Importers, Market Power and Optimal Tariffs*

Jonathan Becker† Corina Boar‡ Virgiliu Midrigan§

September 2026

Abstract

Importers are few and large, have higher labor productivity and pass through cost changes to prices incompletely. We study optimal tariffs in a model consistent with these facts. Firms pay a fixed cost to import and charge markups that increase with size. Market power implies that importers are too few and too small relative to the efficient allocations. Tariffs amplify this distortion. We derive a formula that relates the optimal tariff not only to the foreign export supply elasticity but also to how much distortions amplify the effect of trade costs on welfare. In our calibrated economy the optimal tariff is negative and decreases with country size.

Keywords: importers, variable markups, misallocation, optimal tariffs, gains from trade.

JEL codes: E10, F12, F13, L11, O40.

*We thank Suk Joon Kim for excellent research assistance.

†Stony Brook University, jonathan.becker@stonybrook.edu.

‡New York University and NBER, corina.boar@nyu.edu.

§New York University and NBER, virgiliu.midrigan@nyu.edu.

1 Introduction

Tariffs are back at the center of economic policy. The canonical rationale for tariffs is the terms-of-trade motive: when a country restricts import demand, foreign exporters lower their prices. These terms-of-trade gains must be weighed against the costs of restricting imports. A salient feature of the data is that importing is concentrated among a small number of very large firms. These firms have higher measured labor productivity and pass through cost changes to prices only incompletely, two features that are widely interpreted as evidence of market power. This interpretation suggests that imports are already too low, even absent tariffs, so tariffs may amplify existing distortions and be potentially costly.

Our paper studies the implications of these empirical regularities for optimal tariff policy. We study a model of international trade with two key features. First, heterogeneous firms choose whether to pay a fixed cost to import intermediate inputs. Second, firms charge markups that increase with their size. We show that firm market power implies that importers are too few and too small relative to the efficient allocations, and so is the volume of international trade. Tariffs considerably amplify these distortions, so they are far more costly than in existing models. Optimal policy therefore calls for an import subsidy, in contrast to existing work where the terms-of-trade motive typically calls for a positive tariff. The larger the country is, the costlier are tariffs and the larger the optimal subsidy. Abstracting from either selection into importing or variable markups understates the cost of tariffs and therefore leads to the conclusion that it is optimal to tax trade.

Our analysis builds on three well-known sets of facts. First, as Jones et al. (2025) document, most international trade is in intermediate goods. Second, importers are few and large. Only 5.5% of U.S. firms import, yet they account for nearly half of all employment (Bernard et al., 2007, 2009). Third, importers have 25% higher measured labor productivity than non-importers (Bernard et al., 2018), which is often interpreted as evidence of higher markups (Bernard et al., 2003). Consistent with this interpretation, Amiti et al. (2014, 2019a) document that importers pass through only 62% of changes in their costs to prices.

We develop our argument in two steps. We first study a static economy that allows us to transparently isolate the forces shaping the optimal tariff and to characterize it analytically. We then show that our conclusions extend to a richer dynamic economy in which we compute the optimal tariff taking transition dynamics into account.

The static economy consists of two symmetric countries, Home and Foreign. In each country, a representative household supplies labor inelastically and owns all firms. Firms are heterogeneous in their efficiency and produce differentiated varieties using labor and intermediate inputs. Intermediate inputs can be sourced domestically or imported, but importing requires paying a fixed cost, so only the most efficient firms import. Because importers combine domestic and foreign inputs, they produce at a lower cost than non-importers. A final good is assembled from the differentiated varieties using the Kimball (1995) aggregator, so the demand elasticity a firm faces decreases with its size. Larger firms therefore charge higher markups and pass through cost changes to prices incompletely. The government rebates tariff revenues lump sum to households. The exchange rate is determined by the requirement that trade balances. Throughout the paper, we compare our economy with Kimball demand to an otherwise identical economy with CES demand, in which markups are constant and pass-through is complete.

We calibrate both economies to match the same set of statistics: the aggregate markup, the share of intermediate inputs, the trade elasticity, the import share, the fraction of importers and their employment share. In our Kimball economy we also target the pass-through elasticity of costs to prices. Two differences between these economies are critical for our results. First, the marginal importer is much larger in our economy: it employs 7 times as many workers as the average firm, compared to 3 times in the CES economy. This is because with variable markups employment increases more gradually with firm efficiency, so matching the same fraction and employment share of importers requires that marginal importers be larger. Second, importing confers a larger cost advantage in our economy: it reduces costs by 16%, compared to 8% in the CES economy. This is because the extensive margin of importing responds more to trade costs in our economy, so matching the same trade elasticity requires a lower elasticity of substitution between domestic and foreign inputs.

We characterize the distortions in this economy by comparing the decentralized allocations to those chosen by a world planner who maximizes world consumption. Markups distort the economy along three margins. First, the aggregate markup acts as a tax on inputs, so firms use too few intermediate inputs. Second, productive, high-markup firms produce too little. Third, too few firms import. These distortions are large: eliminating them would raise consumption by 10.1% in our economy, compared to 1.7% in the CES economy. The planner raises the fraction of importers from 5.5% to 8.7% and their employment share from 50% to 96%, more than doubling the import share.

Introducing tariffs greatly amplifies these distortions. To illustrate this, we consider a 10% unilateral Home tariff. In our economy the tariff reduces consumption by 0.75%, whereas in the CES economy it raises it by 0.33%. Importantly, the difference does not arise from the terms of trade: the exchange rate appreciates and imports fall by similar amounts in both economies. It arises instead because aggregate productivity falls by 2.5% in our economy, compared to 1% in the CES economy, a difference entirely accounted for by selection into importing. In our economy marginal importers are large and have a large cost advantage, so their decision to stop importing after a tariff leads to a much larger drop in productivity. Indeed, if we hold the importing cutoff fixed, productivity falls by only 0.7% and consumption increases, as in the CES economy.

To understand these results, we derive an analytical characterization of the response of consumption, net of tariff revenue, to an increase in trade costs. We show that this response is equal to the direct effect that would arise in an efficient economy, amplified by a factor $1 + \Delta$. The amplification term Δ captures the extent to which the higher trade costs move allocations further in the direction in which they are already distorted. If higher trade costs reduce allocations that are already inefficiently low or increase those that are already inefficiently high, Δ is positive. In our economy, the factor $1 + \Delta$ is equal to three. Thus, a given increase in trade costs reduces consumption three times more than it would in the absence of distortions. In contrast, $1 + \Delta$ is only equal to 1.4 in the CES economy. Consequently, the gains from trade are much larger in our economy. Moving from autarky to the current level of trade raises consumption by more than 10%, twice as much as in the CES economy, even though both economies have the same trade elasticity.

We then derive a formula for the optimal tariff in our economy. Absent distortions, the optimal tariff is equal to the inverse of the Foreign export supply elasticity, a standard result. With distortions, the optimal tariff decreases with Δ and is negative when Δ is sufficiently large. In our economy the optimal tariff is -8.1%, so it is optimal to subsidize trade. In contrast, in the otherwise identical CES economy the optimal tariff is positive and equal to 16.2%. Holding the importing cutoff fixed, the optimal tariff is positive in both economies and close to 30%. Selection into importing therefore lowers the optimal tariff in both economies, as in Demidova and Rodríguez-Clare (2009) and Costinot et al. (2020), but on its own it does not overturn the terms-of-trade motive. The interaction between selection and variable markups is therefore critically responsible for our finding that it is optimal to subsidize trade.

We show that our conclusions extend to a richer dynamic economy in which labor supply is elastic, firms accumulate capital, the number of firms is endogenously determined and in which we allow for trade imbalances. We find that the optimal tariff remains negative and equal to -12%. In contrast, it is equal to 10.3% in the otherwise identical CES economy.

Interestingly, we find that the optimal tariff decreases with country size. For example, in our dynamic economy the optimal tariff is -12% when Home accounts for one half of world trade, -3.3% when it accounts for one quarter, and 19.9% when it accounts for 1%. The reason is that when Home is large, its tariff considerably lowers Foreign consumption and therefore Foreign demand for Home exports. Thus, only a small appreciation of the exchange rate is necessary to rebalance trade after an increase in tariffs. This both weakens the terms-of-trade motive and increases the response of trade costs to the tariff, amplifying the effect of distortions. In contrast, when Home is small, Foreign is insulated from the Home tariff, so the exchange rate appreciates much more. This result once again largely reflects the interaction between selection and variable markups, which considerably amplifies the effect of a Home tariff on Foreign consumption. Indeed, in the CES economy with a fixed importing cutoff, the optimal tariff increases with country size.

Related literature. Our paper studies an economy in which variable markups greatly distort firm selection into importing. The optimal tariff we derive extends the classical inverse-elasticity formula[1] by a term that measures the extent to which higher trade costs move allocations further in the direction in which they are already distorted. We quantify this formula and show that the optimal tariff is negative.

A number of papers study the optimal tariff in economies with monopolistic competition and trade in final goods. A classic result is Gros (1987), who derives the optimal tariff in an economy with monopolistic competition and constant markups. As Costinot and Rodríguez-Clare (2014) note, his result coincides with the inverse-elasticity formula. Several papers extend this setting by allowing for exporter selection,[2] which reduces the optimal tariff, or variable markups (Demidova, 2017), which increase it. In all these papers the optimal tariff is positive. One exception is Costinot et al. (2020), who show that, in theory, a uniform import subsidy can be optimal if the non-convexities created by selection abroad are sufficiently strong, but not in the standard CES specification with Pareto productivity and common

[1] The optimal tariff literature is too large for us to do it justice here. A classical reference is Johnson (1953). See Ossa and Redding (2026) for a recent survey. Costinot and Rodríguez-Clare (2014) quantify and Broda et al. (2008) provide evidence for the classical formula. In recent work Alessandria et al. (2025), Auclert et al. (2025), Dávila et al. (2025), Ignatenko et al. (2025), Kocherlakota (2025) and Itskhoki and Mukhin (2026) revisit the optimal tariff in richer economies, which, however, abstract from markup heterogeneity across producers.

[2] See Demidova and Rodríguez-Clare (2009), Felbermayr et al. (2013), Haaland and Venables (2016) and Bagwell and Lee (2020).

fixed exporting costs.

Our paper is more closely related to work that studies the optimal tariff in economies with market power and trade in intermediate inputs (Jung and Köhler, 2021; Caliendo et al., 2023; Antràs et al., 2024; Lashkaripour and Lugovskyy, 2023). These papers assume that markups do not vary with firm size and that there is no selection into importing, the two key ingredients that we focus on and show are quantitatively important. The distortions in these papers can lower the optimal tariff. When quantified, however, the optimal tariff generally remains positive.[3]

Costinot and Werning (2026) offer a Pigouvian perspective on many of these results: their optimal tariff formula depends on the wedge between the social and private cost of a distorted activity times the change in that activity induced by imports. Our decomposition of the welfare effect of trade costs into an efficient component and a term that captures amplification due to distortions is also related to Baqaee and Farhi (2020), who develop formulas for aggregating microeconomic shocks in economies with distortions, as well as to Baqaee and Farhi (2024), who extend this approach to tariffs and trade costs.

Our modeling of import participation builds on models of selection and evidence on heterogeneous importers.[4] Our modeling of markups follows Klenow and Willis (2016) and Edmond et al. (2023). Epifani and Gancia (2011), Edmond et al. (2015) and Arkolakis et al. (2019) study whether trade ameliorates the distortions generated by markups. We study the opposite question: how do these distortions shape trade policy?

2 Motivating Evidence

Our analysis is motivated by three sets of well-known facts. We review them briefly below.

