modify the buyer's payoff index to evaluate prices in logs rather than levels, and assume suppliers are atomistic (a vanishing concentration index, $H \to 0$, defined in Section 3.3), the model collapses to a constant elasticity of substitution (CES) demand system. In this Armington/CES corner (Armington, 1969), no supplier possesses relationship-specific market power, tariffs are passed through completely, micro and macro elasticities coincide, and the welfare gains from trade coincide with the standard constant-elasticity formula (Arkolakis et al., 2012).

Alternatively, if we maintain the log-payoff link, set $\eta = 1$, and introduce a finite upper-tier elasticity representing substitution across broad product categories, the model replicates the nested-CES variable-markup structure of Atkeson and Burstein (2008). In their framework, a supplier's markup rises with its market share because it acts as a large oligopolistic supplier within a narrowly defined, exogenous product nest.

Our baseline framework provides a unified micro-foundation that bridges these extremes without assuming nested utility or modifying the level-payoff structure. In our model, the "nest" is simply the buyer's scarce cognitive attention budget. A marginal supplier with a near-zero share faces a constant elasticity and prices similarly to a fragmented CES firm. Conversely, a dominant supplier that has captured most of a buyer's attention faces heavily inelastic demand and reprices like a large oligopolistic supplier in the Atkeson–Burstein environment. Consequently, whether a trading relationship behaves like a CES variety or an Atkeson–Burstein oligopolist is not a rigid functional assumption, but an endogenous state governed by the sourcing share λ_{xj}.

3.2 Pass-through

How do cost shocks, such as tariffs or productivity changes, transmit into border prices? Consider an individual trading pair (x, j) with an interior posted price. We evaluate the impact of a marginal shock to delivered marginal cost $C_{xj} \equiv \tau_{xj} c_j$. By implicitly differentiating the coupled price and share equations while holding the meeting rate m_{xj} and the rest-of-world aggregate $Z_{x,-j}$ at their initial equilibrium values, we isolate the immediate, partial-equilibrium pricing response.

Proposition 3.3 (Cost and tariff pass-through)**.** *Holding matching rates fixed, the impact pass-through of the marginal cost into prices is:*

$$\frac{dP_{xj}}{dC_{xj}} \;=\; \eta\left(1-\lambda_{xj}\right). \tag{3.3}$$

Let $P_{xj}^{\text{FOB}} \equiv P_{xj}/\tau_{xj}$ *denote the border price and* $\mu_{xj} \equiv P_{xj}/(\tau_{xj}c_j)$ *denote the price relative to the marginal cost. The response of the border price to a proportional tariff shock is:*

$$\frac{d\ln P_{xj}^{\text{FOB}}}{d\ln \tau_{xj}} \;=\; -1 \;+\; \frac{\eta\left(1-\lambda_{xj}\right)}{\mu_{xj}}. \tag{3.4}$$

For marginal suppliers with negligible market share ($\lambda_{xj} \to 0$), the pass-through of a tariff into prices is high (approaching η). Consequently, the border price response is minimal. As buyer sourcing concentration rises ($\lambda_{xj} \to 1$), pass-through into prices declines toward zero. Foreign exporters absorb the tariff shock by compressing their profit markups to protect the captive relationship.

The parameterization trilemma. Drozd et al. (2026) highlight a "parameterization trilemma" in international macroeconomics: standard models struggle to jointly match realistic markups, low short-run trade elasticities, and incomplete pass-through without layering multiple frictions. Our framework resolves this tension naturally. A single friction, the cognitive cost of buyer attention, endogenously generates the concentrated relationship state that simultaneously supports high markups, renders short-run demand inelastic, and suppresses full tariff pass-through.

3.3 Relationship, origin, and home–foreign elasticities

The model contains three distinct elasticities, depending on the breadth of the shock. An impact relationship elasticity moves one supplier's cost while holding rival contracts fixed. A country of origin elasticity moves the cost of all suppliers in an origin simultaneously. A macro home-versus-foreign elasticity moves the entire foreign block and allows domestic incumbents to reprice.

The relationship elasticity. First, we ask how a buyer shifts attention when a single supplier's cost changes. Define the baseline directed elasticity as:

$$\varepsilon_{D,xj} \;\equiv\; \frac{\eta\, m_{xj}\, \tau_{xj} c_j}{\theta_x}. \tag{3.5}$$

Using the pass-through formula (3.3), the relative attention response to a cost shock to one supplier against an unshocked benchmark is damped by the buyer's captivity:

$$\varepsilon^{\text{rel}}_{xj} \;=\; \varepsilon_{D,xj}\, \big(1 - \lambda_{xj}\big). \tag{3.6}$$

This is a one-pair derivative. It cannot be aggregated by applying the same expression to a country's origin share.

The country of origin elasticity. Second, we ask what happens when a common cost shock dC hits all suppliers in a specific country of origin simultaneously. To compute this, we trace the shock through both the demand system and the sellers' pricing responses.

Suppose an origin country contains K symmetric suppliers. Following the symmetric subdivision principle, let their common individual attention share be ℓ and the origin's aggregate share be $\Lambda = K\ell$. Because they face the same shock, all K suppliers change their prices by the same amount dP. The unshocked rest of the world keeps its prices fixed. In a logit demand system, the change in the aggregate price index is therefore ΛdP. The change in each supplier's share is given by $d\ln\ell = -\frac{m}{\theta}(dP - \Lambda dP) = -\frac{m}{\theta}(1-\Lambda)dP$.

However, we must determine dP. Differentiating the optimal price equation (2.9) reveals that a supplier adjusts its price due to the cost shock, but also due to its shifting market share: $dP = \eta dC + \frac{\theta}{m(1-\ell)^2}d\ell$. Substituting the logit response $d\ell = \ell d\ln\ell$ into this pricing condition and solving for dP/dC yields a dampened pass-through rate $\eta\Pi$, where $\Pi = 1/[1+\frac{\ell(1-\Lambda)}{(1-\ell)^2}]$. Multiplying this pass-through by the unconstrained elasticity yields the origin response.

Proposition 3.4 (Origin-wide elasticity)**.** *Assume a common directed elasticity ε_D across K symmetric suppliers in an origin country i. A uniform proportional cost shock to all suppliers in the country of origin produces the aggregate response relative* *to an unshocked benchmark:*

$$\varepsilon_i \;=\; \varepsilon_D\,\Pi(\ell,\Lambda), \qquad \Pi(\ell,\Lambda) \;\equiv\; \frac{(1-\ell)^2}{(1-\ell)^2+\ell(1-\Lambda)}. \tag{3.7}$$

Granularity heavily influences this aggregation. Adding suppliers increases the number of within-origin contracts that move together. As the number of small suppliers grows, the origin elasticity approaches the unconstrained ε_D. A simple approximation based on the aggregate share $(1-\Lambda)$ misses this simultaneous repricing effect.

The home-versus-foreign elasticity. Third, we ask what happens in a macroeconomic experiment that alters all foreign trade barriers. When foreign options become more expensive (e.g., via a global tariff dC hitting all foreign suppliers), expenditure flows back to domestic incumbents.

The difference here is that domestic suppliers face no cost shock, but they recognize their increased market share and raise prices to extract higher markups: $dP_j = \frac{\theta}{m(1-\lambda_j)^2}d\lambda_j > 0$. This cross-price response suppresses the aggregate trade elasticity.

To compute this, we partition the buyer's options into a domestic block h with aggregate share $\lambda_{nn} \equiv \sum_{j:\,i(j)=x}\lambda_{xj}$, where we keep the conventional gravity notation and n denotes the buyer's own country x, and a foreign block F with share $1-\lambda_{nn}$. Define the individual shares within each block as λ_j. We first define a single-variable dampening kernel $\Pi(x)$ which measures how much a given supplier passes through a shock versus absorbing it into their markup.

$$\Pi(x) \;\equiv\; \frac{(1-x)^2}{(1-x)^2+x}, \qquad \bar{\Pi}_h \;\equiv\; \frac{1}{\lambda_{nn}}\sum_{j\in h}\lambda_j\Pi(\lambda_j), \qquad \bar{\Pi}_F \;\equiv\; \frac{1}{1-\lambda_{nn}}\sum_{j\in F}\lambda_j\Pi(\lambda_j). \tag{3.8}$$

By differentiating the individual logit share and optimal price equations simultaneously, we can express each supplier's price change as a function of their own cost shock and the global price index, scaled by their individual $\Pi(\lambda_j)$. Summing these responses across all firms yields the shift in the aggregate price index. The resulting movement in the relative home-to-foreign shares collapses to a function of the block-level share-weighted means $\bar{\Pi}_h$ and $\bar{\Pi}_F$.

Proposition 3.5 (Home-versus-foreign elasticity)**.** *Let a uniform cost shock hit all foreign suppliers. Assume a common directed elasticity ε_D. Allowing every supplier to* *reprice yields the aggregate elasticity:*

$$\varepsilon_{\text{macro}} \equiv -\frac{\partial \ln[(1-\lambda_{nn})/\lambda_{nn}]}{\partial \ln \tau} = \varepsilon_D \frac{\bar{\Pi}_h \bar{\Pi}_F}{\lambda_{nn}\bar{\Pi}_h + (1-\lambda_{nn})\bar{\Pi}_F}. \tag{3.9}$$

The impact elasticity depends on each block only through its damping mean. Any portfolio can therefore be summarized by effective symmetric-granularity indices H^h and H^F, defined by $\Pi(\lambda_{nn}H^h) = \bar{\Pi}_h$ and $\Pi((1-\lambda_{nn})H^F) = \bar{\Pi}_F$. For a genuinely symmetric block with K suppliers, $H = 1/K$; for an arbitrary portfolio, H is a local symmetric-equivalent index rather than a Herfindahl. The index is sufficient for the impact elasticity but not for the entire dynamic path: within the domestic block, smaller incumbents absorb returning expenditure disproportionately, meaning an asymmetric portfolio de-concentrates along the path toward autarky.

Domestic concentration matters heavily because returning expenditure raises each incumbent's individual share and widens its markup. The elasticity is generally non-monotone in the domestic expenditure share. It approaches ε_D when the domestic block is very small. It falls where returning expenditure materially changes incumbent shares, and returns toward ε_D near autarky when several domestic incumbents receive only a small marginal reallocation from the vanishing foreign block.

With a single domestic supplier, the repricing penalty diverges as captivity becomes complete. However, this divergence only continues until the domestic price hits the buyer's valuation ceiling (Assumption 1). At this boundary, the domestic incumbent cannot extract further markups and acts as a fixed-price option. The repricing penalty vanishes, and the welfare derivative switches to the price-ceiling branch ($dW_x = -\theta_x \, d\ln\lambda_{nn}$). The welfare cost of autarky is therefore steep, but bounded by the buyer's fundamental valuation.

The gap between micro and macro trade elasticity estimates is a direct consequence of this aggregation. In microeconomic regressions, buyers substitute across a long tail of small foreign varieties where no individual exporter possesses pricing power. This reveals the unconstrained directed elasticity. In macroeconomic adjustments, a foreign shock forces buyers back toward a concentrated domestic base. Incumbents raise their markups as they absorb the returning demand, muting the measured aggregate response.

3.4 Slow awareness and the horizon of adjustment

Empirical studies document that trade elasticities increase substantially over time following a policy change (Boehm et al., 2023), reflecting sluggish adjustments in firm-to-firm links (Huneeus, 2018). Existing explanations for this term structure typically operate on the supply side, emphasizing gradual capacity accumulation or slow-moving customer capital (Drozd and Nosal, 2012). In our framework, this dynamic instead arises on the buyer side from the slow updating of the attention prior. A parsimonious way to represent that persistence is:

$$\tilde{a}_{xj,t+1} \;=\; (1-\delta-\nu)\tilde{a}_{xj,t} \;+\; \delta\lambda_{xj,t} \;+\; \nu\tilde{a}^0_{xj}, \tag{3.10}$$

where δ converts realized sourcing into future awareness and ν reanchors awareness toward fundamental geography $\tilde{a}^0$.

Linearizing a one-relationship response around a steady state produces a geometric amplification relative to the static impact effect. Let ε_0 denote the impact response. The cumulative horizon profile is:

$$\frac{\varepsilon_h}{\varepsilon_0} \;=\; 1 \;+\; \frac{\delta}{\nu}\left[1-(1-\nu)^h\right]. \tag{3.11}$$

The ratio δ/ν governs the long-run asymptotic amplification; this treats the shocked relationship's share as small, and for a share λ the long-run amplification is $1+\delta(1-\lambda)/(\nu+\delta\lambda)$. This extension generates a substantial difference between short-run and multi-year trade responses even when the underlying cognitive price of attention remains unchanged.

3.5 Welfare and the ACR-consistent path integral

To make the theoretical results directly comparable to the quantitative trade literature, we can compute the consumption-equivalent welfare gains of trade using the domestic expenditure share as a sufficient statistic (Arkolakis et al., 2012).

Rather than tracking every foreign price change to measure welfare, we can evaluate the gains from trade by tracking the buyer's reliance on a domestic anchor. Opening to trade raises welfare through two channels. First, it provides high-surplus foreign matches (the standard physical reallocation effect). Second, it reduces the buyer's dependence on domestic incumbents. This reduced dependence forces domestic suppliers to compress their relationship markups to compete. We can track this welfare change continuously as an economy moves along the path toward autarky.

Proposition 3.6 (Welfare derivative along the trade path)**.** *Hold meeting rates, the attention prior, and the price of attention fixed. Let the domestic block have arbitrary supplier shares with damping mean $\bar{\Pi}_h$. Along any equilibrium path with interior domestic prices, buyer welfare satisfies:*

$$dW_x \;=\; -\theta_x \,\frac{1}{\bar{\Pi}_h}\, d\ln\lambda_{nn}. \tag{3.12}$$

All foreign responses are encoded entirely in the movement of the aggregate domestic share (λ_{nn}). The factor $1/\bar{\Pi}_h$ isolates the allocation distortion introduced by market power. As expenditure returns home, each incumbent's individual share rises. The incumbent recognizes its growing captive base and widens its relationship markup accordingly. The buyer therefore loses more welfare than the physical reallocation alone implies. A highly concentrated domestic portfolio (where $\bar{\Pi}_h$ is severely dampened) leaves the buyer highly exposed to this markup extraction.

To compare with the quantitative trade literature, we adopt the local Arkolakis–Costinot–Rodríguez-Clare (ACR) sufficient-statistic identity as the accounting convention, with the model's state-dependent elasticity substituted in:

$$d\ln\mathcal{W}_n \;=\; -\frac{1}{\varepsilon_{\text{macro}}(\lambda_{nn})} d\ln\lambda_{nn}. \tag{3.13}$$

Here, $\mathcal{W}_n$ denotes the consumption-equivalent aggregate welfare object; it is a buyer-side statistic that excludes seller profits and is distinct from the payoff-denominated derivative of Proposition 3.6. Substituting our home-versus-foreign elasticity schedule (3.9) into this identity and integrating from the currently observed domestic share up to autarky ($\lambda_{nn} = 1$) yields the total consumption-equivalent gain from trade:

$$\ln\frac{\mathcal{W}_n}{\mathcal{W}_n^{\text{aut}}} \;=\; \int_{\lambda_{nn}}^{1} \frac{d\lambda}{\lambda\,\varepsilon_{\text{macro}}(\lambda;\varepsilon_D, H^h, H^F)}. \tag{3.14}$$

Standard sufficient-statistic formulas assume a constant trade elasticity. In our framework, the aggregate elasticity is state-dependent. As an economy moves toward autarky, domestic incumbents become increasingly captive, extract higher markups, and suppress the macroeconomic elasticity along the path. The true gains from trade require integrating the inverse of this state-dependent schedule.

The integration applies the interior schedule up to the share at which the domestic price reaches the buyer's valuation ceiling. Beyond this point, the incumbent acts as a fixed-price option, and the elasticity reverts to the unconstrained directed elasticity. This two-part integration ensures the welfare cost of closing the economy remains bounded.

4 Evidence and Identification

Our theoretical framework yields two distinct empirical predictions. First, at the microeconomic level, a foreign supplier's pass-through of a tariff depends on how captive its buyer is. Second, at the macroeconomic level, the aggregate trade elasticity is severely damped relative to the microeconomic elasticity because concentrated domestic incumbents raise their markups when foreign goods become more expensive.

To quantify the model and calculate the resulting welfare gains from trade, we estimate the structural parameters jointly using three sets of data moments: sector-level trade elasticities, US tariff pass-through regressions, and a cross-country panel of macroeconomic elasticity gaps.

4.1 Data and empirical mapping

We combine bilateral trade data with detailed tariff changes. From the BACI database (Gaulier and Zignago, 2010), we obtain bilateral trade values and quantities at the 6-digit Harmonized System (HS6) level for the years 2015–2024. Dividing value by quantity yields the unit value, which serves as our measure of the border price (P^{FOB}). We merge this with the Global Tariff Database (Teti, 2024; Rodríguez-Clare et al., 2025) to capture the statutory tariff rates (t) applied during the 2018–2024 US trade war.

To avoid endogeneity (where current tariffs might reduce current trade volumes), we construct the "pre-war sourcing share" using average trade flows from 2015 to 2017. While the model is defined at the buyer–seller level, public customs data do not reveal individual firm-to-firm links. We therefore use an origin country's pre-war share of total US imports in a given product category as our proxy for the buyer's attention share

(λ). Expenditure shares accurately track attention shares when baseline conversion rates and prices are relatively uniform.

Finally, we require estimates of trade elasticities. We anchor the baseline directed elasticity (ε_D) using the median sector-based estimate across the 24 WIOD-matched economies, 5.38, from Imbs and Méjean (2017), a value firmly in the range of modern micro estimates (Simonovska and Waugh, 2014; Fontagné et al., 2022). To discipline our macro-aggregation mechanism, we pair these sector-level estimates with the conventional constrained aggregate elasticity series and domestic expenditure shares for their full panel of 28 countries. This conventional aggregate object is distinct from the theoretical welfare-equivalent aggregate elasticity introduced in Imbs and Méjean (2017).

4.2 Tariff pass-through and the pricing parameters

Our first goal is to test the model's prediction that dominant suppliers absorb cost shocks while marginal suppliers pass them on. We estimate how US import unit values respond to tariffs across origin–product pairs:

$$\ln UV_{ikt} \;=\; \beta_1 \ln(1+t_{ikt}) \;+\; \beta_2 \Big[\ln(1+t_{ikt}) \times \lambda_{ik}^{\text{pre}}\Big] \;+\; \mu_{ik} \;+\; \mu_{kt} \;+\; \mu_{it} \;+\; e_{ikt}. \quad (4.1)$$

Here, UV_{ikt} is the border unit value and t_{ikt} is the statutory tariff rate applied to product k from origin country i in year t, while $\lambda_{ik}^{\text{pre}}$ denotes the pre-war sourcing share and e_{ikt} is the residual error. The fixed effects (μ_{ik}, μ_{kt}, and μ_{it}) absorb permanent origin–product characteristics, global product-level price movements, and origin-specific macroeconomic fluctuations, so that identification comes from relative tariff changes within these cells. Standard errors are clustered by HS2 chapter to account for correlated shocks within broad industries.

Table 4.1 reports the results. The baseline coefficient for pairs with negligible sourcing shares ($\lambda \approx 0$) is statistically indistinguishable from zero ($\beta_1 = -0.058$). This means marginal suppliers do not lower their border prices; they pass the tariff fully on to the US buyer. However, the interaction coefficient is strongly negative and significant ($\beta_2 = -0.812$). This confirms our mechanism: origin–product pairs with large pre-war sourcing shares exhibit substantial absorption, as the relationship mechanism predicts: dominant suppliers compress their own profit markups to protect their captive buyers. Excluding China from the sample (Column 2) leaves this dynamic intact, suggesting it

	$\ln(\text{Unit Value})_{ikt}$		$\ln(\text{Value})_{ikt}$
	(1) All origins	(2) Excl. China	(3) All origins
$\ln(1+t_{ikt})$	-0.058	-0.070	-0.726^{***}
	(0.105)	(0.134)	(0.226)
$\ln(1+t_{ikt}) \times \lambda_{ik}^{\text{pre}}$	-0.812^{***}	-0.653^{**}	-2.977^{***}
	(0.228)	(0.297)	(0.780)
Origin × Product FE	Yes	Yes	Yes
Product × Year FE	Yes	Yes	Yes
Origin × Year FE	Yes	Yes	Yes
Observations	1,008,877	979,240	1,008,877
Clusters (HS2)	96	96	96
R^2	0.889	0.888	0.905

Pre-war sourcing shares ($\lambda_{ik}^{\text{pre}}$) are computed from 2015–2017 trade. Standard errors clustered by HS2 chapter are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4.1: US tariff variation and border unit values, 2018–2024.

is a general feature of trade rather than a China-specific anomaly.

To use these regression coefficients in our structural model, we must map them to our theory. Let $r \equiv (v_{xj} - \tau_{xj} c_j)/(\tau_{xj} c_j)$ denote the surplus-to-cost ratio. The theoretical schedule for the border price response is:

$$\frac{d \ln P^{\text{FOB}}}{d \ln \tau} = -1 + \frac{(1-\lambda)^2}{\Gamma(1-\lambda) + 1/\varepsilon_D}, \qquad \Gamma \equiv \frac{1 + (1-\eta) r}{\eta}. \tag{4.2}$$

Our empirical pass-through regression identifies the pricing composite Γ, which bundles the seller's perceived congestion elasticity (η) and the surplus-to-cost ratio (r). However, pinning down the full model requires us to also identify the underlying trade elasticities and the degree of domestic market power. We turn to these next.

4.3 Identifying domestic concentration from the elasticity gap

Our theoretical framework emphasizes that there is no single "trade elasticity." The elasticity depends entirely on the breadth of the shock and which sellers are able to adjust their markups (see Table 4.2). Confusing these margins is a common source of error in the quantitative trade literature.

To quantify these margins, we rely on a well-documented puzzle in international macroeconomics: microeconomic trade elasticities are systematically larger than macroe-

Margin	Shock	Individual pricing feedback	US model implication
Relationship	One supplier	Own share only	$\approx$ 5.4 (small shares)
Origin	All suppliers in origin	Within-origin rivals reprice	below 5.4
Home–Foreign	All foreign suppliers	Domestic incumbents reprice	$\approx$ 2.4

The margins differ according to the set of contracts that move simultaneously with the shock. The last column reports model-implied US values at the estimated parameters.

Table 4.2: The three elasticity margins in the model.

conomic trade elasticities. The Imbs and Méjean (2017) dataset quantifies this gap across countries. We take their micro/constrained-macro discrepancy as the empirical moment to explain. In their data, when buyers substitute across a long tail of small foreign varieties (a microeconomic shock), the median estimated elasticity across countries is approximately 5.4 (4.9 for the United States). No individual exporter possesses pricing power, so this reveals the unconstrained directed elasticity $\varepsilon_D \approx 5.4$.

However, when an economy faces a nationwide macroeconomic shock (e.g. a global tariff), the conventional aggregate trade elasticity drops to a cross-country median of roughly 2.9, and 1.5 for the United States. Why does this happen? In our model, a nationwide tariff makes all foreign goods more expensive, forcing buyers to return to domestic producers. If domestic production is concentrated, those domestic incumbents recognize the surge in captive demand and raise their markups. This cross-price response chokes off the substitution, severely suppressing the aggregate trade elasticity.

We can use this empirical gap between the micro and macro elasticities to infer how concentrated the domestic market must be. In our theory, the severity of the macro elasticity dampening depends on two things: the domestic expenditure share (which is directly observable; 84 percent for the US in their data and 76 percent in the World Input–Output Database) and the unobserved domestic supplier concentration, denoted by the index H^h. The index H^h measures the effective market power of domestic incumbents, where $H^h = 1$ represents a single domestic monopoly and $H^h \to 0$ represents infinite fragmentation.

Because H^h is not directly observable in standard macro data, we treat it as an unknown parameter. We substitute each country's observed micro and macro elasticities and its domestic expenditure share into our theoretical home-versus-foreign elasticity formula (3.9), and invert the equation. This reverse-engineering process solves for the domestic concentration H^h required to generate the empirical elasticity gap.

We perform the same inversion for all 28 economies in the Imbs and Méjean (2017) dataset. Twenty-two of the 28 economies admit a feasible concentration solution ($H^h \leq 1$). The remaining six economies exhibit an even larger elasticity gap than the domestic captivity mechanism can generate alone, which enter our subsequent joint estimation as the upper tail of the concentration distribution.

4.4 Joint estimation of the structural parameters

Because our structural model involves bounded parameters (e.g., $H^h \in (0, 1]$) and nonlinear relationships, conventional unconstrained standard errors are unreliable at the binding boundary. We therefore estimate all parameters jointly using the limited-information quasi-Bayesian estimator of Chernozhukov and Hong (2003), evaluated on a grid. The estimator searches for the parameter values that minimize the distance between the model's predictions and two blocks of empirical moments. The baseline fixes the matching elasticity at $\alpha = 1/2$, the symmetric Cobb–Douglas benchmark, which bounds $\eta \in [1/2, 1]$.

1. **The Pricing Block:** We project the nonlinear theoretical pass-through curve (4.2) onto the linear regression specification used in Table 4.1. The estimator seeks parameters that match the model-implied β_1 and β_2 to our empirical estimates, weighted by the empirical covariance matrix.

2. **The Micro–Macro Block:** We treat the 28 paired elasticity gaps as noisy observations of our theoretical dampening factor. We model the cross-country concentration distribution hierarchically, which allows us to pool information across countries without forcing them to share a single concentration value.

Table 4.3 reports the baseline quasi-posterior estimates. The data push the pricing composite to $\Gamma = 1.07$. Given our theoretical constraints, this places the perceived elasticity parameter η very close to 1 (median 0.97), meaning sellers price as if market congestion is largely fixed. The baseline markup over marginal cost for a marginal supplier (μ_0) is estimated at roughly 22 percent.

The concentration block yields a population median of 0.85, indicating that domestic supply reprices as if it were concentrated in the hands of a few dominant firms across most countries. For the United States specifically, the concentration is $H^h_{US} = 0.86$. When paired with the US domestic expenditure share of 76%, this concentrated portfo-

	5%	Median	95%
Perceived match-volume elasticity η	0.89	0.97	1.00
Pricing composite Γ	1.00	1.07	1.29
Low-share price-cost ratio μ_0	1.18	1.22	1.33
Directed elasticity ε_D	4.66	5.47	6.28
Population median concentration	0.79	0.85	0.90
US concentration H^h_{US}	0.77	0.86	0.93
US damping factor $\varepsilon_{macro}/\varepsilon_D$	0.33	0.43	0.55
US home–foreign elasticity ε_{macro}	1.75	2.36	3.11
Constant-ε_D welfare gain (%)	4.5	5.2	6.1
State-dependent welfare gain (%)	8.2	11.7	18.5
Ratio of actual to standard gains	1.6	2.3	3.5

NOTE: The table reports the posterior restricted to interior domestic pricing at the observed state and integrates the price-ceiling branch (Appendix F). The top panel shows the estimated structural parameters. The middle panel shows the implied US elasticities. The bottom panel shows the implied gains from trade. Actual welfare gains are more than double the standard constant-elasticity formula.

Table 4.3: Joint quasi-posterior parameter estimates and derived US quantities.

lio generates substantial macroeconomic attenuation: a baseline micro elasticity of 5.47 is cut by more than half, resulting in an aggregate macro elasticity of just 2.36. The difference from the 1.5 source-data US aggregate elasticity reflects hierarchical shrinkage in the inferred US concentration, as well as the transport from the 84 percent domestic share in the elasticity data to the 76 percent WIOD goods-absorption share used for the US counterfactual.

These parameter estimates position the real-world data between the theoretical corners of the trade literature. First, the perceived match-volume elasticity is estimated at $\eta = 0.97$. This means sellers price as if market congestion is essentially fixed, far from the Hosios surplus-sharing corner ($\eta \approx 0.5$) in favor of the standard oligopoly pricing assumptions used by Atkeson and Burstein (2008). Second, the strong pass-through interaction ($\beta_2 = -0.81$) also places the data far from the pure CES corner, which would predict border-price pass-through to be invariant to market share.