Fact 1: Most international trade is in intermediate goods. As is well known (Hummels et al., 2001; Yi, 2003; Johnson and Noguera, 2012), trade in intermediate goods accounts for a large fraction of international trade. For example, Jones et al. (2025) document that roughly two-thirds of global trade and 50% of U.S. trade are in intermediate inputs. The

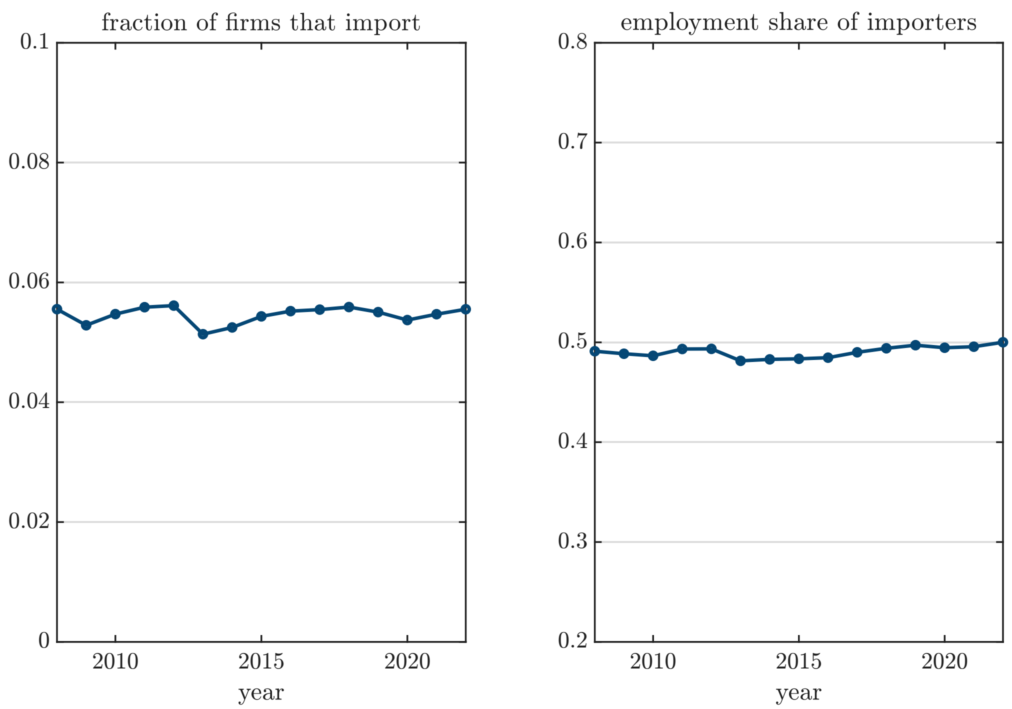
[3]In the 186-country model of Caliendo et al. (2023) the median optimal tariff is 10% and it is negative for only five countries. Antràs et al. (2024) find an optimal tariff of 17% for the U.S.

[4]Halpern et al. (2015) and Antràs et al. (2017) estimate fixed-cost models of sourcing. Gopinath and Neiman (2014) show that a rise in the cost of imported inputs lowers aggregate productivity; Blaum et al. (2018) study the gains from trade in such models. Our model is also consistent with the evidence that tariffs pass through fully into import prices (Amiti et al., 2019b) and incompletely into output prices because firms adjust markups (De Loecker et al., 2016).

remainder of U.S. trade is approximately evenly split between capital and consumption goods.

Fact 2: Importers are few and large. As Bernard et al. (2007) and Bernard et al. (2009) document, only a small fraction of U.S. firms transact directly with foreign suppliers. Despite their small number, these firms account for a large share of employment. We illustrate these facts in Figure 1, which shows that only 5.5% of firms import, but they account for nearly 50% of employment.

Figure 1: Importer Prevalence and Employment Share



Notes: The data are from the U.S. Census Bureau, Business Dynamics Statistics of U.S. Goods Traders.

Fact 3: Markups rise with size and pass-through is incomplete. Bernard et al. (2018) document that importers have 25% higher labor productivity than non-importers. This productivity difference has often been interpreted as evidence of higher markups (Bernard et al., 2003). This interpretation is further supported by the evidence that importers only pass through a fraction of changes in costs to prices. Amiti et al. (2014, 2019a) document, for Belgian firms matched to customs data, a pass-through of only 0.62.

Taken together, these facts motivate our modeling choices below and help guide the quantification. We study an economy in which heterogeneous firms engage in trade in intermediate inputs (Yi, 2003; Gopinath and Neiman, 2014; Blaum et al., 2018). As in Antràs et al. (2017) and Halpern et al. (2015), we rationalize the small fraction of importers by assuming that there is a fixed cost of sourcing from abroad, so only the most productive firms import. We also assume that firms face variable demand elasticities, so the pass-through from costs to prices is incomplete. Moreover, more productive firms charge higher markups and therefore have higher measured labor productivity (Melitz and Ottaviano, 2008; Atkeson and Burstein, 2008; Edmond et al., 2015).

3 Static Model

For clarity, we begin by studying a static environment that allows us to isolate the forces that shape the optimal tariff. We use this setting to derive a formula for the optimal tariff in economies with distortions. We quantify the model to illustrate that the motivating facts described above imply significant distortions that are critical in determining the optimal tariff.

3.1 Environment

The world consists of two countries, Home and Foreign, identical in preferences, technology and initial trade policies. A representative household in each country owns all firms in that country and supplies labor inelastically. Firms are heterogeneous in their efficiency and use labor and intermediate inputs to produce differentiated varieties. Intermediate inputs can be purchased domestically or imported, but importing requires paying a fixed cost. Competitive final good producers aggregate differentiated varieties using a Kimball aggregator. The final good is used for consumption and as an intermediate input. The government in each country collects tariffs on imports and rebates the revenue lump sum to households. Below we mostly focus on describing the problem of the Home country and denote Foreign variables with an asterisk.

3.1.1 Household

The representative household supplies $L = 1$ units of labor inelastically at wage rate W. The budget constraint of the household is

$$C = WL + \Pi + T,$$

where C is aggregate consumption, Π is firm profits, and T denotes lump-sum transfers from the government. We treat the final good in each country as the numeraire and let E denote the exchange rate, expressed as units of the Home final good per unit of the Foreign final good.

3.1.2 Final Goods Producers

Competitive final good producers combine differentiated varieties y_i using the Kimball (1995) aggregator

$$1 = \int_0^1 \Upsilon\Big(\frac{y_i}{Y}\Big)\,\mathrm{d}i. \tag{1}$$

Letting $q_i = y_i/Y$ be the relative quantity of a given variety, cost minimization by final goods producers implies the inverse demand curve

$$p_i = \Upsilon'(q_i)D,$$

where $D = \Big(\int_0^1 \Upsilon'(q_i)\,q_i\,\mathrm{d}i\Big)^{-1}$ is the endogenously determined demand index. We use the Klenow and Willis (2016) functional form for $\Upsilon\,(q)$, which implies that the demand elasticity is $\sigma q^{-\varepsilon/\sigma}$, so that the optimal markup is

$$\mu_i = \frac{\sigma}{\sigma - q_i^{\varepsilon/\sigma}} \tag{2}$$

and increases with firm relative size q_i, more so the larger is the super-elasticity ε/σ. Importantly, this implies that the pass-through of costs to prices is given by

$$\frac{1}{1+\frac{\varepsilon}{\sigma}\mu_i},$$

and is therefore incomplete and declining in size. By setting $\varepsilon = 0$, the model reduces to the standard CES economy with constant markups and complete pass-through.

3.1.3 Producers of Differentiated Varieties

A firm with efficiency z and importer status $j \in \{n, m\}$, where n denotes a non-importer and m an importer, produces output with technology

$$y_j(z) = z\,l_j(z)^{\nu} x_j(z)^{1-\nu},$$

where $l_j(z)$ is labor, $x_j(z)$ is a bundle of intermediate inputs and ν is the elasticity of labor in production. We assume that efficiency $z \geq 1$ is drawn from a Pareto distribution $F(z) = 1 - z^{-\eta}$, with $\eta > 1$.

The firm can access the import market by paying a fixed cost f_m denominated in units of the final output. If the firm does not pay the fixed cost, it only sources intermediate inputs $x_n(z)$ from Home at a price $P_n = 1$. If the firm pays the fixed cost, it can also source intermediate inputs from abroad. Its composite intermediate bundle is then

$$x_m(z) = \left[x_m^h(z)^{\frac{\gamma-1}{\gamma}} + x_m^f(z)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}},$$

where γ is the elasticity of substitution between Home-sourced, $x_m^h(z)$, and Foreign-sourced, $x_m^f(z)$, intermediate inputs.

The price of one unit of Foreign intermediate input expressed in units of the Home final good is

$$\tau = (1+\xi)(1+\kappa)E,$$

where ξ is the Home tariff and κ is an iceberg transportation cost. From now on, we will refer to τ as the *trade cost*. The optimal choice of the domestic and imported intermediate inputs implies that the price of the composite intermediate input faced by an importer is

$$P_m = \left(1 + \tau^{1-\gamma}\right)^{\frac{1}{1-\gamma}}.$$

Since we assume that $\gamma > 1$, the price of the composite intermediate input faced by an importer is lower than the price of the domestic intermediate input, $P_m < P_n = 1$, so importers have a cost advantage relative to non-importers.

Letting

$$\Omega_j = \left(\frac{W}{\nu}\right)^{\nu} \left(\frac{P_j}{1-\nu}\right)^{1-\nu}$$

denote the composite cost of the two inputs in production, the optimal price charged by each producer is a markup over the marginal cost[5]

$$p_j(z) = \mu_j(z)\frac{\Omega_j}{z}.$$

Equivalently, the firm's optimal quantity is implicitly defined by the first-order condition

$$\Upsilon'(q_j(z)) = \mu_j(z)\frac{\Omega_n/D}{z}P_j^{1-\nu}, \tag{3}$$

where the markup $\mu_j(z)$ is given by equation (2).

A firm imports if the profit gain from doing so exceeds the fixed cost. Letting

$$\pi_j(z) = p_j(z)\, y_j(z) \left(1 - \frac{1}{\mu_j(z)}\right)$$

[5]See Appendix A for a derivation of the solution to the firm's problem.

denote operating profits under importer status j, a firm imports if

$$f_m \leq \pi_m(z) - \pi_n(z).$$

Since $\pi_m(z) - \pi_n(z)$ is increasing in z, there exists a cutoff $\bar{z}$ such that all firms with $z \geq \bar{z}$ import and all firms with $z < \bar{z}$ do not. The cutoff is determined by the indifference condition

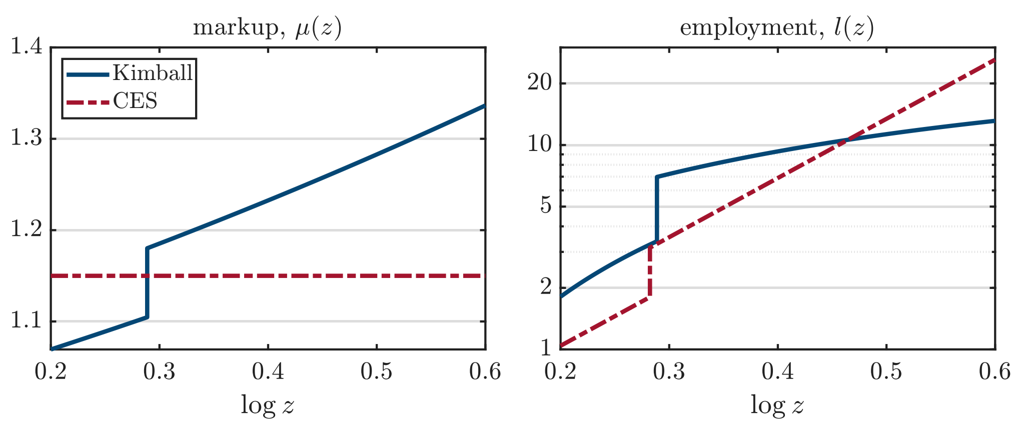
$$f_m = \pi_m(\bar{z}) - \pi_n(\bar{z}) \tag{4}$$

and the fraction of firms that import is

$$\lambda_m = 1 - F(\bar{z}) = \bar{z}^{-\eta}.$$

Figure 2 illustrates the individual decision rules and contrasts the predictions of our model with variable markups with those of a model with CES demand and constant markups. The left panel shows that the markup in our model increases with firm efficiency and jumps discretely when the firm becomes an importer because the importing cost advantage allows the firm to expand output. The right panel shows that employment increases more gradually with efficiency under Kimball demand. More productive firms face lower demand elasticities and pass efficiency gains through to prices only incompletely, so they expand output, and thus employment, by less.

Figure 2: Firm Decision Rules



Notes: The left panel plots markups and the right panel plots employment as functions of firm efficiency under Kimball and CES demand. The plot uses the parameter values in the calibration discussed below. The discrete jumps in both panels reflect the change in importer status.

An implication of this result that will be essential for our policy conclusions and that we will return to later is that the marginal firms under Kimball demand are larger than those under CES demand, provided both models reproduce the same fraction and employment share of importers. Intuitively, the areas under the two employment policy functions to the right of the cutoffs must be the same. Since the policy function in the Kimball economy is flatter, the intercept must be higher to ensure that the areas are equal.

3.2 Equilibrium and Aggregation

We conclude the description of the model by discussing the key equilibrium conditions and aggregating individual firm decision rules. Wages in each country, W and W^*, are pinned down by the respective labor market clearing conditions. For example, in Home, the labor market clearing condition is

$$L = \int_1^{\bar{z}} l_n(z)\,\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} l_m(z)\,\mathrm{d}F(z). \tag{5}$$

The exchange rate E is pinned down by the requirement that trade balances

$$(1+\kappa)\underbrace{\int_{\bar{z}^*}^{\infty} x_m^{*h}(z)\,\mathrm{d}F(z)}_{\equiv \mathrm{EX}} \;=\; (1+\kappa)\,E\underbrace{\int_{\bar{z}}^{\infty} x_m^{f}(z)\,\mathrm{d}F(z)}_{\equiv \mathrm{IM}}, \tag{6}$$

where the left-hand side is the value of Home exports and the right-hand side is the value of Home imports.

The aggregate resource constraint is

$$Y = C \;+\; X \;+\; \lambda_m f_m, \tag{7}$$

where X is the amount of the Home final good that is used as an intermediate input by both Home and Foreign firms and is equal to

$$X = \int_1^{\bar{z}} x_n(z)\,\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} x_m^h(z)\,\mathrm{d}F(z) + (1+\kappa)\int_{\bar{z}^*}^{\infty} x_m^{*h}(z)\,\mathrm{d}F(z).$$

To understand the workings of the model, we find it useful to aggregate firm-level decision rules into aggregate counterparts. We define the aggregate markup $\mathcal{M}$ as the harmonic sales-weighted average of firm-level markups[6]

$$\frac{1}{\mathcal{M}} = D\left(\int_1^{\bar{z}} \frac{\Upsilon'(q_n(z))\,q_n(z)}{\mu_n(z)}\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} \frac{\Upsilon'(q_m(z))\,q_m(z)}{\mu_m(z)}\mathrm{d}F(z)\right).$$

The aggregate markup acts as a uniform tax on labor and intermediate inputs. Specifically, it reduces the labor share

$$\frac{WL}{Y} = \frac{\nu}{\mathcal{M}}$$

[6]See Appendix A for a derivation of all the aggregation results in this section.

as well as the intermediate input share. The intermediate input share is further reduced by tariffs. Imposing that trade must balance, we can express the ratio of intermediate inputs to output as

$$\frac{X}{Y} = \frac{1-\nu}{\bar{\mathcal{M}}},$$

where the wedge

$$\frac{1}{\bar{\mathcal{M}}} = \frac{1}{\mathcal{M}} - \frac{T}{(1-\nu)\,Y}$$

is reduced both by the aggregate markup and by the tariff. Here,

$$T = \xi\,(1+\kappa)\,E \int_{\bar{z}}^{\infty} x_m^f(z)\,\mathrm{d}F(z) \tag{8}$$

is the tariff revenue collected by the Home government.

Aggregating optimal labor and intermediate input choices of individual firms allows us to write an aggregate production function

$$Y = ZL,$$

where aggregate productivity is given by

$$Z = \left(\frac{1-\nu}{\mathcal{M}}\right)^{\frac{1-\nu}{\nu}} \Bigg(\underbrace{\int_1^{\bar{z}} \frac{q_n(z)}{z}\mathrm{d}F(z) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\mathrm{d}F(z)}_{\mathcal{Q}}\Bigg)^{-\frac{1}{\nu}}. \tag{9}$$

The first term captures the fact that markups act as a tax on intermediate inputs, while the second term $\mathcal{Q}$ summarizes the allocation of production across firms with different efficiency and importer status. We will use this representation of aggregate productivity to decompose the distortions in this economy.

3.3 Parameterization

We start from a steady state in which there are no tariffs and the two countries are symmetric. We have three sets of parameters to calibrate, describing market power, technology and trade costs. The parameters describing market power are the demand elasticity at $q=1$, σ, and the curvature of the demand elasticity, ε. The technology parameters are the elasticity of labor in production, ν, the tail of the productivity distribution, η, and the elasticity of substitution between Home and Foreign intermediate inputs, γ. The trade cost parameters are the iceberg cost, κ, and the fixed cost of importing, f_m.

We jointly calibrate these parameters to match seven moments: an aggregate markup of 1.15 (Edmond et al., 2023), a sales-weighted pass-through elasticity of 0.62 (Amiti et al., 2014), a ratio of intermediate inputs to gross output of 0.42, a partial-equilibrium trade elasticity of 4 (Simonovska and Waugh, 2014), an import share of gross output of 8.1%, an importer fraction of 5.55% of firms and an importer employment share of 50%. The intermediate inputs to gross output ratio and the import share are calculated using data from the BEA for 2023, while the fraction of importers and their employment share are as described in Section 2.

We report the calibrated parameter values in Panel A of Table 1 and the targeted moments in Panel B. To illustrate the importance of variable markups, throughout the paper we will compare the Kimball model with an otherwise identical economy with CES demand. We also report the parameter values for the CES model, which is calibrated to match the same set of moments, except for the pass-through elasticity, which is mechanically equal to one.

Table 1: Parameterization

A. Calibrated Parameter Values

		Kimball	CES
σ	demand elasticity at $q = 1$	20.02	7.667
ε/σ	super-elasticity of demand	0.530	0
ν	labor elasticity	0.517	0.517
η	Pareto tail of productivity	10.01	10.24
γ	home–foreign input elasticity	2.346	3.829
$\lambda_m f_m/Y$	importing costs (% of output)	4.360	0.972
κ	iceberg cost	0.409	0.177

B. Targeted Moments

	Data	Kimball	CES
aggregate markup	1.15	1.15	1.15
pass-through elasticity	0.62	0.62	1
intermediate inputs/output	0.42	0.42	0.42
trade elasticity	4.00	4.00	4.00
import share	0.08	0.08	0.08
percent of firms importing	5.55	5.55	5.55
employment share of importers	0.50	0.50	0.50

Notes: Pass-through is mechanically equal to one under CES.

Because the model is exactly identified, the fit is perfect. Although the parameters are jointly identified by all targets, we find it useful to provide some intuition for identification. The aggregate markup and the pass-through elasticity primarily inform σ and ε, while the intermediate input share informs ν. The trade elasticity is informative about γ, the employment share of importers helps pin down η, the fraction of importers helps pin down f_m and the import share helps pin down κ.

We note that the Kimball economy requires a lower γ (2.3 vs. 3.8) to match the same trade elasticity of 4. To understand why, note that the trade elasticity in the model has two components: an intensive margin that captures substitution towards Home inputs among firms that maintain their importer status and an extensive margin that captures changes in the fraction of importers.[7] In the Kimball economy, the extensive margin is stronger for two reasons. First, higher trade costs reduce markups and therefore profit margins, an effect absent in the CES economy. This implies that the import cutoff is more sensitive to a change in trade costs. Second, as discussed above, the marginal importer is larger, so a given change in the importer cutoff has a larger effect on trade. Because both models match the same overall trade elasticity, the intensive margin, which is primarily governed by γ, must be lower in the Kimball economy to offset the stronger extensive margin.

That γ is lower under Kimball demand also implies that importers in this economy have a larger cost advantage. To see why, notice that we can express the price of intermediate inputs faced by importers as

$$P_m = (1-\vartheta)^{\frac{1}{\gamma-1}},$$

where

$$\vartheta \equiv \frac{\tau x_m^f(z)}{P_m x_m(z)} = \frac{\tau^{1-\gamma}}{1+\tau^{1-\gamma}}$$

is the share of imported intermediate inputs in an importer's total spending on intermediates. This share is immediately pinned down by three of our data targets: the import share, the employment share of importers, which, absent tariffs, is equivalent to their intermediate input share, and the intermediate input share. Since ϑ is the same in both economies, a lower γ implies a lower P_m and therefore a larger cost advantage for importers in the Kimball economy. Because of this, the Kimball economy requires higher trade costs to match the same import share and fraction of importers.

We also note that the value of the tail parameter η is similar in the two economies. This parameter is pinned down by the employment share of importers. As discussed above, even

[7]See Appendix B for a detailed derivation of the trade elasticity. There we analytically decompose the trade elasticity into these two margins and discuss and quantify the forces that shape each of them.

though employment increases less with productivity in the Kimball economy, the marginal importer is larger. Therefore, the model does not need much more dispersed productivity to match the employment share of importers.

Our model is consistent with several additional features of the data that we have not directly targeted in our calibration. First, the model reproduces well the evidence in Bernard et al. (2018) that importers have higher labor productivity. In particular, we find that the value added per worker of importers is 31% larger than that of non-importers, which is in line with the 25% premium documented by Bernard et al. (2018). Second, the elasticity of substitution between Home and Foreign intermediate inputs, γ, which we estimate to be 2.3 in our Kimball economy, is consistent with the estimates in Antràs et al. (2017) and Blaum et al. (2018), which put it in the neighborhood of 2.5.

3.4 Distortions

We next characterize and quantify the distortions in this economy. To do so, we solve the problem of a world planner who maximizes the equally weighted sum of Home and Foreign consumption, subject to the same technology as in the decentralized equilibrium. The planner chooses the importer cutoff, the allocation of labor, intermediate inputs and relative quantities across firms in each country, subject to the final good production function (1), the labor resource constraint (5) and the aggregate resource constraint (7).[8]

Comparing the optimal choices of the planner with the allocations in the decentralized equilibrium allows us to highlight three distortions, all of which lower aggregate productivity and consumption. First, firms in the decentralized economy use too few intermediate inputs. Namely, the intermediate input share chosen by the planner is $1-\nu$, higher than the $\frac{1-\nu}{\mathcal{M}}$ in the decentralized equilibrium. Second, productive, high-markup firms in the decentralized equilibrium produce too little. To see this, we note that the efficient allocations chosen by the planner satisfy the first-order condition

$$\Upsilon'(q_j(z)) = \frac{\zeta}{z} P_j^{1-\nu},$$

where ζ is the geometric weighted average of the multipliers on the labor and aggregate resource constraints, $P_n = 1$ and $P_m = \left(1 + (1+\kappa)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$. Comparing this condition with the decentralized first-order condition (3) shows that the decentralized allocations feature wedges $\mu_j(z)$ that increase with productivity.