Instead, the data reveal an asymmetry between foreign and domestic supply. Marginal foreign suppliers with negligible shares pass tariffs through almost fully, behaving like fragmented CES firms. In contrast, the US domestic market concentration is estimated at $H^h_{US} = 0.86$. On a scale where 0 is the atomistic CES world and 1 is a single captive domestic monopoly, the US economy sits 86 percent of the way toward the captive corner.

This state-dependence explains the gap between the micro and macro elasticities, and the resulting welfare implications. Foreign imports operate largely in the fragmented CES regime, yielding a high directed elasticity. However, concentrated domestic supply operates closer to the captive Atkeson–Burstein regime. A nationwide tariff forces expenditure out of the competitive CES world and into the captive domestic world, granting domestic incumbents immense pricing power along the transition path. This is why the aggregate macro elasticity falls to 2.36.

Robustness and direct validation Before evaluating the macroeconomic and welfare implications of these parameters, we provide several overidentifying checks on the estimation. The model predicts a specific, convex shape for the pass-through schedule. If we estimate the tariff response non-parametrically across discrete share bins, the data exhibits the steepening predicted by the model through the mid-tier shares, flattening out only at the very highest shares.

We test the robustness of these estimates across several alternative specifications detailed in the Appendix. Whether we exclude China from the pass-through panel, widen the priors on the underlying surplus ratios, or anchor the elasticity to alternative datasets, the core mechanism remains intact. Across all robustness specifications, the proportional welfare correction generated by domestic captivity remains between 2.3 and 2.8.

While these results are highly robust within aggregate data, it is important to note that public customs data measure origin–product shares rather than true buyer–seller dependence. Direct validation of the mechanism ultimately requires matched firm-to-firm transaction records. Matched data can replace the origin-proxy with the buyer's actual pre-shock supplier share, providing a more direct test of relationship-level pass-through and domestic concentration. Proceeding with our validated baseline estimates, we now quantify the macroeconomic consequences of state-dependent market power.

4.5 The quantitative gains from trade

With the structural parameters jointly estimated, we can evaluate the true welfare gains from trade. We transport the estimated US symmetric-equivalent concentration index ($H^h_{US} = 0.86$) to the US goods-absorption share of $\lambda_{nn} = 0.76$, constructed from the World Input–Output Database (Timmer et al., 2015). We conservatively assume

Domestic Concentration Index (H^h)	0.30	0.50	0.70	**0.86**	0.90
Symmetric-equivalent suppliers $1/H^h$	3.33	2.00	1.43	**1.16**	1.11
Open macro elasticity	4.93	4.35	3.40	**2.33**	2.03
Interior-path welfare gain	5.5%	6.1%	7.7%	**12.5%**	15.9%
Gain relative to constant ε_D benchmark	1.06	1.16	1.46	**2.39**	3.04

NOTE: Gains are consumption-equivalent levels. The constant directed-elasticity (ACR) gain is 5.2% across all columns. The bold column represents the posterior median US concentration estimated from the cross-country elasticity gaps.

Table 4.4: US implications conditional on the domestic concentration index H^h.

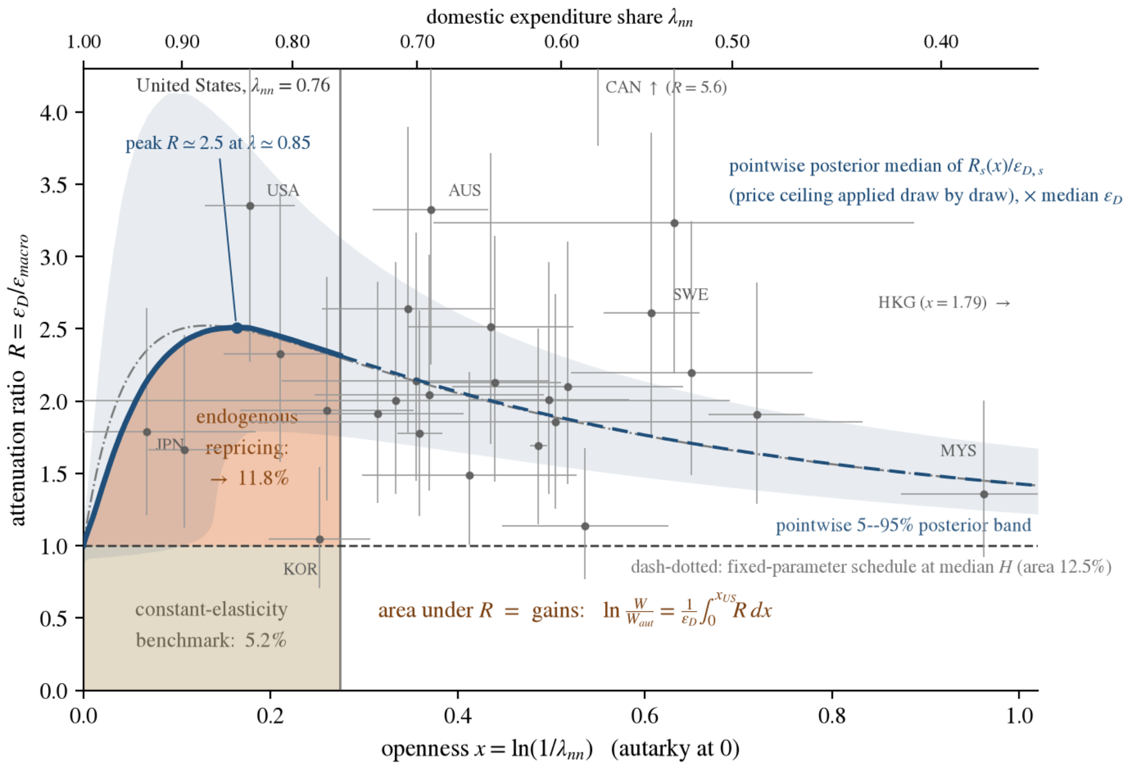
granular foreign supply ($H^F = 0$), which isolates the domestic repricing effect.

The standard ACR benchmark evaluates the gains using a constant directed elasticity ($\varepsilon_D \approx 5.38$) along the entire path to autarky. This traditional approach yields a US consumption-equivalent gain of 5.2 percent.

Table 4.4 demonstrates why domestic concentration drastically alters this assessment. A dispersed theoretical benchmark ($H^h = 0.5$) produces a macro elasticity of 4.35 and a welfare gain of only 6.1 percent. But the empirical data imply a much higher effective concentration. At $H^h_{US} = 0.86$, the state-dependent path integration forces buyers back into concentrated domestic relationships. This revives local monopoly power, steepening the welfare cost of closure. Accounting for the full posterior distribution (and the valuation ceiling), the estimated consumption-equivalent gain rises to 11.7 percent. The true welfare gains are more than double the standard estimate.

Figure 4.1 visually synthesizes the paper's identifying data, its central aggregation mechanism, and its welfare implication in a single picture. The horizontal axis measures openness, so autarky sits at the origin. The vertical axis measures the attenuation ratio ($\varepsilon_D/\varepsilon_{\text{macro}}$). Each cross represents one economy from the Imbs and Méjean (2017) dataset. The solid curve is the posterior schedule, the pointwise median across draws of the welfare integrand with each draw's price ceiling imposed, so that the shaded area corresponds to the posterior gain rather than to a single parameter value.

Because of the mathematical properties of the ACR identity, the area under this curve, scaled by $1/\varepsilon_D$, corresponds to the welfare gains from trade. The standard constant-elasticity benchmark assumes the attenuation ratio is always 1 (the flat rectangle), yielding 5.2 percent. However, the data identify a hump-shaped schedule. Attenuation peaks at intermediate openness, where foreign competition still exists but domestic incumbents are already highly dominant. Integrating the area under this pos-



NOTE: Each cross is one economy in the paired elasticity sample, plotted against openness $x = \ln(1/\lambda_{nn})$. Bars are diagnostic ranges, not sampling standard errors: vertical bars are ±one residual standard deviation of the hierarchical model; horizontal bars are ±half the gap between the elasticity-source and WIOD measurements of the domestic share. The solid curve is the posterior schedule: at each openness level, the median across posterior draws of the welfare integrand $R_s(x)/\varepsilon_{D,s}$, with the price ceiling applied draw by draw, rescaled by the posterior-median ε_D so that it reads in units of $R = \varepsilon_D/\varepsilon_{\text{macro}}$. The thin dash-dotted curve is the fixed-parameter schedule $R(\lambda; H) = 1 + H\lambda(1-\lambda)/(1-H\lambda)^2$ at the posterior-median $H^h_{US} = 0.86$. The shaded band spans the pointwise 5–95 percent posterior range. Because $dx = -d\ln\lambda_{nn}$, the area under the schedule between autarky (origin) and the observed US openness, divided by ε_D, equals the log consumption-equivalent gain from trade. The flat rectangle of height 1 represents the standard constant-elasticity ACR benchmark (5.2%). The shaded region above it represents the additional welfare loss from endogenous domestic repricing. The total The total area under the solid curve gives an 11.8% gain, within 0.1 percentage point of the posterior-median gain of 11.7%; the fixed-parameter schedule gives 12.5%

Figure 4.1: The elasticity schedule and the gains from trade.

terior schedule incorporates the massive domestic repricing penalty and delivers the 11.7 percent gain; the fixed-parameter schedule at the posterior-median concentration index, shown dash-dotted, gives 12.5 percent.

4.6 Policy implications

Beyond the static welfare evaluation, the theoretical framework provides a unified, systematic rationale for the complex anatomy of recent trade wars. Empirical evaluations of successive tariff regimes point to significant heterogeneity in pass-through and macroeconomic incidence. For instance, the targeted 2018–2019 tariffs were passed through almost fully at the border yet exhibited relatively muted effects on aggregate US consumer prices, driven by retail margin compression and rapid offshore substitution (Amiti et al., 2019b; Fajgelbaum et al., 2020; Cavallo et al., 2021; Fajgelbaum and Khandelwal, 2022). In contrast, the application of broader baseline tariffs in the mid-2020s exerted much stronger, gradually accumulating upward pressure on the domestic price level (e.g., Cavallo et al., 2025; Hacıoğlu-Hoke et al., 2026; Minton et al., 2026). Furthermore, the literature increasingly observes that tariff pass-through is non-linear, steepening as the scale of the intervention grows.

Standard quantitative models, which predominantly assume a constant elasticity of substitution (CES), treat pass-through as log-linear and invariant to the scale or breadth of the shock. Our information-cost framework departs from this benchmark, mapping these empirical phenomena directly to the endogenous state of buyer captivity.

The non-linearity of pass-through. Under standard CES preferences, a 5 percent tariff and a 20 percent tariff exhibit the same proportional pass-through rate at the border. In our framework, pass-through is state-dependent: $dP/dC = \eta(1 - \lambda)$. When a small tariff is initially levied on a dominant foreign supplier (large λ), that supplier possesses a substantial "markup cushion." Because the supplier extracts significant rents from the buyer's concentrated attention, it finds it optimal to compress its markup and absorb the cost shock rather than risk losing the captive relationship. Thus, a marginal hike from 0 to 5 percent sees low pass-through.

However, as the tariff climbs, the foreign supplier's fundamental cost advantage shrinks, and its market share λ inevitably falls. As λ drops, the barrier term $1/(1 - \lambda)$ governing the relationship markup weakens. The supplier exhausts its markup cushion and can no longer afford to absorb the cost. Consequently, the marginal pass-through schedule steepens. A tariff increase from 15 to 20 percent will bite the domestic buyer significantly harder than the initial 0 to 5 percent increase.

Targeted versus blanket tariffs. The model also formally distinguishes between the macroeconomic incidence of a targeted tariff (such as the China-specific duties of 2018) and a blanket tariff (universal duties on all foreign origins).

When policymakers levy a targeted tariff, buyers can substitute toward un-tariffed foreign competitors. The relevant substitution margin is the origin-wide elasticity of Proposition 3.4, which approaches the unconstrained directed elasticity ($\varepsilon_D \approx 5.4$) when the targeted origin contains many granular suppliers. Because alternative foreign options remain untariffed, domestic incumbents experience little captive surge in demand and have little room to raise their prices. The targeted foreign suppliers absorb a portion of the shock, buyers offshore the remainder, and the aggregate domestic inflation effect is highly muted.

A blanket tariff alters this dynamic by shutting down the foreign substitution margin. Buyers are forced to redirect expenditure homeward. As demand returns to the United States, it flows into a highly concentrated domestic supplier base ($H^h_{US} = 0.86$). Domestic incumbents recognize that foreign alternatives are now universally expensive, rendering the US buyer captive. Rather than simply absorbing the new volume at constant prices, these incumbents aggressively hike their relationship markups. This cross-price response suppresses the aggregate macro elasticity to 2.36. The domestic consumer consequently faces a dual penalty: paying the tariff on imported goods while simultaneously absorbing newly elevated monopoly markups on domestic goods.

The horizon of adjustment. Finally, the framework provides intuition for how these price effects evolve over time. In our setting, dynamics are governed by the slow-moving cognitive geography of the attention prior ($\tilde{a}$).

For targeted tariffs, the long-run effect on domestic prices is largely zero-reverting. Initially, buyers are anchored to their historical awareness of the targeted country. Over time, as buyers force themselves to allocate attention to alternative foreign hubs, the attention prior gradually updates, consistent with the slow rise of trade elasticities over time documented by Boehm et al. (2023). As these new relationships formalize, the originally targeted firms lose whatever residual captivity they retained, and the temporary price disruptions wash out as supply chains reorganize offshore.

For blanket tariffs, the time-path is the opposite: the effects accumulate. When all foreign goods are tariffed and buyers are forced home, their awareness of foreign suppliers literally depreciates over time. The domestic attention prior ($\tilde{a}_{\text{home}}$) solidifies. As domestic incumbents become increasingly entrenched in the buyers' cognitive priors, domestic captivity rises, and domestic markups widen further. Rather than dissipating, the inflation shock locks in.