[8] See Appendix C for the analysis of the planner's problem.

Third, the fraction of importers is distorted in the decentralized equilibrium. The efficient cutoff satisfies

$$\frac{f_m}{DY} = (\epsilon_m(\bar{z}) - 1)\,\Upsilon'(q_m(\bar{z}))\,q_m(\bar{z}) - (\epsilon_n(\bar{z}) - 1)\,\Upsilon'(q_n(\bar{z}))\,q_n(\bar{z}),$$

where

$$\epsilon_j(z) = \frac{\Upsilon(q_j(z))}{\Upsilon'(q_j(z))\,q_j(z)}$$

is the inverse elasticity of the Kimball aggregator Υ with respect to the relative quantity of a given variety. In contrast, the decentralized cutoff in equation (4) can be written as

$$\frac{f_m}{DY} = \frac{1}{\mu_m(\bar{z})}(\mu_m(\bar{z}) - 1)\,\Upsilon'(q_m(\bar{z}))\,q_m(\bar{z}) - \frac{1}{\mu_n(\bar{z})}(\mu_n(\bar{z}) - 1)\,\Upsilon'(q_n(\bar{z}))\,q_n(\bar{z}).$$

This differs from the planner's condition along two dimensions. First, the markup $\mu_j(\bar{z})$ that shapes the decision to import in the decentralized equilibrium is not equal to the inverse elasticity $\epsilon_j(\bar{z})$ that shapes the planner's choice of which firms import. Second, the decentralized condition features an additional wedge $1/\mu_j(\bar{z})$ that reduces the gains from importing. With CES demand, $\mu_j(\bar{z}) = \epsilon_j(\bar{z}) = \frac{\sigma}{\sigma-1}$, so the markup acts as a tax on importing and too few firms import. More generally, with Kimball demand, the fraction of importers may be too high or too low.

We quantify the importance of these distortions in Table 2 by reporting the consumption gains from implementing the planner's allocation as well as those from fixing one distortion at a time. The overall consumption gains from removing distortions are much larger in the Kimball economy (10.1% vs. 1.7%). While the intermediate input choice is distorted by a similar magnitude in both economies, the allocation of output across firms and the importing decision are much more severely distorted under Kimball demand. Also note that while the planner increases the fraction of importers by similar amounts in both economies (from 5.5% to 8.7% in Kimball and to 8.1% in CES), the employment share of importers increases much more in the Kimball economy (from 50% to 96%) than in CES (from 50% to 56%), reflecting the larger cost advantage of importers in the former. Consequently, relative to the baseline value of 0.08, the efficient import share rises to 0.18 in the Kimball economy and only to 0.10 in the CES economy.

4 Optimal Tariff Policy

We next turn to optimal tariff policy. We first illustrate the impact of tariffs on welfare using a quantitative example. We then provide an analytical characterization of the impact

Table 2: Effect of Removing Distortions

	Baseline	Kimball	CES
Consumption gains (%)			
overall		10.05	1.71
intermediate inputs only		1.70	1.60
misallocation only		6.63	0
fraction of importers only		2.00	0.10
Efficient allocations			
percent of firms importing	5.55	8.74	8.14
employment share of importers	0.50	0.96	0.56
import share	0.08	0.18	0.10

Notes: The top panel reports the consumption gains, in percent, from removing all distortions and each distortion in isolation. The bottom panel reports the fraction of firms that import, their employment share, and the import share in the decentralized baseline and in the efficient allocations under Kimball and CES demand.

of tariffs on welfare. The response of welfare to tariffs is critically shaped by the extent to which distortions amplify the effect of changes in trade costs on welfare. Lastly, we derive a formula for the optimal tariff and show that it depends not only on the Foreign export supply elasticity, but also on the amplification due to distortions. We use the calibrated model to quantify the optimal tariff and show that it is negative.

4.1 Impact of Tariffs: Quantitative Example

We next illustrate the impact of tariffs quantitatively by considering a 10% unilateral Home tariff, the revenue from which is rebated as a lump-sum transfer to the representative Home household. Table 3 reports the effects of the tariff in both the Kimball and CES models and shows that our baseline Kimball economy predicts a 0.75% decline in consumption, in sharp contrast to the CES economy, which predicts a 0.33% increase in consumption. This discrepancy arises despite a similar exchange rate appreciation and decline in imports. It is driven instead by a much larger drop in productivity (2.5% vs. 1%). This larger decline occurs even though production reallocates toward low-markup non-importers, reducing the aggregate markup and partially offsetting the productivity loss. Two forces are responsible for the decline in productivity. First, the employment share of importers falls more in the Kimball economy. Second, as already discussed, importers have a much larger cost advantage.

Table 4 reinforces the point that the decline in productivity and consumption is due to

Table 3: Effects of a 10% Tariff

	Baseline	Kimball	CES
consumption change, %		−0.75	0.33
productivity change, %		−2.49	−1.00
exchange rate change, %		−5.07	−5.19
import share	0.081	0.062	0.064
aggregate markup	1.150	1.143	1.150
percent of firms importing	5.55	4.66	4.78
employment share of importers	0.50	0.43	0.47

Notes: The table reports the effects of a 10% Home tariff. Changes are relative to the initial steady state and all remaining entries are levels. The baseline column reports initial steady-state values. The Kimball and CES columns report statistics from the equilibrium with tariffs in the respective models.

the decline in the fraction of importers. In the columns labeled "fixed $\bar{z}$", we consider a counterfactual scenario in which the importing cutoff $\bar{z}$ is held fixed at its initial steady-state value.[9] For comparison, the columns labeled "moving $\bar{z}$" reproduce the results of the baseline model in which the importing cutoff adjusts. Notice that the decline in productivity in the fixed cutoff scenario is much smaller, 0.7%, and is similar to that in the CES economy. As a consequence, consumption increases in the Kimball economy as well.

Table 4: Role of the Moving Cutoff

	Kimball		CES	
	moving $\bar{z}$	fixed $\bar{z}$	moving $\bar{z}$	fixed $\bar{z}$
consumption change, %	−0.75	0.38	0.33	0.47
productivity change, %	−2.49	−0.71	−1.00	−0.69
exchange rate change, %	−5.07	−5.84	−5.19	−5.39

Notes: The moving cutoff columns allow $\bar{z}$ to adjust in response to a 10% Home tariff. The fixed cutoff columns hold $\bar{z}$ at its initial steady-state value.

4.2 Impact of Tariffs: Analytical Characterization

We next analytically characterize the response of Home consumption to a change in tariffs, a response that will shape the optimal tariff formula that we derive in Section 4.3 below. We do so in stages. First, we provide a general characterization of how consumption responds

[9] One way to interpret this scenario is that the tariff is accompanied by the introduction of a subsidy, financed by lump-sum taxes on Home households, that partially offsets the fixed cost of importing.

to changes in trade costs, $\tau = (1+\xi)(1+\kappa)E$, in economies with distortions. Second, we specialize this characterization to our model. Third, we characterize how the exchange rate and therefore the trade cost respond to changes in tariffs. Fourth, we combine all these to describe the overall effect of tariffs on Home consumption.[10]

4.2.1 Response of Consumption to Changes in Trade Costs

We first describe how consumption responds to changes in trade costs. We find it useful to isolate the component of consumption that does not include tariff revenue, $C - T$. We can write this generally as a function of allocations a and trade costs τ

$$C - T = \bar{C}(a, \tau),$$

where a is a vector that includes all the allocations in the decentralized economy.[11] The total effect of a change in trade costs on consumption net of tariff revenue can then be written as

$$\frac{\mathrm{d}\bar{C}(a,\tau)}{\mathrm{d}\tau} = \frac{\partial \bar{C}(a,\tau)}{\partial \tau} + \nabla_a \bar{C}(a,\tau)' \frac{\mathrm{d}a}{\mathrm{d}\tau},$$

where $\nabla_a \bar{C}(a,\tau)$ is the gradient of $\bar{C}$ with respect to the allocation vector a and describes how Home consumption changes as we change each component of the decentralized allocations. Holding allocations fixed, an increase in the trade cost τ affects consumption only through its direct effect on import costs, so

$$\frac{\partial \bar{C}(a,\tau)}{\partial \tau} = -\mathrm{IM}.$$

Therefore, we can write the total effect of a change in trade costs on consumption as the sum of its direct effect and the effect through changes in allocations

$$\frac{\mathrm{d}(C-T)}{\mathrm{d}\tau} \equiv -(1+\Delta)\,\mathrm{IM}, \tag{10}$$

where

$$\Delta = -\frac{1}{\mathrm{IM}} \nabla_a \bar{C}(a,\tau)' \frac{\mathrm{d}a}{\mathrm{d}\tau}.$$

At the efficient allocations $\nabla_a \bar{C}(a,\tau)$ is zero and therefore so is Δ. If $\nabla_a \bar{C}(a,\tau)$ is positive in a given direction, increasing that particular allocation would increase consumption. An increase in trade costs that instead reduces that allocation moves it further away in the

[10]See Appendix D for the derivation of the results in this section.

[11]We use the vector language for convenience, but formally a is in a function space and the gradients we introduce should be understood as functional derivatives.

direction in which it is already distorted, amplifying the negative effect on consumption. Thus Δ captures the extent to which higher trade costs move the economy further away from the initial efficient allocations, amplifying the response of consumption to trade costs. This interpretation is closely related to the Pigouvian logic underlying the formulas in Costinot and Werning (2026), which weight changes in distorted activities induced by trade by the wedges between their social and private marginal costs. Our decomposition also parallels Baqaee and Farhi (2020), who show that the response of aggregate output to microeconomic shocks separates into a term that obtains in an efficient economy and a term that arises only in the presence of distortions.

We can further relate the amplification due to distortions, Δ, to the underlying distortions themselves.[12] Let a^p denote the efficient allocation in the absence of tariffs, $\hat{a} \equiv a - a^p$ the gap between the decentralized and the efficient allocations and $\mathcal{L} \equiv \bar{C}(a^p, \tau) - \bar{C}(a, \tau)$ the level of inefficiencies. At the initial steady state without tariffs, the level of inefficiencies $\mathcal{L}$ is simply the consumption loss due to distortions that we report in Table 2. A second-order approximation around a^p gives

$$\mathcal{L} \simeq \frac{1}{2}\hat{a}' H \hat{a},$$

where $H = -\nabla^2_{aa}\bar{C}(a^p, \tau)$ is the Hessian of consumption evaluated at the efficient allocation. Combining this approximation with the expression for Δ above yields

$$\Delta \simeq \frac{2\mathcal{L}}{\mathrm{IM}}\hat{\beta},$$

where

$$\hat{\beta} \equiv \frac{\hat{a}' H \frac{\mathrm{d}a}{\mathrm{d}\tau}}{\hat{a}' H \hat{a}}$$

is the weighted projection of the allocation response $\frac{\mathrm{d}a}{\mathrm{d}\tau}$ on the preexisting allocation gap $\hat{a}$. It is positive when higher trade costs lower precisely those allocations that are already below their efficient levels or when they raise those that are already too high. Thus, $\mathcal{L}$ measures how far the economy is from efficiency, while $\hat{\beta}$ measures the extent to which higher trade costs move allocations further away from the initial efficient allocation a^p.

To summarize, the consumption response to changes in trade costs is amplified when two conditions are simultaneously satisfied: the economy is distorted to begin with, so $\mathcal{L}$ is high, and changes in trade costs move the economy further away from the initial efficient allocation, so $\hat{\beta}$ is also high. Consequently, a large $\mathcal{L}$ need not imply a large Δ if changes in

[12] See Appendix D.1 for a derivation.

trade costs do not move the economy further away from the initial efficient allocation. We next describe the determinants of Δ in our model.

4.2.2 Determinants of Δ in Our Model

To characterize the response of consumption to a change in trade costs in our model, first note that consumption net of tariff revenue is given by

$$C - T = WL + \Pi = \left(1 - \frac{1-\nu}{\mathcal{M}}\right) Z - \lambda_m f_m. \tag{11}$$

We proceed in two steps. First, we describe how aggregate productivity Z responds to changes in trade costs. Second, we use this to characterize the overall response of consumption.

Response of Aggregate Productivity to Changes in Trade Costs. Totally differentiating the expression for aggregate productivity Z in equation (9) with respect to τ allows us to write

$$\begin{aligned}\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau} = & -\underbrace{\frac{1-\nu}{\nu}(1-\lambda)}_{\text{ACR}} \quad - \quad \underbrace{\frac{1-\nu}{\nu}\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}}_{\text{markup response}} \quad - \quad \underbrace{\frac{1}{\nu\,\mathcal{Q}}}_{\text{quantity reallocation}}\mathcal{R} \\ & -\underbrace{\frac{1}{\nu}\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\frac{l_m(\bar{z})}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)}_{\text{loss of importing technology}},\end{aligned} \tag{12}$$

where λ is the domestic intermediate spending share and where

$$\mathcal{R} = \int_1^{\bar{z}} \frac{q_n(z)}{z}\frac{\mathrm{d}\log q_n(z)}{\mathrm{d}\log\tau}\,\mathrm{d}F(z) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\frac{\mathrm{d}\log q_m(z)}{\mathrm{d}\log\tau}\,\mathrm{d}F(z) + F'(\bar{z})\Big[q_n(\bar{z}) - q_m(\bar{z})\Big]\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}.$$

Equation (12) allows us to isolate the channels through which a change in trade costs affects aggregate productivity. The first term captures the direct effect of higher trade costs on productivity and depends on the share of intermediates that are imported, $1-\lambda$, and the intermediate input share, $1-\nu$. Notice that this first term can be equivalently expressed as $-\frac{1-\nu}{\nu}\frac{1}{\mathcal{A}}\frac{\mathrm{d}\log\lambda}{\mathrm{d}\log\tau}$, where $\mathcal{A}$ is the general-equilibrium trade elasticity, so this term has the same form as the sufficient statistic in Arkolakis et al. (2012) and therefore we refer to it as ACR. The second term captures the direct effect of a change in markups on productivity. The third term captures the effect of reallocation of production among firms that do not change importer status (the first two terms in $\mathcal{R}$) and from firms that stop importing (the last term in $\mathcal{R}$). Finally, the fourth term captures the effect of a change in the fraction of importers on productivity and the resulting loss in the importers' cost advantage.

We illustrate the relative importance of these channels in Table 5 by evaluating the total derivative in equation (12) and each of its components. We scale each derivative so the results can be interpreted as the response to a 10% increase in τ. As the table shows, a 10% increase in trade costs leads to a much larger drop in productivity in the Kimball economy than in the CES economy, −5.8% vs. −2.5%. The difference is primarily driven by the loss of importing technology, which reduces productivity by 5.3% in the Kimball economy and only by 1% in CES. Though the fall in the aggregate markup increases productivity in the Kimball economy by 1.3%, its effect is not strong enough to overturn the loss of importing technology.

To further illustrate the importance of importer exit, we repeat this decomposition in a counterfactual where we keep the cutoff $\bar{z}$ fixed. In this case, the drop in productivity is similar across the two economies. Moreover, the ACR term now accounts for the bulk of the productivity losses even in the Kimball economy because the positive effect of a lower aggregate markup is offset by an increase in misallocation.

Table 5: Productivity Response to an Increase in Trade Costs

	Kimball		CES	
	moving $\bar{z}$	fixed $\bar{z}$	moving $\bar{z}$	fixed $\bar{z}$
overall	−5.81	−1.96	−2.45	−1.72
ACR	−1.72	−1.72	−1.72	−1.72
markup response	1.32	0.54	0.00	0.00
quantity reallocation	−0.12	−0.78	0.26	0.00
loss of importing technology	−5.28	0.00	−0.99	0.00

Notes: The table reports $\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau}$ and its components, evaluated at the initial steady state. The numbers are scaled by $100\log(1.10)$ and can be interpreted as the impact of a 10% increase in τ.

To understand why the loss of importing technology is so consequential in the Kimball economy, notice that it is the product of three terms

$$-\frac{1}{\nu}\underbrace{\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)}_{\substack{\text{change fraction}\\ \text{importers}}} \quad \underbrace{\frac{l_m(\bar{z})}{L}}_{\substack{\text{employment}\\ \text{marginal importer}}} \quad \underbrace{\left(\frac{1}{P_m^{1-\nu}}-1\right)}_{\substack{\text{cost advantage}\\ \text{importers}}}.$$

Intuitively, a given drop in the fraction of importers is more consequential if marginal importers are large and if importing confers a larger cost advantage. As Table 6 shows, the fraction of importers falls by a similar amount in both economies. However, in the Kimball economy the marginal importer is much larger: 6.98 vs. 3.15 times the size of the average firm. Moreover, the cost advantage of importers is also larger: the decision to import reduces costs by 16% in Kimball and only 8% in CES. Altogether, these two factors imply that losing a fraction of importers is much more costly in our Kimball economy than under CES.

Table 6: Loss of Importing Technology

	Kimball	CES
loss of importing technology	−5.28	−0.99
decline in fraction of importers	2.04	1.86
employment of marginal importer	6.98	3.15
cost advantage of importers	0.16	0.08

Notes: The table evaluates each of the components of the last term in the expression for $\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau}$.

Components of Δ. Totally differentiating equation (11) and using equations (10) and (12), we can separate the various components of Δ in our framework as follows

$$\Delta = \underbrace{\frac{\mathcal{M}-1}{\nu}}_{\text{markup level}} + \underbrace{\frac{Z}{\tau\,\mathrm{IM}}\frac{1-\nu}{\nu}\left(1-\frac{1}{\mathcal{M}}\right)\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}}_{\text{markup response}} + \underbrace{\frac{Z}{\tau\,\mathrm{IM}}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{1}{\nu\mathcal{Q}}\mathcal{R}}_{\text{quantity reallocation}}$$

$$+\underbrace{\frac{1}{\tau\,\mathrm{IM}}\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\left[\frac{Z}{\nu}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{l_m(\bar{z})}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)-f_m\right]}_{\text{importer exit}}.$$

The first component captures the level of markups. If markups are greater than one, imports are inefficiently low, so further reducing them through higher trade costs amplifies the consumption decline. The second component captures the response of markups. If the initial markup is greater than one, further increasing it exacerbates the consumption drop. The third component captures the effect of reallocating production across firms. Finally, the last component is the additional loss from importer exit. If there are too few importers to begin with, an increase in trade costs, which induces a decrease in the mass of importers, further reduces consumption.

Table 7 quantifies each of these components. Domestic distortions amplify the consumption responses to changes in trade costs much more in our baseline Kimball economy compared to the CES economy. In response to a given increase in trade costs, consumption falls by $1 + \Delta = 3$ times as much as in the absence of distortions in the Kimball economy. In contrast, in the CES economy, in response to the same increase in trade costs, consumption falls by only 1.4 times as much as in the absence of distortions. The bulk of this difference is accounted for by the exit of importers, whose mass falls even more relative to the efficient allocation. Eliminating this margin by fixing the cutoff $\bar{z}$ reduces the amplification due to domestic distortions in both economies.

Table 7: Amplification Due to Distortions

	Kimball		CES	
	moving $\bar{z}$	fixed $\bar{z}$	moving $\bar{z}$	fixed $\bar{z}$
overall amplification, Δ	2.05	0.79	0.41	0.29
markup level	0.29	0.29	0.29	0.29
markup response	−0.22	−0.09	0.00	0.00
quantity reallocation	0.09	0.59	−0.20	0.00
net loss from importer exit	1.89	0.00	0.32	0.00

Notes: The table reports Δ and its four components, evaluated at the initial steady state.

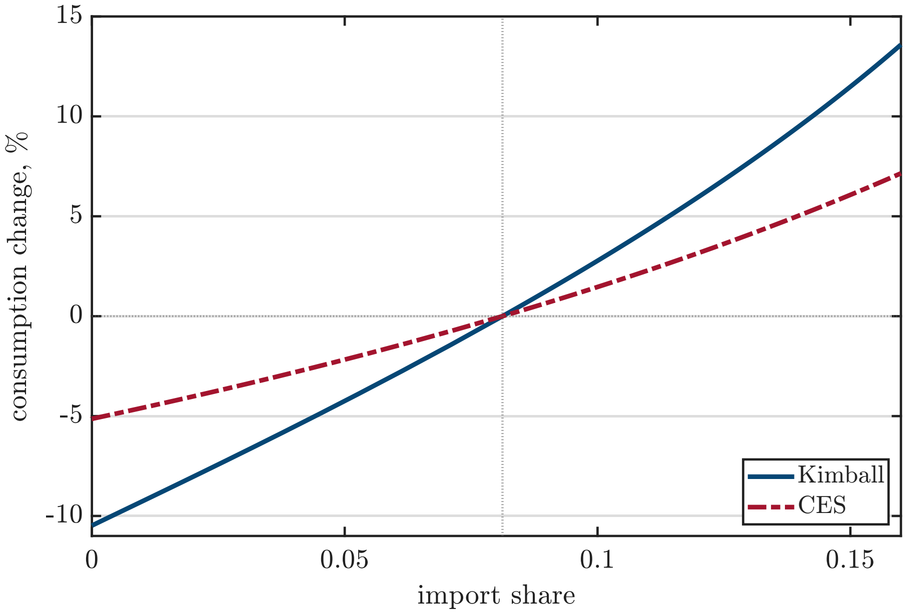
That consumption responds much more to changes in trade costs in the Kimball economy implies that the gains from trade are also larger. We illustrate this in Figure 3, where we trace out how consumption varies in the two economies as we change the iceberg cost κ and thus the import share. In the Kimball economy, moving from autarky to the current level of trade increases consumption by more than 10%, whereas in the CES economy the increase is only 5%. Similarly, doubling the import share from its current level increases consumption by 14% in the Kimball economy, nearly twice as much as in the CES economy. These differences are large, despite the two economies having the same trade elasticity.

4.2.3 Response of Trade Costs to Changes in Tariffs

So far, we have characterized how consumption responds to changes in trade costs. To complete the analysis, we next show how trade costs themselves respond to changes in tariffs. Since $\tau = (1+\xi)(1+\kappa)E$, the response of trade costs to a change in the Home tariff ξ depends on how the exchange rate responds to the change in the tariff,

$$\frac{\mathrm{d}\log\tau}{\mathrm{d}\log(1+\xi)} = 1 + \frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)}. \tag{13}$$

Figure 3: Gains from Trade



Notes: The figure reports the consumption change relative to the steady state as a function of the import share as we vary κ. The dotted vertical line marks the calibrated steady state.