A natural qualification is that, over sufficiently long horizons, permanently elevated domestic markups should induce the entry of new domestic suppliers. Such dynamic entry would dilute the domestic concentration index (H^h) and alleviate the repricing penalty. However, to the extent that domestic production features high fixed costs, regulatory barriers, or requires scarce cognitive attention from buyers, existing incumbents can extract these state-dependent rents for a considerable time. Recognizing this endogeneity of market power is therefore essential for evaluating the true incidence of macro-trade policy.

5 Conclusion

We develop a model of international trade that replaces the mechanical assumption of a "love of variety" with a more realistic cost of information processing. When buyers face cognitive constraints in exploring and maintaining global supplier relationships, they naturally concentrate their sourcing. Suppliers, in turn, exploit the cost of shifting this attention to extract relationship-specific markups. This single friction bridges the gap between fragmented constant-elasticity models and variable-markup oligopoly models, demonstrating that pricing behavior is not a rigid functional assumption, but an endogenous state governed by market share.

Our framework provides a unified explanation for persistent empirical puzzles in international macroeconomics. Using recent US tariff data, we show that pass-through is highly non-linear: dominant foreign suppliers absorb cost shocks to protect their captive relationships, while marginal suppliers pass them on to the buyer. Furthermore, by mapping the attenuation between micro and macro trade elasticities to the domestic market structure, we reveal that US domestic supply is characterized by entrenched incumbents.

Recognizing this domestic concentration alters our understanding of modern trade wars. While targeted tariffs allow buyers to substitute across a competitive fringe of offshore suppliers, blanket baseline tariffs force buyers to return expenditure home. Captive domestic incumbents exploit this lack of foreign competition by raising their markups. State-dependent repricing chokes off aggregate substitution, halving the macroeconomic trade elasticity. When we integrate this state dependence along the path to a closed economy, we find that the true welfare cost of US autarky is 11.7 percent, more than double the standard constant-elasticity prediction.

Ultimately, our results caution against treating trade elasticities and pass-through rates as invariant parameters. As global trade policy shifts from selective interventions to broad closures, assuming constant substitution masks the inflationary power of domestic monopolies. Because sourcing is governed by slow-moving cognitive awareness, the price disruptions of blanket tariffs do not dissipate over time, but accumulate. As buyers forget offshore alternatives, domestic market power locks in, leaving consumers paying the price of deglobalization long after the initial shock.

References

George Alessandria, Horag Choi, and Kim J. Ruhl. Trade Adjustment Dynamics and the Welfare Gains from Trade. *Journal of International Economics*, 131:103458, 2021. doi: 10.1016/j.jinteco.2021.103458

George A. Alessandria, Shafaat Y. Khan, Armen Khederlarian, Kim J. Ruhl, and Joseph B. Steinberg. Recovering Credible Trade Elasticities from Incredible Trade Reforms. Technical report, NBER Working Paper 33568, 2025. doi: 10.3386/w33568

Treb Allen. Information Frictions in Trade. *Econometrica*, 82(6):2041–2083, 2014. doi: 10.3982/ECTA10984

Vanessa Alviarez, Michele Fioretti, Ayumu Ken Kikkawa, and Monica Morlacco. Two-Sided Market Power in Firm-to-Firm Trade. Technical report, NBER Working Paper 31253, 2023. doi: 10.3386/w31253

Mary Amiti, Oleg Itskhoki, and Jozef Konings. International Shocks, Variable Markups, and Domestic Prices. *Review of Economic Studies*, 86(6):2356–2402, 2019a. doi: 10.1093/restud/rdz005

Mary Amiti, Sebastian Heise, and David E. Weinstein. The Anatomy of Tariff Pass-Through into Consumer Prices. NBER Working Paper 35561, 2026. doi: 10.3386/w35561

Mary Amiti, Stephen J. Redding, and David E. Weinstein. The Impact of the 2018 Tariffs on Prices and Welfare. *Journal of Economic Perspectives*, 33(4):187–210, 2019b. doi: 10.1257/jep.33.4.187

James E. Anderson and Eric van Wincoop. Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192, 2003. doi: 10.1257/000282803321455214

Pol Antràs, Teresa C. Fort, and Felix Tintelnot. The Margins of Global Sourcing: Theory and Evidence from US Firms. *American Economic Review*, 107(9):2514–2564, 2017. doi: 10.1257/aer.20141685

Costas Arkolakis, Arnaud Costinot, and Andrés Rodríguez-Clare. New Trade Models, Same Old Gains? *American Economic Review*, 102(1):94–130, 2012. doi: 10.1257/aer.102.1.94

Costas Arkolakis, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. The Elusive Pro-Competitive Effects of Trade. *Review of Economic Studies*, 86(1):46–80, 2019. doi: 10.1093/restud/rdx075

Costas Arkolakis, Federico Huneeus, and Yuhei Miyauchi. Production Network Formation, Trade, and Welfare. Technical report, NBER Working Paper 30954, revised December 2025. First circulated February 2023. doi: 10.3386/w30954

Paul S. Armington. A Theory of Demand for Products Distinguished by Place of Production. *IMF Staff Papers*, 16(1):159–178, 1969. doi: 10.2307/3866403

Andrew Atkeson and Ariel Burstein. Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review*, 98(5):1998–2031, 2008. doi: 10.1257/aer.98.5.1998

Xiwen Bai, Jesús Fernández-Villaverde, Yiliang Li, Ricardo Marto, and Francesco Zanetti. Shipping to America. Manuscript, April 2026. https://www.ricardomarto.com/publication/wp_shipping2america/

Andrew B. Bernard, Andreas Moxnes, and Yukiko U. Saito. Production Networks, Geography, and Firm Performance. *Journal of Political Economy*, 127(2):639–688, 2019. doi: 10.1086/700764

Andrew B. Bernard, Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes. The Origins of Firm Heterogeneity: A Production Network Approach. *Journal of Political Economy*, 130(7):1765–1804, 2022. doi: 10.1086/719759

Daniel M. Bernhofen and John C. Brown. An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan. *American Economic Review*, 95 (1):208–225, 2005. doi: 10.1257/0002828053828491

Christoph E. Boehm, Andrei A. Levchenko, and Nitya Pandalai-Nayar. The Long and Short (Run) of Trade Elasticities. *American Economic Review*, 113(4):861–905, 2023. doi: 10.1257/aer.20210225

Christian Broda, Nuno Limão, and David E. Weinstein. Optimal Tariffs and Market Power: The Evidence. *American Economic Review*, 98(5):2032–2065, 2008. doi: 10.1257/aer.98.5.2032

Alberto Cavallo, Gita Gopinath, Brent Neiman, and Jenny Tang. Tariff Pass-Through at the Border and at the Store: Evidence from US Trade Policy. *American Economic Review: Insights*, 3(1):19–34, 2021. doi: 10.1257/aeri.20190536

Alberto Cavallo, Paola Llamas, and Franco M. Vazquez. Tracking the Short-Run Price Impact of U.S. Tariffs. NBER Working Paper 34496, 2025. doi: 10.3386/w34496

Anton Cheremukhin and Paulina Restrepo-Echavarria. An Information-Based Theory of Monopsony Power. Federal Reserve Bank of Dallas Working Paper 2518 (revised 2026), 2026. doi: 10.24149/wp2518

Anton Cheremukhin, Paulina Restrepo-Echavarria, and Antonella Tutino. Targeted Search in Matching Markets. *Journal of Economic Theory*, 185:104956, 2020. doi: 10.1016/j.jet.2019.104956

Victor Chernozhukov and Han Hong. An MCMC Approach to Classical Estimation. *Journal of Econometrics*, 115(2):293–346, 2003. doi: 10.1016/S0304-4076(03)00100-3

Emmanuel Dhyne, Ayumu Ken Kikkawa, and Glenn Magerman. Imperfect Competition in Firm-to-Firm Trade. *Journal of the European Economic Association*, 20(5):1933–1970, 2022. doi: 10.1093/jeea/jvac013

Rudiger Dornbusch, Stanley Fischer, and Paul A. Samuelson. Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review*, 67(5):823–839, 1977. https://www.jstor.org/stable/1828066

Lukasz A. Drozd and Jaromir B. Nosal. Understanding International Prices: Customers as Capital. *American Economic Review*, 102(1):364–395, 2012. doi: 10.1257/aer.102.1.364

Lukasz A. Drozd, Marcin Kolasa, and Jaromir B. Nosal. Pricing-to-Market in Business Cycle Models. *Review of Economic Dynamics*, 60:101333, 2026. doi: 10.1016/j.red.2026.101333

Jonathan Eaton and Samuel Kortum. Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779, 2002. doi: 10.1111/1468-0262.00352

Chris Edmond, Virgiliu Midrigan, and Daniel Yi Xu. Competition, Markups, and the Gains from International Trade. *American Economic Review*, 105(10):3183–3221, 2015. doi: 10.1257/aer.20120549

Pablo D. Fajgelbaum, Pinelopi K. Goldberg, Patrick J. Kennedy, and Amit K. Khandelwal. The Return to Protectionism. *Quarterly Journal of Economics*, 135(1):1–55, 2020. doi: 10.1093/qje/qjz036

Pablo D. Fajgelbaum and Amit K. Khandelwal. The Economic Impacts of the US–China Trade War. *Annual Review of Economics*, 14:205–228, 2022. doi: 10.1146/annurev-economics-051420-110410

Robert C. Feenstra, Philip Luck, Maurice Obstfeld, and Katheryn N. Russ. In Search of the Armington Elasticity. *Review of Economics and Statistics*, 100(1):135–150, 2018. doi: 10.1162/rest_a_00696

Lionel Fontagné, Houssein Guimbard, and Gianluca Orefice. Tariff-Based Product-Level Trade Elasticities. *Journal of International Economics*, 137:103593, 2022. doi: 10.1016/j.jinteco.2022.103593

Mogens Fosgerau, Emerson Melo, André de Palma, and Matthew Shum. Discrete Choice and Rational Inattention: A General Equivalence Result. *International Economic Review*, 61(4):1569–1589, 2020. doi: 10.1111/iere.12469

Guillaume Gaulier and Soledad Zignago. BACI: International Trade Database at the Product-Level. The 1994–2007 Version. CEPII Working Paper 2010-23, 2010. doi: 10.2139/ssrn.1994500

Pinelopi K. Goldberg and Giovanni Maggi. Protection for Sale: An Empirical Investigation. *American Economic Review*, 89(5):1135–1155, 1999. doi: 10.1257/aer.89.5.1135

Gita Gopinath and Brent Neiman. The Incidence of Tariffs: Rates and Reality. *Journal of Economic Perspectives*, 40(3): 123–144, 2026. doi: 10.1257/jep.20251486

Sinem Hacıoğlu-Hoke, Sarojini Malladi, and Leo Feler. The Slow Climb: How Tariffs Gradually Raised Retail Prices in 2025. FEDS Notes, Board of Governors of the Federal Reserve System, March 5, 2026. doi: 10.17016/2380-7172.4010

Keith Head and Thierry Mayer. Gravity Equations: Workhorse, Toolkit, and Cookbook. In *Handbook of International Economics, Vol. 4*, pages 131–195. Elsevier, 2014. doi: 10.1016/B978-0-444-54314-1.00003-3

Sebastian Heise. Firm-to-Firm Relationships and the Pass-Through of Shocks: Theory and Evidence. *Review of Economics and Statistics*, forthcoming, 2024. doi: 10.1162/rest_a_01502

Thomas J. Holmes, Wen-Tai Hsu, and Sanghoon Lee. Allocative Efficiency, Markups, and the Welfare Gains from Trade. *Journal of International Economics*, 94(2): 195–206, 2014. doi: 10.1016/j.jinteco.2014.07.002

Arthur J. Hosios. On the Efficiency of Matching and Related Models of Search and Unemployment. *Review of Economic Studies*, 57(2):279–298, 1990. doi: 10.2307/2297382

Wen-Tai Hsu, Yi Lu, and Guiying Laura Wu. Competition, Markups, and Gains from Trade: A Quantitative Analysis of China between 1995 and 2004. *Journal of International Economics*, 122:103266, 2020. doi: 10.1016/j.jinteco.2019.103266

Federico Huneeus. Production Network Dynamics and the Propagation of Shocks. Working paper, Princeton University, 2018. https://www.fedehuneeus.com/s/JMP_FHL.pdf

Anna Ignatenko, Ahmad Lashkaripour, Luca Macedoni, and Ina Simonovska. Making America Great Again? The Economic Impacts of Liberation Day Tariffs. *Journal of International Economics*, 157:104138, 2025. doi: 10.1016/j.jinteco.2025.104138

Jean Imbs and Isabelle Méjean. Elasticity Optimism. *American Economic Journal: Macroeconomics*, 7(3):43–83, 2015. doi: 10.1257/mac.20130231

Jean Imbs and Isabelle Méjean. Trade Elasticities. *Review of International Economics*, 25(2):383–402, 2017. doi: 10.1111/roie.12270

Douglas A. Irwin. The Welfare Cost of Autarky: Evidence from the Jeffersonian Trade Embargo, 1807–09. *Review of International Economics*, 13(4):631–645, 2005. doi: 10.1111/j.1467-9396.2005.00527.x

Timothy J. Kehoe and Kim J. Ruhl. How Important is the New Goods Margin in International Trade? *Journal of Political Economy*, 121(2):358–392, 2013. doi: 10.1086/670272

Paul Krugman. Pricing to Market When the Exchange Rate Changes. In S. W. Arndt and J. D. Richardson, editors, *Real-Financial Linkages among Open Economies*, pages 49–70. MIT Press, Cambridge, MA, 1987. NBER Working Paper 1926 version, 1986: doi: 10.3386/w1926.