The exchange rate, in turn, is pinned down by the balanced trade condition (6). Letting

$$\mathcal{D} = -\frac{\mathrm{d}\log \mathrm{IM}}{\mathrm{d}\log \tau}$$

denote the Home import demand elasticity and

$$\mathcal{E} = -\frac{\mathrm{d}\log \mathrm{EX}/E}{\mathrm{d}\log \tau^*}$$

denote the Foreign export supply elasticity, the trade balance condition implies that

$$\frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)} = -\frac{\mathcal{D}}{\mathcal{D}+\mathcal{E}}. \tag{14}$$

To gain some intuition, suppose that ξ increases by 1%. At the old exchange rate, the trade cost τ also increases by 1%, which reduces Home imports by $\mathcal{D}$%, creating a trade surplus. To restore the trade balance, exports have to fall, which requires a drop (appreciation) in the exchange rate. The larger is the Home import demand elasticity $\mathcal{D}$, the larger is the drop in Home imports, so the larger is the required exchange rate appreciation. The larger is the Foreign export supply elasticity $\mathcal{E}$, the larger is the drop in exports, so the smaller the required appreciation.

Finally, equations (13) and (14) give the response of trade costs to a change in tariffs

$$\frac{\mathrm{d}\log \tau}{\mathrm{d}\log(1+\xi)} = \frac{\mathcal{E}}{\mathcal{D}+\mathcal{E}}. \tag{15}$$

This expression shows that the share of the increase in tariffs that is passed through to trade costs depends on the relative elasticities of Home import demand and Foreign export supply. A higher Home import demand elasticity reduces this pass-through, while a higher Foreign export supply elasticity increases it.

4.2.4 Response of Consumption to Changes in Tariffs

Finally, combining equations (8), (10) and (15) allows us to write the response of consumption to a change in tariffs as

$$\frac{\mathrm{d}\log C}{\mathrm{d}\log(1+\xi)} = \frac{(1+\kappa)E\,\mathrm{IM}}{C}\Bigg(\underbrace{-\frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)}}_{\text{terms of trade}} + \underbrace{\xi\frac{\mathrm{d}\log \mathrm{IM}}{\mathrm{d}\log(1+\xi)}}_{\text{tariff wedge}} - \underbrace{\Delta(1+\xi)\frac{\mathrm{d}\log\tau}{\mathrm{d}\log(1+\xi)}}_{\text{distortions}}\Bigg). \tag{16}$$

An increase in tariffs has three distinct effects on consumption. First, it improves the terms of trade, which increases consumption. Second, if the tariff is positive to begin with, so that imports are too low because of the tariff, further increasing it worsens this inefficiency. These two forces are the standard tradeoffs that govern tariff policy in an undistorted economy ($\Delta = 0$). Notice that $\xi = 0$ is suboptimal in this case because $\frac{\mathrm{d}\log C}{\mathrm{d}\log(1+\xi)} > 0$. Thus an increase in tariffs raises consumption and the optimal tariff is positive in the absence of distortions.

As equation (10) shows, domestic distortions add a channel through which tariffs reduce consumption. Namely, if $\Delta > 0$, the increase in trade costs resulting from an increase in tariffs amplifies the drop in consumption. This channel is captured by the last term in equation (16). In the presence of this channel, $\frac{\mathrm{d}\log C}{\mathrm{d}\log(1+\xi)}$ is not necessarily positive at $\xi = 0$. Thus the effect tariffs have on consumption depends on the extent to which distortions amplify the response of consumption to changes in trade costs.

4.3 Optimal Tariff

We now turn to the determination of the optimal tariff. Setting the derivative of consumption with respect to the tariff in equation (16) equal to zero and rearranging gives the following expression for the optimal tariff

$$\xi = \frac{1}{\mathcal{E}}\frac{\mathcal{D} - \mathcal{E}\Delta}{\mathcal{D} + \Delta}. \tag{17}$$

Absent distortions ($\Delta = 0$), the optimal tariff is positive and it is equal to $1/\mathcal{E}$, the standard result from the terms-of-trade manipulation literature. More generally, the optimal tariff is decreasing in Δ. We note that equation (17) is an implicit expression for the optimal tariff since $\mathcal{D}$, $\mathcal{E}$ and Δ also depend on the tariff ξ, so the optimal tariff is the fixed point of this equation.

Table 8 reports the resulting optimal tariff. The optimal tariff is negative in the baseline Kimball economy and equal to -8.1%, in sharp contrast to the CES economy, where it is positive and equal to 16.2%. Once again, the difference between Kimball and CES is due to the extensive margin of importing. In the counterfactual Kimball economy with a fixed cutoff $\bar{z}$, the optimal tariff is positive and equal to 29.6%, similar to that in the CES economy with a fixed cutoff $\bar{z}$. Note also that selection into importing reduces the optimal tariff in both the CES and Kimball economies, as in Demidova and Rodríguez-Clare (2009) and Costinot et al. (2020). On its own, however, the extensive margin distortion does not offset the terms-of-trade motive for a positive optimal tariff in the CES economy.

The table also reports the resulting welfare change in each country. Implementing the positive optimal tariff in the CES economy or in the counterfactual economies without an extensive margin adjustment increases Home welfare at the expense of Foreign welfare. In contrast, implementing the negative optimal tariff in our baseline economy raises both Home and Foreign welfare and, in fact, Foreign gains much more. This is because the optimal tariff policy leads to an exchange rate depreciation, which benefits Foreign without it having to incur the tax cost of financing import subsidies.

Table 8: Optimal Tariff

	Kimball		CES	
	moving $\bar{z}$	fixed $\bar{z}$	moving $\bar{z}$	fixed $\bar{z}$
optimal tariff	-0.081	0.296	0.162	0.277
Home consumption change, %	0.25	0.61	0.37	0.71
Foreign consumption change, %	2.29	-3.93	-1.46	-2.18
exchange rate change, %	4.78	-15.1	-8.04	-13.2

Notes: The table reports the optimal Home tariff, the resulting consumption change in each country and the change in the exchange rate.

5 Dynamic Model

We have purposely kept the model simple in order to highlight the key tradeoffs that govern the optimal tariff. We next enrich the model by introducing four additional ingredients. First, we assume that labor supply is elastic. Second, we allow for capital accumulation. Third, we model endogenous firm entry so that tariffs also affect the degree of competition in the economy. Collectively, these three ingredients imply a more elastic response of production to changes in tariffs. Fourth, we allow for trade imbalances, so exchange rates are no longer pinned down by the restriction that imports are equal to exports period by period. We consider one-time, permanent changes in Home tariffs and calculate the optimal tariff taking transition dynamics into account.

5.1 Environment

Since the dynamic model shares many of the features of the static model, we describe only the new elements of the environment and focus on the Home economy.

5.1.1 Household

The representative household in each country maximizes the discounted sum of utility defined over consumption and labor supply. The household's lifetime utility is

$$\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\theta}}{1-\theta} - \frac{L_t^{1+\phi}}{1+\phi} \right),$$

where C_t is consumption, L_t is labor supply and the parameters β, θ and ϕ are the discount factor, the relative risk aversion and the (inverse) elasticity of labor supply of the household, respectively. The household's budget constraint is

$$C_t + A_{t+1} = (1 + r_{t-1}) A_t + W_t L_t + T_t,$$

where A_t is household wealth, r_t is the interest rate, W_t is the wage rate and T_t is tariff revenue.

5.1.2 Technology

Competitive final goods producers assemble final output from N_t varieties according to

$$1 = \int_0^{N_t} \Upsilon \left(\frac{y_{it}}{Y_t} \right) \mathrm{d}i.$$

A firm with efficiency z_i now produces a differentiated variety using labor and intermediate inputs, as well as capital, with technology

$$y_{it} = z_i \left(l_{it}^{\alpha} k_{it}^{1-\alpha} \right)^{\nu} x_{it}^{1-\nu},$$

where α determines the elasticity of labor in production. As in the static model, markups increase with firm size and importing intermediates requires paying a fixed cost.

We assume that firms exit with exogenous probability φ and are replaced by a mass $\tilde{N}_t$ of new entrants who draw their efficiency from the same distribution $F(z)$ as the incumbent firms. The mass of firms evolves according to

$$N_{t+1} = (1-\varphi)\, N_t + \tilde{N}_t.$$

Entrants pay a fixed cost Φ_t to enter. We follow Gutiérrez et al. (2021) in assuming that entry costs increase with the mass of entrants. Specifically,

$$\Phi_t = \Phi \left(\frac{\tilde{N}_t}{\varphi} \right)^{\phi_n},$$

where Φ determines the average level of entry costs and ϕ_n determines the elasticity of entry rates to the value of firms. Letting Q_t denote the value of a firm, the mass of entrants is pinned down by the free entry condition

$$\Phi_t \geq Q_t,$$

with strict equality if $\tilde{N}_t > 0$.

5.1.3 Financial Intermediaries

Perfectly competitive financial intermediaries use the wealth of Home and Foreign households to finance the purchase of physical capital and shares of stocks in both countries. Because we assume no aggregate uncertainty, the returns on these assets are equalized and the portfolio composition of households is indeterminate.

No-arbitrage implies that

$$R_t = r_{t-1} + \delta \quad \text{and} \quad R_t^* = r_{t-1}^* + \delta,$$

where R_t and R_t^* are the rental rates of Home and Foreign capital, respectively, and δ is the depreciation rate. Moreover, the value of the firms in each country is the present discounted value of expected future profits

$$Q_t = \frac{1}{1+r_t} \left[(1-\varphi)\, Q_{t+1} + \pi_{t+1} \right] \quad \text{and} \quad Q_t^* = \frac{1}{1+r_t^*} \left[(1-\varphi)\, Q_{t+1}^* + \pi_{t+1}^* \right],$$

where

$$\pi_t = \int_1^{\bar{z}_t} \pi_{nt}(z)\, \mathrm{d}F(z) + \int_{\bar{z}_t}^{\infty} \pi_{mt}(z)\, \mathrm{d}F(z) - \lambda_{mt} f_m$$

are the expected profits of a firm in Home and π_t^* is similarly defined for Foreign. The uncovered interest parity condition implies that the Home and Foreign interest rates are related by

$$1 + r_t = \frac{E_{t+1}}{E_t}\left(1 + r_t^*\right) - \psi \frac{NFA_{t+1}}{Y_t},$$

where $\psi > 0$ is a small cost of holding net foreign assets that ensures that in the long run the net foreign asset position is zero (Schmitt-Grohé and Uribe, 2003).

These no-arbitrage conditions hold in all periods other than the period in which Home unexpectedly implements a change in tariffs. Since in that period the rates of return on assets are not equalized, we need to take a stand on the portfolio composition of Home and Foreign households in the initial steady state. We assume that Home households only hold Home assets and Foreign households only hold Foreign assets.

5.1.4 Current Account

The world asset market clearing condition,

$$\underbrace{A_{t+1} - \left(K_{t+1} + Q_t N_{t+1}\right)}_{NFA_{t+1}} = -E_t \underbrace{\left[A_{t+1}^* - \left(K_{t+1}^* + Q_t^* N_{t+1}^*\right)\right]}_{NFA_{t+1}^*},$$

requires that the Home and Foreign net foreign asset positions add up to zero. Equivalently, the sum of the wealth of Home and Foreign households must be equal to the sum of Home and Foreign physical capital and the stock market value of firms.

The current account identity,

$$NFA_{t+1} = \left(1 + r_{t-1}\right) NFA_t + \underbrace{\left(1 + \kappa\right) N_t^* \int_{\bar{z}_t^*}^{\infty} x_{mt}^{*h}\left(z\right) \mathrm{d}F\left(z\right)}_{\text{exports}} - \underbrace{\left(1 + \kappa\right) E_t N_t \int_{\bar{z}_t}^{\infty} x_{mt}^{f}\left(z\right) \mathrm{d}F\left(z\right)}_{\text{imports}},$$

requires that the change in the Home net foreign asset position is equal to the current account balance, which consists of net interest income and net exports.