Rodney D. Ludema and Anna Maria Mayda. Do Terms-of-Trade Effects Matter for Trade Agreements? Theory and Evidence from WTO Countries. *Quarterly Journal of Economics*, 128(4):1837–1893, 2013. doi: 10.1093/qje/qjt016

Filip Matějka and Alisdair McKay. Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review*, 105(1): 272–298, 2015. doi: 10.1257/aer.20130047

Marc J. Melitz and Gianmarco I. P. Ottaviano. Market Size, Trade, and Productivity. *Review of Economic Studies*, 75(1):295–316, 2008. doi: 10.1111/j.1467-937X.2007.00463.x

Marc J. Melitz and Stephen J. Redding. New Trade Models, New Welfare Implications. *American Economic Review*, 105(3):1105–1146, 2015. doi: 10.1257/aer.20130351

Robert Minton, Madeleine Ray, and Mariano Somale. Detecting Tariff Effects on Consumer Prices in Real Time – Part II. FEDS Notes, Board of Governors of the Federal Reserve System, April 8, 2026. doi: 10.17016/2380-7172.4040

B. Ravikumar, Ana Maria Santacreu, and Michael Sposi. Capital Accumulation and Dynamic Gains from Trade. *Journal of International Economics*, 119:93–110, 2019. doi: 10.1016/j.jinteco.2019.04.009

Andrés Rodríguez-Clare, Feodora Teti, Mauricio Ulate, José P. Vásquez, and Román D. Zárate. The 2025 Trade War: Dynamic Impacts Across U.S. States and the Global Economy. CESifo Working Paper 12179, October 2025. doi: 10.2139/ssrn.5571078

Kim J. Ruhl. The International Elasticity Puzzle. Working paper, NYU Stern WP 08-30, 2008. https://ideas.repec.org/p/ste/nystbu/08-30.html

Ina Simonovska and Michael E. Waugh. The Elasticity of Trade: Estimates and Evidence. *Journal of International Economics*, 92(1):34–50, 2014. doi: 10.1016/j.jinteco.2013.10.001

Ina Simonovska and Michael E. Waugh. Trade Models, Trade Elasticities, and the Gains from Trade. *Journal of International Economics*, 162:104268, 2026. doi: 10.1016/j.jinteco.2026.104268

Christopher A. Sims. Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003. doi: 10.1016/S0304-3932(03)00029-1

Meredith Startz. The Value of Face-to-Face: Search and Contracting Problems in Nigerian Trade. Technical report, Working paper (SSRN 3096685; PEDL), 2021. doi: 10.2139/ssrn.3096685

Feodora A. Teti. Missing Tariffs. CESifo Working Paper 11590, 2024. doi: 10.2139/ssrn.5097020

Marcel P. Timmer, Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. An Illustrated User Guide to the World Input–Output Database: The

Case of Global Automotive Production. *Review of International Economics*, 23(3): 575–605, 2015. doi: 10.1111/roie.12178

World Trade Organization. European Union: Safeguard Measures on Certain Steel Products (DS595). Panel report WT/DS595/R, circulated 29 April 2022, adopted 31 May 2022. doi: 10.30875/25189832-595

A Mathematical Proofs for the Core Model

This section provides the complete derivations for the buyer attention problem, the seller pricing problem, the equilibrium potential function, and the granularity invariance result.

A.1 Proof of Proposition 2.1 (Buyer Sourcing)

The buyer maximizes expected net surplus minus the relative entropy cost of attention. The expected payoff term is linear in the attention vector λ_x. The relative entropy cost is convex on the compact simplex. The objective function is therefore strictly concave. The marginal cost of attention diverges to negative infinity as any probability share approaches zero ($\lim_{\lambda_{xj}\to 0^+} \ln(\lambda_{xj}/\tilde{a}_{xj}) = -\infty$). The optimal attention allocation is therefore interior for all options with a positive prior $\tilde{a}_{xj} > 0$.

We formulate the Lagrangian with multiplier ξ_x associated with the unit-sum constraint $\sum_j \lambda_{xj} = 1$:

$$\mathcal{L}_x(\lambda_x, \xi_x) \;=\; \sum_{j\in J} \lambda_{xj}\, m_{xj}\left(v_{xj} - P_{xj}\right) \;-\; \theta_x \sum_{j\in J} \lambda_{xj} \ln\frac{\lambda_{xj}}{\tilde{a}_{xj}} \;-\; \xi_x\Big(\sum_{j\in J} \lambda_{xj} - 1\Big).$$

The first-order condition with respect to the attention share λ_{xj} is:

$$m_{xj}\left(v_{xj} - P_{xj}\right) \;-\; \theta_x\Big(\ln\frac{\lambda_{xj}}{\tilde{a}_{xj}} + 1\Big) \;-\; \xi_x \;=\; 0.$$

Rearranging terms isolates the log attention share:

$$\ln\frac{\lambda_{xj}}{\tilde{a}_{xj}} \;=\; \frac{m_{xj}\left(v_{xj} - P_{xj}\right)}{\theta_x} \;-\; \Big(1 + \frac{\xi_x}{\theta_x}\Big).$$

$$\lambda_{xj} \;=\; \tilde{a}_{xj}\, \exp\Big(\frac{m_{xj}\left(v_{xj} - P_{xj}\right)}{\theta_x}\Big)\, \exp\Big(-1 - \frac{\xi_x}{\theta_x}\Big).$$

Summing across all options $j \in J$ and imposing the unit-sum condition $\sum_{j\in J} \lambda_{xj} = 1$ identifies the Lagrange multiplier term:

$$\exp\Big(1 + \frac{\xi_x}{\theta_x}\Big) \;=\; \sum_{k\in J} \tilde{a}_{xk}\, \exp\Big(\frac{m_{xk}\left(v_{xk} - P_{xk}\right)}{\theta_x}\Big) \;\equiv\; Z_x.$$

Substituting the partition function Z_x back into the attention share yields the prior-tilted multinomial logit sourcing formula.

To evaluate the maximized expected welfare W_x, we substitute the optimal attention shares back into the objective function:

$$\begin{aligned} W_x &= \sum_{j\in J} \lambda_{xj}\, m_{xj}\left(v_{xj} - P_{xj}\right) - \theta_x \sum_{j\in J} \lambda_{xj}\left[\frac{m_{xj}\left(v_{xj}-P_{xj}\right)}{\theta_x} - \ln Z_x\right] \\ &= \theta_x \ln Z_x \sum_{j\in J} \lambda_{xj} = \theta_x \ln Z_x . \end{aligned}$$

This establishes the closed-form inclusive value representation.

To derive the payoff semi-elasticity, we differentiate the natural logarithm of the sourcing share $\ln\lambda_{xj} = \ln\tilde{a}_{xj} + \frac{m_{xj}(v_{xj}-P_{xj})}{\theta_x} - \ln Z_x$ with respect to the price P_{xj}. We hold the meeting rate fixed:

$$z_{xj} \equiv -\frac{\partial \ln \lambda_{xj}}{\partial P_{xj}} = \frac{m_{xj}}{\theta_x} + \frac{1}{Z_x}\frac{\partial Z_x}{\partial P_{xj}}.$$

Using the derivative of the partition function $\frac{\partial Z_x}{\partial P_{xj}} = -\lambda_{xj} Z_x \frac{m_{xj}}{\theta_x}$, we obtain:

$$z_{xj} = \frac{m_{xj}}{\theta_x} - \lambda_{xj}\frac{m_{xj}}{\theta_x} = \frac{m_{xj}}{\theta_x}\left(1-\lambda_{xj}\right).$$

This completes the proof. □

A.2 Proof of Proposition 2.2 (Seller Pricing)

The seller maximizes operating profit $V_{xj}(P_{xj} - \tau_{xj} c_j)$, where realized match volume is $V_{xj} \equiv \mu_x \lambda_{xj} m_{xj}$. Under the aggregate matching technology, the meeting rate is $m_{xj} = A_{xj}\lambda_{xj}^{\alpha-1}$. A seller internalizing congestion at intensity ρ perceives the volume elasticity:

$$\frac{\partial \ln V_{xj}}{\partial \ln \lambda_{xj}} = \eta \equiv 1 - \rho(1-\alpha), \qquad \frac{\partial \ln m_{xj}}{\partial \ln \lambda_{xj}} = \eta - 1.$$

Lowering the price P_{xj} attracts attention, congests the submarket, and lowers the meeting rate. This feedback dampens the attention response. We differentiate the buyer's rule $\ln\lambda_{xj} = \ln\tilde{a}_{xj} + m_{xj}(v_{xj} - P_{xj})/\theta_x - \ln Z_x$ while carrying this feedback. Let

$D_{xj} \equiv -\partial \ln \lambda_{xj} / \partial P_{xj}$ denote the perceived demand slope:

$$D_{xj} \;=\; z_{xj}\Big[1 + \big(\eta - 1\big)\big(v_{xj} - P_{xj}\big)D_{xj}\Big] \quad\Longrightarrow\quad D_{xj} \;=\; \frac{z_{xj}}{1 + \big(1-\eta\big)\, z_{xj}\,\big(v_{xj} - P_{xj}\big)}.$$

Differentiating the profit objective $V_{xj}(P_{xj} - \tau_{xj}c_j)$ with respect to P_{xj} and dividing by V_{xj} yields the first-order condition:

$$\frac{\partial \ln V_{xj}}{\partial \ln \lambda_{xj}} \frac{\partial \ln \lambda_{xj}}{\partial P_{xj}}\big(P_{xj} - \tau_{xj}c_j\big) \;+\; 1 \;=\; 0 \quad\Longrightarrow\quad \eta\, D_{xj}\big(P_{xj} - \tau_{xj}c_j\big) \;=\; 1.$$

Substituting the expression for D_{xj} and clearing the denominator gives:

$$\eta\, z_{xj}\big(P_{xj} - \tau_{xj}c_j\big) \;=\; 1 \;+\; \big(1-\eta\big) z_{xj}\big(v_{xj} - P_{xj}\big).$$

Dividing by z_{xj} and collecting the P_{xj} terms yields:

$$\eta\, P_{xj} \;+\; (1-\eta) P_{xj} \;=\; \eta\, \tau_{xj}c_j \;+\; (1-\eta) v_{xj} \;+\; \frac{1}{z_{xj}}.$$

Adding and subtracting $(1-\eta)\tau_{xj}c_j$ from the right hand side isolates the marginal cost:

$$P^*_{xj} \;=\; \tau_{xj}c_j \;+\; (1-\eta)\big(v_{xj} - \tau_{xj}c_j\big) \;+\; \frac{1}{z_{xj}}.$$

This completes the derivation of the optimal posted price (2.9). The congestion feedback converts the multiplicative dial η on the margin into a surplus sharing rule, leaving the relationship markup $1/z_{xj}$ additive and undampened. $\square$

A.3 Proof of Proposition 2.4 (Equilibrium Potential)

We verify that the gradient of $\Phi_\rho(\lambda)$ on the strategy space reproduces the equilibrium system. Because the aggregate matching function is Cobb–Douglas in buyer attention ($M_{xj} \propto \lambda^\alpha_{xj}$), its derivative is $\frac{\partial M_{xj}}{\partial \lambda_{xj}} = \alpha \frac{M_{xj}}{\lambda_{xj}} = \alpha \mu_x m_{xj}$. Differentiating the potential function Φ (Equation 2.10) with respect to λ_{xj} (with Lagrange multiplier $\mu_x \frac{\alpha}{\eta} \theta_x \xi^\Phi_x$ on $\sum_j \lambda_{xj} = 1$) and dividing by $\mu_x \frac{\alpha}{\eta}$ yields:

$$\eta\,\big(v_{xj} - \tau_{xj}c_j\big)\, m_{xj} \;-\; \frac{\theta_x}{1 - \lambda_{xj}} \;-\; \theta_x\Big(\ln \frac{\lambda_{xj}}{\tilde{a}_{xj}} + 1\Big) \;=\; \theta_x\, \xi^\Phi_x.$$

From the buyer's semi-elasticity, $\frac{\theta_x}{1-\lambda_{xj}} = \frac{m_{xj}}{z_{xj}}$. Factoring out m_{xj} from the first two terms gives:

$$m_{xj}\Big[\eta\big(v_{xj} - \tau_{xj}c_j\big) \;-\; \frac{1}{z_{xj}}\Big] \;-\; \theta_x\Big(\ln\frac{\lambda_{xj}}{\tilde{a}_{xj}} + 1\Big) \;=\; \theta_x\,\xi_x^{\Phi}.$$

From the seller's equilibrium pricing contract (2.9), the buyer's net surplus is $v_{xj} - P^*_{xj} = \eta(v_{xj} - \tau_{xj}c_j) - 1/z_{xj}$. Substituting this back directly yields:

$$m_{xj}\,\big(v_{xj} - P^*_{xj}\big) \;-\; \theta_x\Big(\ln\frac{\lambda_{xj}}{\tilde{a}_{xj}} + 1\Big) \;=\; \theta_x\,\xi_x^{\Phi}.$$

This coincides with the first-order condition of the buyer's optimal sourcing problem.

The potential function $\Phi(\lambda)$ is the sum of three terms. The physical surplus term $\sum(v_{xj} - \tau_{xj}c_j)M_{xj}$ is strictly concave on the active set because M_{xj} exhibits decreasing returns to scale with respect to individual buyer attention λ_{xj} ($\alpha \in (0,1)$). The captivity penalty $\sum \ln(1 - \lambda_{xj})$ is also strictly concave in λ. The negative relative entropy costs are strictly concave in λ on the relative interior of the simplex. The sum is therefore strictly concave on the relative interior of the convex compact strategy space: a maximizer exists, is unique, and coincides with the unique interior solution of the first-order-condition system. □

A.4 Proof of Granularity Invariance and Aggregation

Suppose an aggregate sourcing option (such as an origin country) with total prior mass $\tilde{A}$ is partitioned into $K \geq 1$ symmetric, independently optimizing exporting firms. The buyer views each firm k as a separate option with price P_k and prior $\tilde{a}_k = \tilde{A}/K$. Under symmetry, the buyer directs equal attention $\lambda_k = \ell = \Lambda/K$ to each firm, where Λ is the aggregate share of the entire group. The relative entropy cost across all K firms within this group is:

$$C_x(\lambda) \;=\; \theta_x \sum_{k=1}^{K} \frac{\Lambda}{K} \ln\frac{\Lambda/K}{\tilde{A}/K} \;=\; \theta_x\,\Lambda \ln\frac{\Lambda}{\tilde{A}}.$$

The sub-option count K cancels completely inside the logarithm. Dividing an aggregate category into symmetric sub-options leaves the underlying attention price θ_x and aggregate information cost unchanged.