5.2 Parameterization

We assume that there are no tariffs in the initial steady state, so the two countries are symmetric. A period in the model is one year. We assign conventional values to the discount factor, the relative risk aversion and the elasticity of labor supply, $\beta = 0.98$, $\theta = 1$ and $1/\phi = 0.5$. We set the firm exit rate to $\varphi = 0.04$ (Boar and Midrigan, 2025), the entry cost parameter to $\phi_n = 1.5$ (Gutiérrez et al., 2021) and normalize the entry cost Φ so that the number of firms is equal to one in the steady state. We set $\psi = 0.001$, in line with the small premium used by Schmitt-Grohé and Uribe (2003).

We calibrate the remaining parameters following the same strategy as in the static model. Here, we only highlight the new parameters: the parameter α that determines the labor elasticity in production and the depreciation rate δ. We target two additional moments that identify these parameters: the labor share and the capital-output ratio. Table 9 reports the results of the calibration. The original set of parameters is the same as in the static model. The new parameters are $\alpha = 0.671$ and $\delta = 0.086$.

Table 9: Dynamic Model Parameterization

A. Calibrated Parameter Values

		Kimball	CES
σ	demand elasticity at $q = 1$	20.02	7.667
ε/σ	super-elasticity of demand	0.530	0
ν	value-added elasticity	0.517	0.517
α	labor elasticity	0.671	0.671
δ	depreciation rate	0.086	0.086
η	Pareto tail of productivity	10.01	10.24
γ	home–foreign input elasticity	2.346	3.829
$\lambda_m f_m/Y$	importing costs (% of output)	4.360	0.972
κ	iceberg cost	0.409	0.177

B. Targeted Moments

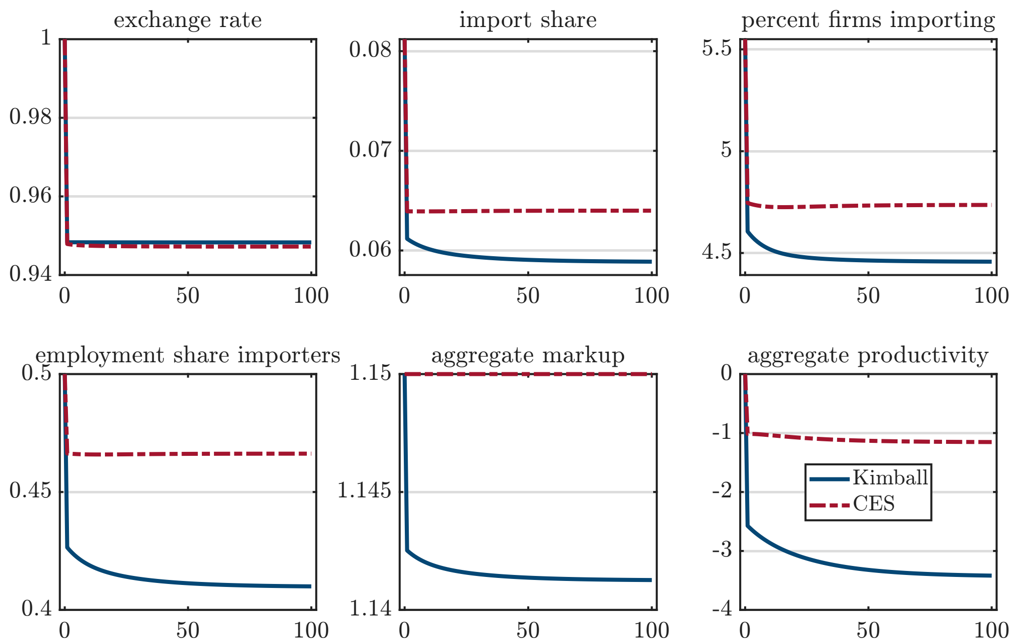
	Data	Kimball	CES
aggregate markup	1.15	1.15	1.15
pass-through elasticity	0.62	0.62	1
intermediate inputs/output	0.42	0.42	0.42
labor share in value added	0.52	0.52	0.52
capital/value added	2.41	2.41	2.41
trade elasticity	4.00	4.00	4.00
import share	0.08	0.08	0.08
percent of firms importing	5.55	5.55	5.55
employment share of importers	0.50	0.50	0.50

Notes: Pass-through is mechanically equal to one under CES.

5.3 Transition Dynamics After a 10% Increase in Home Tariffs

Before computing the optimal tariff, we study the transition dynamics following the introduction of a one-time, permanent 10% unilateral Home tariff. Figure 4 corroborates the mechanism in the static model. In response to a tariff, the exchange rate appreciates by an equal amount in both the Kimball and CES economies, but the decline in aggregate productivity is much more pronounced in the Kimball economy. This is because the tariff reduces the fraction of firms importing and their employment share, pushing the economy further away from the efficient allocation.

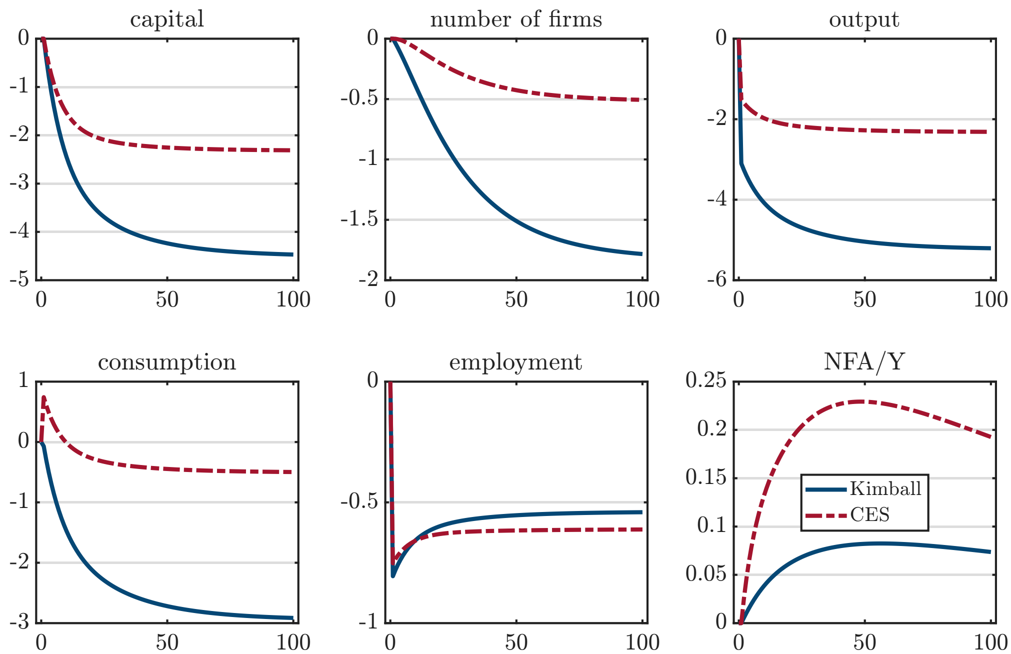
Figure 4: Transition Dynamics After a 10% Tariff



Notes: The figure reports the transition dynamics following a 10% Home tariff. All panels report Home variables. Aggregate productivity is expressed as a percentage change from the initial steady state and the remaining variables are reported in levels.

Consequently, as Figure 5 shows, the drop in output is much larger in the Kimball economy. Consumption increases on impact in the CES economy, but eventually falls below the initial steady state as the economy decumulates capital and firms. In contrast, in the Kimball economy consumption falls on impact and experiences a much larger eventual decline. The Home net foreign asset position improves in both economies, owing to the Home households' desire to smooth consumption, but the change is small.

Figure 5: Aggregate Dynamics After a 10% Tariff



Notes: The figure reports the transition dynamics following a 10% Home tariff. All panels report Home variables. Capital, the number of firms, output, consumption and employment are expressed as percentage changes from the initial steady state. The NFA is expressed as a percentage of output.

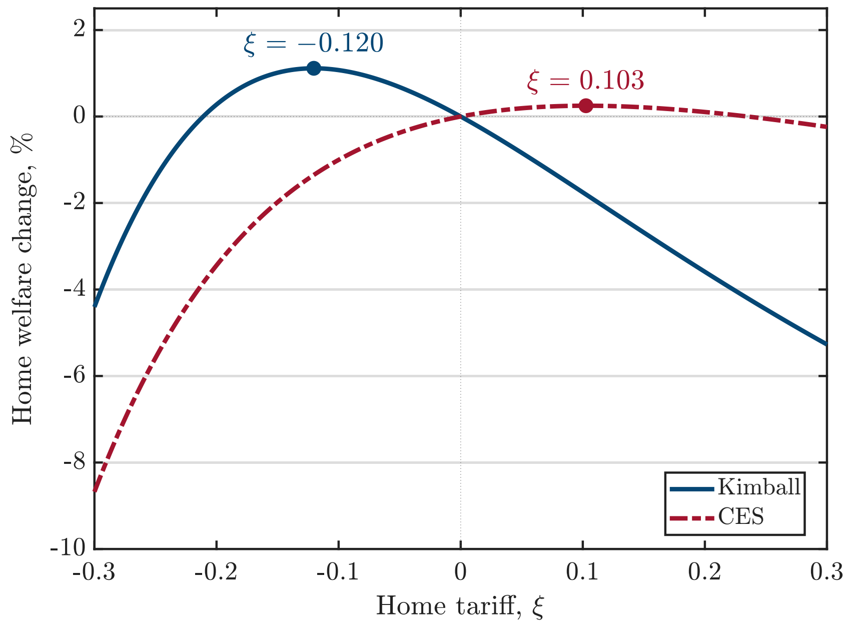
5.4 Optimal Tariff

We next compute the optimal tariff by maximizing the lifetime utility of Home households. Figure 6 illustrates how consumption-equivalent welfare in Home changes with the Home tariff. In our baseline Kimball economy, the optimal tariff is negative and equal to -12%, delivering a welfare gain of 1.11%. In sharp contrast, in the CES economy the optimal tariff is positive and equal to 10.3%, while the increase in welfare is equal to 0.25%. Relative to the static model, the additional margins of adjustment in the dynamic model (elastic labor supply, capital accumulation and firm entry) reduce the optimal tariff in both economies. Our conclusions therefore stand: the distortions associated with the extensive margin of importing in the Kimball economy are strong enough to outweigh the terms-of-trade motive, making it optimal to subsidize imports.

5.5 Role of Country Size

So far we have considered two symmetric trade partners, implying that Home accounts for 50% of world trade. We next ask how the optimal tariff varies with the size of the Home

Figure 6: Home Welfare Across Tariffs



Notes: The figure reports how the consumption-equivalent welfare change for Home varies with tariffs. The dots indicate the tariffs that maximize welfare.

economy. To that end, we extend the model to $1 + S$ economies that are symmetric in the initial steady state without tariffs. In this model, Home accounts for $1/(1+S)$ of world trade.

With multiple trading partners, the importers' bundle of intermediate inputs is now

$$x_{mt}(z) = \left(x_{mt}^{h}(z)^{\frac{\gamma-1}{\gamma}} + S x_{mt}^{f}(z)^{\frac{\gamma-1}{\gamma}}\right)^{\frac{\gamma}{\gamma-1}},$$

where $x_{mt}^{f}(z)$ is the amount of imports from each of the S foreign economies. The price of intermediate inputs for importers is now

$$P_{mt} = \left(1 + S\tau_t^{1-\gamma}\right)^{\frac{1}{1-\gamma}},$$

where, as before, $\tau_t = (1+\xi_t)(1+\kappa)E_t$ is the trade cost. Notice that as S increases, the price of imported intermediate inputs faced by Home importers decreases, which increases Home's import share. To keep the import share constant as we vary S, we adjust the iceberg cost κ so that the import share is unchanged and reproduces the data.