The semi-elasticity of buyer attention with respect to the price of individual firm k is:

$$z_k \;\equiv\; -\frac{\partial \ln \ell}{\partial P_k} \;=\; b\left(1-\ell\right) \;=\; b\left(1-\frac{\Lambda}{K}\right),$$

where $b \equiv m/\theta_x$. Now consider a uniform cost shock dC to all K symmetric suppliers in the origin. Because the shock is common, every member of the group reprices together by the same amount dP, while the unshocked rest of the world keeps prices fixed. The logit differential for each member is therefore $d\ln\ell = -b(dP - \Lambda dP) = -b(1-\Lambda)dP$.

Each supplier also reprices optimally through its markup. Differentiating the posted-price equation $P = C + (1-\eta)(v-C) + 1/[b(1-\ell)]$ under the common cost shock dC:

$$\begin{aligned}
dP \;&=\; \eta\, dC \;+\; \frac{1}{b}\,\frac{d\ell}{(1-\ell)^2} \;=\; \eta\, dC \;-\; \frac{\ell(1-\Lambda)}{(1-\ell)^2}\, dP,\\
\frac{dP}{dC} \;&=\; \eta\,\frac{(1-\ell)^2}{(1-\ell)^2+\ell(1-\Lambda)}.
\end{aligned}$$

The relative group-share response is $-b\tau_{xj}c_j\, dP/d\ln\tau_{xj}$. Substituting the directed elasticity $\varepsilon_D = \eta b\tau_{xj}c_j$ delivers the origin-wide elasticity $\varepsilon_i = \varepsilon_D\Pi(\ell,\Lambda)$ of Proposition 3.4. A shortcut that applies the one-relationship response to each member while holding the other members of the origin fixed omits this simultaneous within-origin repricing and does not aggregate correctly. □

B Endogenous Seller Screening and Separation

In the main text, we assume sellers possess enough capacity to accommodate incoming inquiries, subsuming seller capacity into the aggregate matching technology. In an extended formulation, the seller endogenously chooses its capacity allocation $q_j(x)$ subject to an entropy cost. Because the price equation conditions on the resulting matching state, the pricing results are mathematically identical whether screening capacity is endogenous or subsumed. This section proves that separation.

Let the seller allocate limited screening capacity $q_j(x)$ across buyer types subject to an exogenous benchmark distribution $\tilde{s}_{xj}$. The capacity restriction is $\sum_x \tilde{s}_{xj}q_j(x) = 1$. The seller pays a relative entropy cost $\theta_j \sum_x \tilde{s}_{xj}q_j(x)\ln q_j(x)$, where $\theta_j > 0$ is its price of screening capacity. The meeting rate becomes $m_{xj} = A_{xj}\lambda_{xj}^{\alpha-1}q_j(x)^{1-\alpha}$, of which the seller perceives the elasticity $1-\rho\alpha$ with respect to $q_j(x)$.

The seller jointly chooses P_{xj} and $q_j(x)$ to maximize profit. Because the first-order condition for P_{xj} is taken holding $q_j(x)$ (and thus m_{xj}) fixed in the partial derivative, the derivation of P^*_{xj} in Appendix A holds, yielding $P^*_{xj} = \tau_{xj}c_j + (1-\eta)(v_{xj} - \tau_{xj}c_j) + 1/z_{xj}$.

Taking the derivative of the seller's profit objective with respect to $\ln q_j(x)$ yields:

$$\begin{aligned}\frac{\partial \mathcal{L}_j}{\partial \ln q_j(x)} \;&=\; \mu_x\,\lambda_{xj}\,m_{xj}\Big((1-\rho\alpha)+\eta\,L_{xj}\Big)\big(P_{xj}-\tau_{xj}c_j\big) \\ &\quad -\;\mu_j\,\tilde{s}_{xj}\,q_j(x)\Big[\theta_j\big(\ln q_j(x)+1\big)+\zeta_j\Big] \;=\; 0,\end{aligned}$$

where ζ_j is the multiplier on the capacity restriction and $L_{xj} \equiv \partial \ln \lambda_{xj}/\partial \ln q_j = z_{xj}(v_{xj} - P_{xj})(1-\rho\alpha)/[1 + (1-\eta)z_{xj}(v_{xj} - P_{xj})]$.

Evaluated at the optimal price P^*_{xj}, the first product simplifies significantly. We know that $P^*_{xj} - \tau_{xj}c_j = (1-\eta)(v_{xj} - \tau_{xj}c_j) + 1/z_{xj}$. Substituting this margin completely cancels the endogenous attention-elasticity term L_{xj}, collapsing the bracket to:

$$\Big((1-\rho\alpha)+\eta\,L_{xj}\Big)\big(P^*_{xj}-\tau_{xj}c_j\big) \;=\; (1-\rho\alpha)\big(v_{xj}-\tau_{xj}c_j\big).$$

The seller's optimal screening return relies purely on the primitive joint surplus $v_{xj} - \tau_{xj}c_j$, entirely independent of its own price markup. The share of surplus conceded to the buyer to offset congestion is perfectly balanced by the additional attention that concession attracts.

Consequently, the first-order condition for screening separates cleanly from the pricing decision. The cross-sectional pricing behavior, tariff pass-through schedules, and macro-aggregation rules are invariant to whether q is dynamically chosen by the seller or treated as a fixed component of the matching function.

C Efficiency and the Harberger Wedge

At the full-congestion benchmark ($\rho = 1 \implies \eta = \alpha$) and in the absence of fiscal tariffs, we compare the core potential to the allocation chosen by a utilitarian world planner. The planner values match surplus minus information costs:

$$\Omega^{\text{plan}}(\lambda) \;=\; \sum_{xj}\big(v_{xj}-\tau_{xj}c_j\big)M_{xj}(\lambda) \;-\; \sum_x \mu_x\theta_x \sum_j \lambda_{xj}\ln\frac{\lambda_{xj}}{\tilde{a}_{xj}}. \tag{C.1}$$

Because $\eta = \alpha$ at this benchmark, the market potential $\Phi(\lambda)$ decomposes as:

$$\Phi(\lambda) \; = \; \Omega^{\text{plan}}(\lambda) \; + \; \sum_{xj} \mu_x \theta_x \ln(1 - \lambda_{xj}). \tag{C.2}$$

The second term $\mathcal{P}(\lambda) = \sum_{xj} \mu_x \theta_x \ln(1 - \lambda_{xj})$ is the captivity penalty. It captures the reduced-form footprint of relationship pricing. The gradient of the captivity penalty links directly to the physical relationship markup: $\frac{\partial \mathcal{P}}{\partial \lambda_{xj}} = -\frac{\mu_x \theta_x}{1-\lambda_{xj}} = -(M_{xj}/\lambda_{xj})(1/z_{xj})$.

A second-order Taylor expansion around the planner allocation yields the deadweight welfare loss:

$$\mathcal{L}^{\text{DWL}} \; \simeq \; \frac{1}{2} \nabla \mathcal{P}' \, \mathcal{H}^{-1} \, \nabla \mathcal{P}, \tag{C.3}$$

where $\mathcal{H} = -\nabla^2 \Omega^{\text{plan}}$ is the planner's Hessian projected onto the simplex tangent space. The second-order loss is therefore a quadratic form in the pricing wedge, weighted by the curvature of the planner's objective. A markup that is uniform across the options in a buyer's choice set cancels in the logit normalization, so dispersion in relationship markups is the unique source of the allocative loss. Opening trade improves this allocation by compressing the relative relationship wedges faced by domestic buyers.

D Elasticities, Pass-Through, Horizon Dynamics

D.1 Ricardian Rents in the Frictionless Limit

Multiply the dominance equation (3.1) by θ_x / m_{xj^*}. As $\theta_x \to 0$, the terms $\frac{\theta_x}{m_{xj^*}} \ln \tilde{a}$ and $\frac{\theta_x}{m_{xj^*}} \ln \lambda$ vanish. By the log-sum-exp limit, $\frac{\theta_x}{m_{xj^*}} \ln Z_{x,-j^*} \to \max_{k \neq j^*} \frac{m_{xk}}{m_{xj^*}} \eta(v_{xk} - \tau_{xk} c_k)$. Since losing sellers' captivity vanishes ($\lambda_{xk} \to 0$), their markups vanish and $v_{xk} - P_{xk} \to \eta(v_{xk} - \tau_{xk} c_k)$. Collecting terms yields the positive limit markup for the winner:

$$\frac{1}{z_{xj^*}} \;\longrightarrow\; \eta \left(v_{xj^*} - \tau_{xj^*} c_{j^*} \right) \; - \; \frac{m_{xj_2}}{m_{xj^*}} \, \eta \left(v_{xj_2} - \tau_{xj_2} c_{j_2} \right).$$

D.2 Cost Pass-Through Schedule

Total differentiation of the posted price $P_{xj} = \tau_{xj} c_j + (1-\eta)(v_{xj} - \tau_{xj} c_j) + 1/z_{xj}$, holding matching rates fixed, yields:

$$dP_{xj} \; = \; d(\tau_{xj} c_j) \; - \; (1-\eta) d(\tau_{xj} c_j) \; + \; d\Big(\frac{1}{z_{xj}}\Big) \; = \; \eta \, d(\tau_{xj} c_j) \; + \; d\Big(\frac{1}{z_{xj}}\Big).$$

Differentiating the markup term with respect to the attention share λ_{xj} gives $d(1/z_{xj}) = -[\lambda_{xj}/(1-\lambda_{xj})]dP_{xj}$. Substituting this back into the price differential and factoring out dP_{xj} yields $dP_{xj} = \eta(1-\lambda_{xj})d(\tau_{xj}c_j)$. □

D.3 Home-Versus-Foreign Elasticity with Arbitrary Portfolios

Differentiating the buyer logit for option j under contract changes $\{dP_k\}$ gives $d\ln\lambda_j = -b_j dP_j + \sum_k \lambda_k b_k dP_k$, with supplier-specific $b_j = m_j/\theta_x$ and delivered costs C_j satisfying $\eta b_j C_j = \varepsilon_D$. A domestic supplier faces no direct cost shock and reprices only through its markup, $b_j dP_j = d\lambda_j/(1-\lambda_j)^2$. A foreign supplier adds the common cost term, $b_j dP_j = \varepsilon_D d\ln\tau + d\lambda_j/(1-\lambda_j)^2$. Substituting, and writing $S \equiv \sum_k \lambda_k b_k dP_k$ and $u \equiv \varepsilon_D d\ln\tau$:

$$d\lambda_j\Big[\frac{1}{\lambda_j} + \frac{1}{(1-\lambda_j)^2}\Big] \;=\; S \;-\; u\,\mathbf{1}\{j\in F\} \quad\Longrightarrow\quad d\lambda_j \;=\; \big(S - u\,\mathbf{1}\{j\in F\}\big)\,\lambda_j\,\Pi(\lambda_j),$$

using $1/\Pi(x) = 1 + x/(1-x)^2$. Let $B_h = \lambda_{nn}\bar{\Pi}_h$ and $B_F = (1-\lambda_{nn})\bar{\Pi}_F$ denote the block sums of $\lambda_j\Pi(\lambda_j)$. Adding up within blocks, $d\lambda_{nn} = SB_h$ and $-d\lambda_{nn} = (S-u)B_F$, so $S = uB_F/(B_h+B_F)$ and $d\lambda_{nn} = u\,B_hB_F/(B_h+B_F)$. Dividing by $\lambda_{nn}(1-\lambda_{nn})\,d\ln\tau$ delivers the macroeconomic elasticity of Proposition 3.5. The within-block share-evolution equation follows as $d\lambda_j/d\lambda_{nn} = \lambda_j\Pi(\lambda_j)/B_h$ for $j\in h$. In the symmetric case every domestic share equals $\lambda_{nn}H^h$, so $\bar{\Pi}_h = \Pi(\lambda_{nn}H^h)$, and similarly for the foreign block. □

D.4 Elasticity Term Structure

Consider a two-option choice environment between a shocked origin and an unshocked domestic alternative. The within-period relative share responds to a tariff shock as $x_t = (1-\lambda)(A_t+u)$, where u is the payoff perturbation and A_t is the relative log deviation of the awareness prior. Linearizing the prior's law of motion around a symmetric steady state yields $A_{t+1} = (1-\delta-\nu)A_t + \delta x_t$; for a small relationship share, $x_t \simeq A_t + u$, and the recursion becomes a first-order linear difference equation:

$$A_{t+1} \;=\; (1-\nu)A_t \;+\; \delta\,u.$$

With initial condition $A_0 = 0$, the solution at horizon h is $A_h = \frac{\delta u}{\nu}[1 - (1 - \nu)^h]$. Substituting A_h back into the share response yields the cumulative horizon profile ε_h.

□

E Welfare Derivations and Details

E.1 Welfare Derivative and the Price-Ceiling Branch

Taking logs of the sourcing rule (2.4) for a domestic reference option h and using the inclusive value $W_x = \theta_x \ln Z_x$ gives the domestic anchor identity $W_x = m_{xh}\big(v_{xh} - P_{xh}\big) - \theta_x \ln \lambda_{xh} + \theta_x \ln \tilde{a}_{xh}$. Differentiating this identity holding meeting rates, the awareness prior, and attention prices fixed yields $dW_x = -m_{xh} dP_{xh} - \theta_x d \ln \lambda_{xh}$. The domestic seller faces no direct cost shock. Its price moves only through the relationship markup. Substituting the markup derivative yields:

$$dW_x \;=\; -\theta_x \Big[1 + \frac{\lambda_{xh}}{(1 - \lambda_{xh})^2}\Big] d \ln \lambda_{xh} \;=\; -\frac{\theta_x}{\Pi(\lambda_{xh})}\, d \ln \lambda_{xh}.$$

Along the equilibrium path, every domestic supplier satisfies $d \ln \lambda_{xh} = S\, \Pi(\lambda_{xh})$ with S common across suppliers. This implies $dW_x = -\theta_x S$, independent of the chosen anchor. Substituting $S = d \ln \lambda_{nn} / \bar{\Pi}_h$ delivers the aggregate welfare derivative. □

At the price ceiling the same algebra switches branch. The seller's interior pricing candidate is $P^* = \tau c + (1 - \eta)(v - \tau c) + 1/z$. When this candidate implies a negative net surplus for the buyer ($v - P^* < 0$), the voluntary trade constraint $P \leq v$ of Assumption 1 binds, the price sits at the buyer valuation $P = v$, and complementary slackness makes $dP_j = 0$ for shocks that would push the price above v. A capped incumbent therefore behaves as a fixed-price option: its damping weight switches from $\Pi(\lambda_j)$ to one, and the anchor identity directly delivers $dW_x = -\theta_x\, d \ln \lambda_{nn}$. In terms of the surplus-to-cost ratio $r = (v - \tau c)/\tau c$, the interior condition for a domestic relationship with share λ_j is $r > 1/[\varepsilon_D(1 - \lambda_j)]$; along the symmetric-family path it first fails at $\lambda^*_{nn} = [1 - 1/(r\varepsilon_D)]/H^h$, and the welfare integration switches to the ceiling branch at that point. With an asymmetric portfolio, incumbents reach the ceiling at different shares and the block formula is re-solved with mixed damping weights; the headline calculation uses the symmetric-equivalent continuation.

E.2 Closed-Form Linear Benchmark

The numerical integral of the nonlinear home-versus-foreign schedule has no closed-form solution. We present the analytic result for its linear counterpart $\varepsilon(\lambda) = \varepsilon_D(1 - b\lambda)$, where $b \equiv H^h \in (0,1)$ represents the domestic sourcing concentration. Integrating the local ACR identity from autarky ($\lambda = 1$) to the trade equilibrium ($\lambda = \lambda_{nn}$) yields:

$$\ln\Big(\frac{\mathcal{W}_n}{\mathcal{W}_n^{\text{aut}}}\Big) \;=\; \int_{\lambda_{nn}}^{1} \frac{d\lambda}{\lambda\, \varepsilon_D(1 - b\lambda)} \;=\; \frac{1}{\varepsilon_D}\int_{\lambda_{nn}}^{1} \Big[\frac{1}{\lambda} + \frac{b}{1 - b\lambda}\Big] d\lambda.$$

Evaluating the integral using partial fractions gives $\ln\lambda - \ln(1 - b\lambda)$. Exponentiating produces the captivity-corrected gains formula:

$$\frac{\mathcal{W}_n}{\mathcal{W}_n^{\text{aut}}} \;=\; \lambda_{nn}^{-1/\varepsilon_D}\Big(\frac{1 - b\lambda_{nn}}{1 - b}\Big)^{1/\varepsilon_D}.$$

For this linear schedule the arc elasticity always lies below the open-economy point elasticity; the nonlinear two-block schedule used in the main text does not obey that ordering in general.

E.3 Historical Comparisons

The welfare calculation holds the directed elasticity ε_D and the domestic concentration H^h fixed along the path.

Historical estimates provide rough scale comparisons for the calculated gains from trade. Irwin (2005) estimates that the near-total shutdown of US foreign commerce under the 1807 Embargo Act reduced national income by approximately 5 percent. Bernhofen and Brown (2005) evaluate Japan's nineteenth-century opening, obtaining an empirical upper bound on static gains of about 8 to 9 percent of GDP. These national-income objects are not directly comparable to the buyer-side consumption-equivalent statistic reported in the main text; they serve as historical scale references.

E.4 Path Tracking for Asymmetric Portfolios

The analytical welfare integral holds the domestic concentration index H^h fixed. This assumes the damping statistic evolves according to $\Pi(\lambda_{nn}H^h)$. An asymmetric portfolio naturally de-concentrates as expenditure returns home, because smaller incumbents absorb returning demand more elastically than the dominant incumbent.

We track a literal two-supplier domestic portfolio calibrated to the same impact damping factor as the symmetric index 0.86 (at a domestic share of 0.76). This requires an initial split of 94/6 between the dominant and secondary supplier. Integrating the equilibrium share system continuously to autarky, the secondary supplier gains ground. The portfolio ends at an 80/20 split. The resulting consumption-equivalent gain using the asymmetric path is 8.8 percent, compared to 12.5 percent for the symmetric family. Alternative asymmetric continuations matched to the same impact damping give similar values: a three-supplier 94/3/3 portfolio yields 8.7 percent, and a dominant supplier with a twenty-firm fringe yields 8.6 percent. These values lie within the 90 percent credible interval reported in the main text, so portfolio-structure uncertainty is comparable to the reported parameter uncertainty.

F Empirical Estimation and Robustness

F.1 Joint Estimation Mechanics and the Projection Mapping

We formulate the parameter estimation using a quasi-Bayesian objective

$$Q(\vartheta) = Q_P(\eta, r, \varepsilon_D) + Q_M(\mu_H, \tau_H, \sigma_e).$$

The pricing block is the quadratic form

$$Q_P \;=\; \tfrac{1}{2}\big[\hat{\beta} - m(\Gamma, \varepsilon_D)\big]' \, \hat{V}^{-1}\big[\hat{\beta} - m(\Gamma, \varepsilon_D)\big],$$

where $\hat{\beta}$ is the estimated tariff coefficient vector, $\hat{V}$ its full HS2-cluster covariance matrix, and $m(\cdot)$ the model-implied coefficient mapping defined below. The pass-through schedule $\pi(\lambda)$ is a nonlinear function of the relationship share. The fixed effects in the regression spread the identifying variation across a range of shares. Evaluating the theoretical schedule at a share of zero misstates the estimand.

We write the structural pass-through schedule in partial fractions:

$$\pi(\lambda) \;=\; a_0 \;+\; a_1\lambda \;+\; a_2\,\frac{1}{1-\lambda+c}, \qquad c = \frac{1}{\Gamma\varepsilon_D},$$

with coefficients $a_0 = -1 + 1/\Gamma - 1/(\varepsilon_D\Gamma^2)$, $a_1 = -1/\Gamma$, and $a_2 = 1/(\varepsilon_D^2\Gamma^3)$. We define the pseudo-true coefficient vector of the regression by mathematically projecting these basis functions onto the residualized design matrix $\tilde{X}$:

$$m(\Gamma, \varepsilon_D) \;=\; a_0 A_0 \;+\; a_1 A_1 \;+\; a_2 A(c), \qquad A(w) \equiv (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\big[w(\lambda)\ln(1+t)\big].$$

The projection matrices A_0, A_1, and $A(c)$ are computed directly from the data. This provides an analytical mapping from any parameter value to the regression coefficients.

The micro–macro block treats the 28 paired elasticity gaps $y_c \equiv \ln(\varepsilon_{macro,c}/\varepsilon_{micro,c})$ as noisy observations of the model's damping factor with country-specific concentration. The concentration parameters H_c^h are drawn from a logit-normal distribution governed by hyperparameters μ_H and τ_H, the residual noise $e_c \sim N(0, \sigma_e^2)$ absorbs sampling error in the published elasticities, and the country effects are integrated out by quadrature:

$$e^{-Q_M} \;=\; \prod_{c=1}^{28} \int_0^1 \varphi_{\sigma_e}\big(y_c - \ln D_c(H)\big)\, dF_{\mu_H,\tau_H}(H),$$

where D_c is the damping factor at country c's domestic share, φ_{σ_e} is the normal density, and F_{μ_H,τ_H} is the logit-normal concentration distribution.

F.2 Priors, Restrictions, and Computation

The priors are $\eta \sim U[\alpha, 1]$ with $\alpha = 1/2$, the technological bounds of the internalization elasticity; $r \sim U[0, 3]$ restricted to the interior-pricing region $r > 1/\varepsilon_D$; and $\varepsilon_D \sim N(5.38, 0.5^2)$, an empirical-Bayes anchor centered on the median sector-based estimate across the 24 WIOD-matched economies, with a standard deviation exceeding the bootstrap standard error of that median (0.37). The hyperpriors are $\mu_H \sim N(1.5, 1.5^2)$ on the logit scale, τ_H half-normal with scale 0.75, and σ_e half-normal with scale 0.5. Because the pricing block involves only (η, r, ε_D) and the concentration block only the hyperparameters, the quasi-posterior factorizes, and each factor is computed on a fine grid: $251 \times 301 \times 161$ points for the pricing parameters and $91 \times 50 \times 48$ hyperparameter points with a 500-point quadrature grid in H. Posterior summaries use 40,000 draws with within-cell jittering; the posterior for H_{US}^h is drawn per hyperparameter draw from its conditional density given the US gap.

The pricing likelihood identifies the composite Γ; the split between η and r away from the boundary $\Gamma = 1$ is resolved by the structural bounds and the prior on r. Widening the upper support of the r prior from three to six moves the posterior median of η from 0.97 to 0.98. When the posterior blocks are combined for welfare, one consistency restriction couples them: the observed macro damping requires interior domestic pricing at the observed state, $r > 1/[\varepsilon_D(1-\lambda_{nn}H^h_{US})]$, since a price-capped incumbent generates no repricing damping. This removes about a quarter of the draws, almost all with low prior-driven values of r. Under the restricted posterior, the welfare path remains interior all the way to autarky with probability 0.58, and the median ceiling location lies above $\lambda_{nn} = 0.98$.

The projection mapping is exact for the local pass-through schedule evaluated at predetermined sourcing shares. Integrating the finite-change structural response observation by observation and projecting it through the same residualized design moves the model-implied coefficient pair by only 0.06 and 0.41 cluster standard errors at trade-war tariff levels, in the direction of the estimated interaction, so the local mapping is adequate for the estimation sample.

F.3 Robustness Variants and Specification Checks

Table F.1 reports the joint posterior across alternative specifications. The pass-through interaction magnitude depends on sample construction: -0.81 in the baseline panel and -0.21 in the sample restricted to cells present in both the US and the EU; China supplies about 61 percent of the coefficient-specific identifying variation in the baseline interaction (Table F.4). The joint estimation reflects this through the full covariance matrix and the exclusion variant.

A common-concentration least-squares fit of the damping formula to the 28 gaps gives 0.85 with a nonparametric bootstrap 90 percent interval $[0.81, 0.88]$, consistent with the hierarchy's population median. Halving or doubling the prior scales for τ_H and σ_e, or tripling the dispersion of the μ_H prior, moves the H^h_{US} median by less than 0.01 and the US gain by less than 0.2 percentage points. The frequentist counterpart of the pricing block is a constrained minimum-distance fit, which reaches the boundary $\Gamma = 1$ in every sample; the minimized distances, referred conservatively to χ^2, do not reject (linear moments: $J = 1.73$, $p = 0.19$; excluding China: $J = 3.26$, $p = 0.07$; share bins with full covariance: $J = 1.53$ on three degrees of freedom, $p = 0.68$, for the first four bins, and $J = 3.37$ on four, $p = 0.50$, with the top bin included).

Variant	η	μ_0	ε_D	H^h_{US}	ε_{macro}	Gain (%)
Baseline	0.97 $[0.87, 1.00]$	1.21	5.46	0.86	2.34	12.6 $[8.7, 22.0]$
Excl. China	0.96 $[0.83, 1.00]$	1.22	5.45	0.86	2.34	12.6 $[8.7, 22.0]$
Share bins 1–4	0.98 $[0.90, 1.00]$	1.20	5.47	0.86	2.34	12.5 $[8.7, 22.0]$
FGO elasticity prior	0.96 $[0.86, 1.00]$	1.15	8.35	0.86	3.57	8.0 $[5.6, 14.1]$
US-specific ε_D prior	0.97 $[0.88, 1.00]$	1.23	5.01	0.86	2.14	13.8 $[9.5, 24.3]$
$H^F = 0.3$	0.97 $[0.87, 1.00]$	1.21	5.46	0.85	2.41	12.2 $[8.5, 21.0]$
WIOD shares, 24 economies	0.97 $[0.87, 1.00]$	1.21	5.46	0.89	2.16	14.5 $[8.9, 31.6]$

Table F.1: Robustness of the joint quasi-posterior. Entries are posterior medians; brackets are 90 percent credible intervals for η and the state-dependent US gain. Parameter entries are the unrestricted joint posterior, and gains are computed on the interior branch for comparability across variants; the observed-state interiority restriction and the price-ceiling branch used for the main-text headline trim the baseline to 11.7 $[8.2, 18.5]$. "Share bins" replaces the two linear moments with the share-bin coefficient vector and its full cluster covariance. The "FGO elasticity prior" centers the elasticity prior on the tariff-based estimates of Fontagné et al. (2022); the "US-specific ε_D prior" centers it on the US sector-based estimate 4.91; the last row re-estimates the concentration hierarchy on the 24 WIOD-matched economies with WIOD domestic shares. The concentration posterior is unaffected by the pricing variants because the posterior factorizes.

F.4 Pass-Through Bins

Estimating the pass-through regression separately across discrete share bins broadly follows the predicted steepening of the structural schedule through the fourth bin, while the top bin flattens; the covariance-based test below does not reject the schedule. The estimated FOB price responses across five unweighted mean shares (0.005, 0.09, 0.23, 0.47, 0.80) are -0.060, -0.120, -0.281, -0.557, and -0.410. The schedule steepens through the fourth bin and flattens out for the highest-share relationships. At the posterior median ($\Gamma = 1.07, \varepsilon_D = 5.47$), the projected model values for these bins are $-0.21, -0.29, -0.41, -0.63$, and -0.86. Every empirical bin sits within 1.5 standard errors of its projected model value.

F.5 Third-Market Reallocation and the EU Control

We employ European Union imports as a third-market comparison. In a stacked panel, US relative prices fall significantly in response to the tariff. Because the EU leg itself responds to the treatment, we read these estimates as descriptive rather than as a difference-in-differences design. In the extension of Appendix B, screening capacity is rivalrous across destinations. When a US tariff lowers the payoff of the US market, the optimal screening rule shifts capacity toward third markets.

Capacity withdrawn from the taxed destination thins the US market and thickens the EU market. The model predicts this moves relationship markups in opposite directions. Table F.2 isolates these two channels by estimating the price response on each leg separately. We interact the US tariff with the origin's US product share (buyer captivity) and the origin's export dependence on the US market (seller reallocation).

	US leg		EU leg	
Dependent variable: $\ln uv_{ikt}$	(1)	(2)	(3)	(4)
$\ln(1+\tau_{ikt}^{\mathrm{US}})$	-0.024	-0.223^{*}	0.062	0.248^{***}
	(0.085)	(0.123)	(0.056)	(0.075)
$\ln(1+\tau_{ikt}^{\mathrm{US}}) \times \lambda_{ik}^{\mathrm{pre}}$	-0.213	-0.289^{*}	-0.672^{***}	-0.601^{***}
buyer-side captivity	(0.168)	(0.166)	(0.122)	(0.121)
$\ln(1+\tau_{ikt}^{\mathrm{US}}) \times \sigma_{i\to\mathrm{US}}$		1.534^{**}		-1.428^{***}
seller-side reallocation		(0.707)		(0.293)
Origin × Product FE	Yes	Yes	Yes	Yes
Product × Year FE	Yes	Yes	Yes	Yes
Origin × Year FE	Yes	Yes	Yes	Yes
Observations	611,670	611,670	611,670	611,670

Table F.2: Separating buyer-side captivity from seller-side screening reallocation. Both legs are estimated on the same sample of cells present in both the US and the EU. Export dependence σ changes sign across destinations, consistent with rivalrous screening capacity. Standard errors clustered at the HS2 level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The export dependence coefficient carries a sign pattern consistent with the screening-reallocation mechanism: positive on the US leg (+1.534), where withdrawn capacity would thin the market, and negative on the EU leg (−1.428), where redirected capacity would thicken it. Because the EU comparison market responds to the treatment through this channel, differencing the US leg against the EU leg nets the two effects against each other. Four features keep these estimates descriptive: the both-destination sample conditions on joint survival in the two markets; the buyer-side interaction exhibits a pre-trend in this selected sample; the country-wide dependence result does not survive value weighting (Table F.3); and a product-specific dependence measure closer to the structural screening object is weak. The EU leg was also subject to contemporaneous European steel-safeguard measures (World Trade Organization, 2022).

	(1) Baseline	(2) Excl. China	(3) σ 2015–17 avg	(4) Cluster Origin	(5) Balanced Cells	(6) Value-weighted
Panel A. US leg	*Dependent variable:* $\ln uv_{ikt}^{\mathrm{US}}$					
$\ln(1+\tau)\times\lambda^{\mathrm{pre}}$	-0.289^{*}	-0.993^{***}	-0.284^{*}	-0.289^{**}	-0.209	-0.292
	(0.166)	(0.359)	(0.166)	(0.125)	(0.164)	(0.444)
$\ln(1+\tau)\times\sigma_{i\to\mathrm{US}}$	1.534^{**}	1.442^{**}	1.441^{**}	1.534^{**}	1.669^{**}	-0.393
	(0.707)	(0.725)	(0.681)	(0.623)	(0.804)	(0.871)
Panel B. EU leg	*Dependent variable:* $\ln uv_{ikt}^{\mathrm{EU}}$					
$\ln(1+\tau)\times\lambda^{\mathrm{pre}}$	-0.601^{***}	-0.466	-0.602^{***}	-0.601^{***}	-0.662^{***}	-1.310^{***}
	(0.121)	(0.400)	(0.121)	(0.116)	(0.143)	(0.389)
$\ln(1+\tau)\times\sigma_{i\to\mathrm{US}}$	-1.428^{***}	-1.457^{***}	-1.408^{***}	-1.428^{***}	-0.851^{*}	-0.827
	(0.293)	(0.356)	(0.290)	(0.375)	(0.510)	(0.763)

Table F.3: Robustness of the screening-reallocation estimates. The sign reversal across destinations is stable across exclusions, clustering choices, and balanced panels in the unweighted specifications. The channel does not survive value weighting, indicating the reallocation operates primarily across the large number of smaller origin-product cells.

		China's share of . . .		
Design	Regressor	observations	import value	residual variation
Main US panel	$\ln(1+\tau)$	3.7%	19.3%	15.4%
	$\ln(1+\tau)\times\lambda^{\mathrm{pre}}$			61.3%
Stacked panel	$\ln(1+\tau)\times\mathbf{1}_{\mathrm{US}}$	6.6%	25.6%	32.4%
	$\ln(1+\tau)\times\mathbf{1}_{\mathrm{US}}\times\lambda^{\mathrm{pre}}$			77.1%
Single leg	$\ln(1+\tau)$	6.6%	24.0%	4.6%
	$\ln(1+\tau)\times\lambda^{\mathrm{pre}}$			75.3%
	$\ln(1+\tau)\times\sigma_{i\to\mathrm{US}}$			10.6%

Table F.4: China's leverage in each estimation sample. The last column is China's share of the sum of squared residuals of the regressor after partialling out both the design's fixed effects and every other regressor. Chinese cells are a small share of observations everywhere, but dominate the variation identifying the buyer-side interactions.

G Optimal Tariffs and Strategic Leverage

In standard neoclassical trade policy, the terms-of-trade optimal tariff is $t^{*} = 1/\varepsilon^{x}$, where ε^{x} is the foreign export-supply elasticity. Broda et al. (2008) document that inverse export-supply elasticities are finite and heterogeneous, and that countries unconstrained by the WTO set systematically higher tariffs where those elasticities are larger. In the extension of Appendix B, ε^x is an endogenous equilibrium object determined by the foreign sellers' optimal allocation of capacity.

When foreign exporters optimize capacity across destinations, their supply elasticity to a specific market depends on their export dependence on that market. We parameterize the export-supply elasticity as $\varepsilon^x = C(1-\sigma)$, where C is an aggregate thickness composite and σ is the foreign export dependence on the destination. Exporters concentrated on a single destination ($\sigma \to 1$) possess limited capacity to redirect sales elsewhere. They supply the destination inelastically.

Imposing a tariff restricts market access and alters the exporter's outside alternatives. We represent this dependence response by the linear rule $\sigma(t) = \sigma_0 + \kappa t$, where κ represents the cornering intensity. The terms-of-trade optimal tariff is the solution to an endogenous fixed-point problem.

Proposition G.1 (Optimal Tariff Fixed Point)**.** *The terms-of-trade optimal tariff satisfies the condition* $t^* = 1/[C(1-\sigma_0-\kappa t^*)]$. *This corresponds to the quadratic equation:*

$$C\kappa\,(t^*)^2 \;-\; C(1-\sigma_0)\,t^* \;+\; 1 \;=\; 0. \tag{G.1}$$

In the diversification regime ($\kappa \leq 0$), a unique positive optimal tariff exists and lies below the constant-dependence benchmark $1/[C(1-\sigma_0)]$. *In the cornering regime ($\kappa > 0$), a finite interior optimal tariff exists if and only if the cornering intensity does not exceed the tipping threshold* $\kappa^\dagger \equiv C(1-\sigma_0)^2/4$.

Proof. Substituting $\sigma(t) = \sigma_0 + \kappa t$ into the Lerner rule $t = 1/[C(1-\sigma(t))]$ produces the quadratic. The discriminant is $\mathcal{D} = C^2(1-\sigma_0)^2 - 4C\kappa$. For $\kappa > 0$, real roots exist if and only if $\mathcal{D} \geq 0$, which defines the threshold $\kappa^\dagger$. Solving the quadratic yields a cornering multiplier $\mathcal{M}(\hat{\kappa}) = 2(1-\sqrt{1-\hat{\kappa}})/\hat{\kappa}$, where $\hat{\kappa} = \kappa/\kappa^\dagger$. The optimal tariff factorizes into the baseline Lerner tariff multiplied by this cornering multiplier. □

The fixed point characterizes how terms-of-trade leverage varies with partner dependence: optimal leverage is largest against partners that direct most of their exports to the destination and smallest against diversified partners. We treat this as a comparative-static implication of the parameterization rather than as a quantitative account of observed tariff schedules. Observed schedules reflect much more than terms-of-trade optimality: political organization shapes protection even for governments that weight welfare far more heavily than contributions (Goldberg and Maggi, 1999), and terms-of-trade internalization through WTO negotiations is stronger where exporter concentration is greater (Ludema and Mayda, 2013).

Dynamic Restraint. Tariff leverage erodes over time. Imposing a tariff encourages the targeted partner to establish alternative commercial networks. This erodes future dependence according to the law of motion $\sigma_2 = \sigma_1 - d\,t_1$, where d represents the partner's diversification agility.

Let the hub maximize current tariff revenue $R(t)$ plus the discounted continuation value of market power $\delta_H V(\sigma_2)$. The first-order condition is $R_t(t_1; \sigma_1) - \delta_H V'(\sigma_2) d = 0$. To second order around the static optimum, the dynamically optimal tariff satisfies $t^{\text{dyn}} = t^{\text{static}} - [\delta_H d V'(\sigma_2)]/[-R_{tt}]$. Patient policymakers exercise tariff restraint to preserve partner dependence.

H The 2025 Tariff Surge and CPI Evaluation

Between January and September 2025, the United States executed a broad expansion of import tariffs. Applied rates rose to 15.6% on Mexico, 20.1% on Canada, and 54.4% on China. We evaluate the transmission of this tariff surge into consumer price inflation. Ignatenko et al. (2025) evaluate the same episode in a multi-country general-equilibrium model with flexible aggregate tariff pass-through; our decomposition complements theirs by letting absorption vary with relationship-level sourcing shares. The percentage change in consumer prices at horizon h decomposes as:

$$\begin{aligned}\Delta \ln P^{\text{cons}}(h) \;=\; & \beta_c \left(1 - \lambda_{nn}\right) \sum_{k,i} \varpi_{ik}(h) \Big[1 \;-\; a(h)\, \mathcal{A}_{ik} \Big] \Delta \ln(1 + \tau_{ik}) \\ & +\; \beta_c\, \lambda_{nn}\, \chi\, \Delta \ln P^M(h). \end{aligned} \tag{H.1}$$

The first term represents the direct imported-goods channel. The parameter β_c is the goods share of household consumption. The term $a(h)\mathcal{A}_{ik}$ captures the staggered phase-in of foreign tariff absorption. This relationship-level absorption operates strictly within origin–product cells, complementing cross-route general equilibrium channels, such as the shipping-cost decongestion highlighted by Bai et al. (2026).

The second term represents the domestic markup echo. When tariffs raise the delivered cost of foreign goods, the competitive discipline on domestic producers weakens. Domestic suppliers face a more captive customer base and expand their markups. We derive this echo coefficient directly from the pricing equation rather than importing it as a reduced-form parameter.

A rise in foreign delivered prices lowers the rest-of-world index facing each domestic incumbent. In a two-domestic-supplier benchmark ($K^h = 2$, $\lambda_{nn} = 0.76$, individual domestic share $\Lambda = 0.38$), the incumbent's outside attention set contains imports with conditional weight $w_M = 0.24/0.62 \simeq 0.39$ and the other domestic supplier with weight $w_D \simeq 0.61$; the other incumbent reprices as well, so the echo is a fixed point rather than a one-way pass. Writing $\Theta \equiv \Lambda/\mu_D$ for the conversion from the captivity wedge into the domestic price, with μ_M and μ_D the import and domestic price-cost ratios, the echo coefficient is:

$$\chi^{\text{model}} \equiv \frac{d \ln P_{\text{dom}}}{d \ln P^M} = \frac{\Theta\, w_M\, \mu_M}{1 - \Theta\, w_D\, \mu_D}.$$

Evaluating this expression at the US calibration ($\mu_M = 1.42$, $\mu_D = 1.45$) gives $\chi^{\text{model}} \simeq 0.19$. This benchmark uses two equal incumbents and anchor-row price-cost ratios rather than the estimated concentration index, and is indicative.

Origin	Share of US Imports	Statutory $\Delta\tau$ (pp)	Sourcing Captivity $\bar{\lambda}$	Foreign Absorption	Border-Price Rise Year 1 / Full (%)	Direct CPI Contrib. (pp)
China	0.149	38.0	0.46	0.43	22.0 / 15.5	0.13
Mexico	0.164	16.2	0.44	0.41	12.0 / 8.9	0.08
Canada	0.127	19.6	0.44	0.42	14.1 / 10.4	0.07
Japan	0.048	18.1	0.19	0.21	14.5 / 12.6	0.03
Germany	0.052	15.6	0.17	0.19	12.7 / 11.2	0.03
All	**1.000**	**19.1**	**0.30**	**0.30**	**13.7 / 10.9**	**0.55**

Table H.1: Consumer price index decomposition of the 2025 US tariff wave by partner country. Foreign absorption evaluates the schedule $\mathcal{A}_{ik} = 0.058 + 0.812\lambda_{ik}$. Statutory changes are import-weighted across products within each origin. The direct CPI contribution isolates the imported-goods channel, is evaluated at initial import weights over the first year, and excludes the domestic markup echo.

At initial import weights, the direct imported-goods contributions in the table sum to 0.55 percentage points over the first year. Allowing sourcing to divert away from tariffed origins along the elasticity phase-in path lowers the direct channel to 0.50 percentage points at the one-year point. The domestic markup echo adds 0.30 percentage points, for a first-year total of approximately 0.8 percentage points. Because foreign absorption is concentrated in high-share links, border prices for diversified origins (Germany, Korea) pass through at about 80 percent. Tariffs on diversified manufacturing origins therefore generate larger consumer price increases per dollar of statutory tariff than tariffs on concentrated links.

I Replication and Data Notes

BACI bilateral trade flows (Gaulier and Zignago, 2010) define the product categories at the HS6 level using the 2007 classification nomenclature. We filter low-value cells at a $1,000 threshold to reduce unit-value noise. Applied statutory tariffs combine the Global Tariff Database baseline (Teti, 2024) with the GTD US trade-war extension (Rodríguez-Clare et al., 2025). For the 2018–2024 estimation panel, daily changes are aggregated into day-weighted annual rates; the 2025 policy exercise uses the September 1, 2025 point-in-time snapshot.

The joint estimation utilizes three dedicated scripts. The first estimates the share-bin regression and stores the joint covariance matrix. The second computes the projection objects A_0, A_1, and $A(c)$ from the residualized design matrix. The third evaluates the two quasi-posterior blocks over a dense grid, draws from the posterior, and generates the resulting credible intervals for the macro elasticity and welfare gains.