One important difference from the two-country model is that when Home imposes a tariff, the price of imported intermediate inputs faced by Foreign countries increases by less. To see this, notice that the price of imported intermediate inputs for Foreign countries is now

$$P_{mt}^{*} = \left[1 + (S-1)(1+\kappa)^{1-\gamma} + \left(\frac{1+\kappa}{E_t}\right)^{1-\gamma}\right]^{\frac{1}{1-\gamma}},$$

where the first term is the price of domestic intermediate inputs, the second term is the price of imports from other Foreign countries and the third term is the price of imports from Home. When S is large, imports from Home account for a small share of intermediate inputs, so each Foreign country is insulated from changes in the exchange rate E_t between Home and Foreign.

We assume that Home imposes a tariff on all of its S trading partners and that they do not retaliate. Table 10 shows how the optimal Home tariff varies with the size of Home. As the table shows, the optimal tariff decreases with the size of Home in all economies except for the CES economy with a fixed cutoff. For example, in our baseline Kimball economy, the optimal tariff is -12% when Home accounts for 50% of world trade, -3.3% when Home accounts for 25% of world trade and 19.9% when Home accounts for only 1% of world trade.

Table 10: Optimal Tariff and Country Size

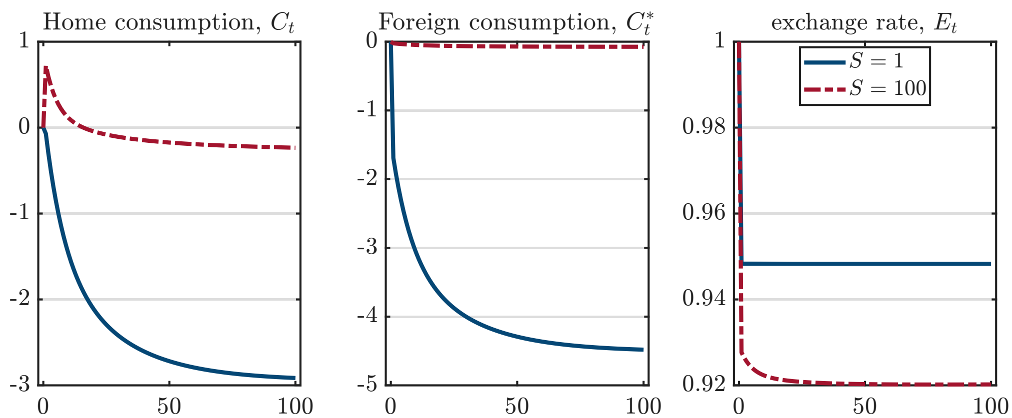
	Kimball		CES	
	moving $\bar{z}$	fixed $\bar{z}$	moving $\bar{z}$	fixed $\bar{z}$
$S = 1$	-0.120	0.212	0.103	0.243
$S = 3$	-0.033	0.294	0.158	0.233
$S = 100$	0.199	0.341	0.185	0.228

Notes: The table reports the optimal Home tariff for different values of S.

To understand the role of country size, Figure 7 shows the transition dynamics following a one-time, permanent 10% Home tariff for two versions of the Kimball economy, one with $S = 1$, and another with $S = 100$. The Home tariff leads to a much larger decline in Foreign consumption when Home is large, implying that a smaller exchange rate appreciation is needed to satisfy the current account identity. In contrast, when Home is small, the decline in Foreign consumption is much smaller, so a larger exchange rate appreciation is needed to satisfy the current account identity. As a consequence, Home's trade cost, $\tau_t = (1+\xi_t)(1+\kappa)E_t$, increases by less, strengthening the terms-of-trade effect and weakening the impact of domestic distortions. Therefore, consumption actually increases in the short run when Home is small, making a positive tariff optimal. Thus, if Home is large, it recognizes that, by restricting imports, it negatively impacts Foreign consumption and thus Foreign demand for Home exports, which in turn hurts Home as well, a force that Ostry and Rose (1989) refer to as the repercussion effect.

As Table 10 shows, the optimal Home tariff decreases less with country size in the Kimball economy with a fixed cutoff and even less in the CES economy. Moreover, the optimal tariff actually increases with country size in the CES economy with a fixed cutoff. Relative to our baseline Kimball economy, the domestic distortions are increasingly less severe in these alternative economies, so Foreign experiences smaller reductions in consumption after an increase in Home tariffs when Home is large. We illustrate this in Figure 8, which shows that in the CES economy with a fixed cutoff, the decline in Foreign consumption is much smaller when Home is large compared to our baseline Kimball economy.

Figure 7: Transition Dynamics and Country Size



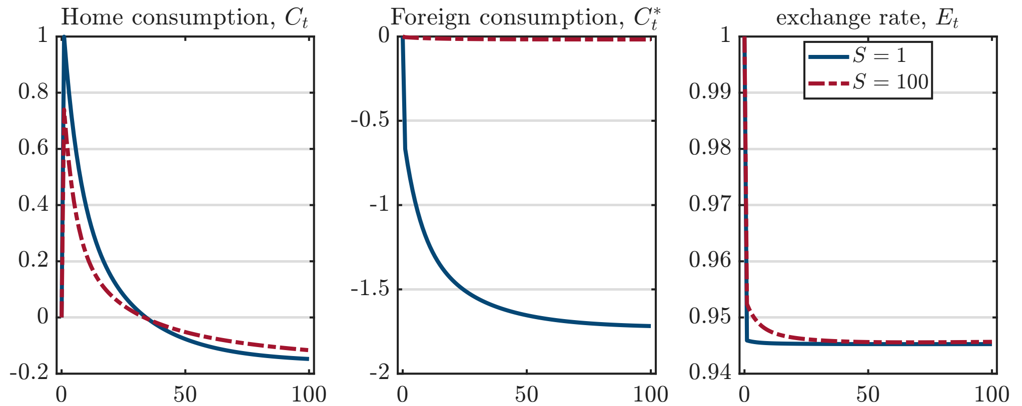
Notes: The figure reports transition dynamics in the baseline Kimball economy following a permanent 10% Home tariff. The blue solid lines correspond to $S = 1$ and the red dash-dotted lines to $S = 100$. Home and Foreign consumption are expressed as percentage changes from the initial steady state and the exchange rate is reported in levels.

6 Conclusion

We study optimal tariff policy in an economy consistent with salient features of importers in the data: they are few and large, have higher labor productivity and pass through cost changes to prices incompletely. Motivated by these facts, we develop a model with international trade in intermediate inputs that features two key ingredients. First, firms must pay a fixed cost to import intermediate inputs, so only the most productive firms import. Second, firms charge markups that increase with their size. We derive a formula that relates the optimal tariff not only to the Foreign export supply elasticity, but also to the extent to which distortions amplify the effect of trade costs on consumption.

In our quantitative analysis, we find that the optimal tariff is negative. Firm market power implies that importers are too few and too small relative to the efficient allocations,

Figure 8: Transition Dynamics and Country Size: CES with a Fixed Cutoff



Notes: The figure reports transition dynamics in the CES economy with a fixed importer cutoff following a permanent 10% Home tariff. The blue solid lines correspond to $S = 1$ and the red dash-dotted lines to $S = 100$. Home and Foreign consumption are expressed as percentage changes from the initial steady state and the exchange rate is reported in levels.

so the volume of international trade is inefficiently low. Tariffs substantially amplify this distortion by leading marginal importers to stop importing. Because these firms are large and lose a sizable cost advantage, productivity and consumption fall significantly, in contrast to economies with constant markups or without selection into importing. The same forces imply that, in contrast to the standard terms-of-trade logic, larger countries find it optimal to impose smaller tariffs, since their tariffs depress foreign demand for their exports and are thus more costly.

References

Alessandria, George A., Jiaxiaomei Ding, Shafaat Y. Khan, and Carter B. Mix, "The Tariff Tax Cut: Tariffs as Revenue," Working Paper 33784, National Bureau of Economic Research 2025.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "Importers, Exporters, and Exchange Rate Disconnect," *American Economic Review*, 2014, *104* (7), 1942–1978.

_ , _ , and _ , "International Shocks, Variable Markups, and Domestic Prices," *Review of Economic Studies*, 2019, *86* (6), 2356–2402.

_ , Stephen J. Redding, and David E. Weinstein, "The Impact of the 2018 Tariffs on Prices and Welfare," *Journal of Economic Perspectives*, 2019, *33* (4), 187–210.

Antràs, Pol, Teresa C. Fort, Agustín Gutiérrez, and Felix Tintelnot, "Trade Policy and Global Sourcing: An Efficiency Rationale for Tariff Escalation," *Journal of Political Economy Macroeconomics*, 2024, *2* (1), 1–44.

_ , _ , and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, 2017, *107* (9), 2514–2564.

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

_ , _ , Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," *Review of Economic Studies*, 2019, *86* (1), 46–80.

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

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub, "The Macroeconomics of Tariff Shocks," Working Paper 33726, National Bureau of Economic Research 2025.

Bagwell, Kyle and Seung Hoon Lee, "Trade Policy under Monopolistic Competition with Firm Selection," *Journal of International Economics*, 2020, *127*, 103379.

Baqaee, David Rezza and Emmanuel Farhi, "Productivity and Misallocation in General Equilibrium," *Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

_ and _ , "Networks, Barriers, and Trade," *Econometrica*, 2024, *92* (2), 505–541.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott, "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods," in Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data*, University of Chicago Press, 2009, pp. 513–552.

_ , _ , Stephen J. Redding, and Peter K. Schott, "Firms in International Trade," *Journal of Economic Perspectives*, 2007, *21* (3), 105–130.

_ , _ , _ , and _ , "Global Firms," *Journal of Economic Literature*, 2018, *56* (2), 565–619.

_ , Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Blaum, Joaquin, Claire Lelarge, and Michael Peters, "The Gains from Input Trade with Heterogeneous Importers," *American Economic Journal: Macroeconomics*, 2018, *10* (4), 77–127.

Boar, Corina and Virgiliu Midrigan, "Markups and Inequality," *Review of Economic Studies*, 2025, *92* (5), 2828–2860.

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

Caliendo, Lorenzo, Robert C. Feenstra, John Romalis, and Alan M. Taylor, "A Second-Best Argument for Low Optimal Tariffs on Intermediate Inputs," *Journal of International Economics*, 2023, *145*, 103824.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in "Handbook of International Economics," Vol. 4, Elsevier, 2014, pp. 197–261.

_ **and Iván Werning**, "Should We Tax Trade? A Pigouvian Perspective," *Journal of Economic Perspectives*, 2026, *40* (3), 101–122.

_ **, Andrés Rodríguez-Clare, and Iván Werning**, "Micro to Macro: Optimal Trade Policy with Firm Heterogeneity," *Econometrica*, 2020, *88* (6), 2739–2776.

Dávila, Eduardo, Andrés Rodríguez-Clare, Andreas Schaab, and Stacy Tan, "A Dynamic Theory of Optimal Tariffs," Working Paper 33898, National Bureau of Economic Research 2025.

Demidova, Svetlana, "Trade Policies, Firm Heterogeneity, and Variable Markups," *Journal of International Economics*, 2017, *108*, 260–273.

_ **and Andrés Rodríguez-Clare**, "Trade Policy under Firm-Level Heterogeneity in a Small Economy," *Journal of International Economics*, 2009, *78* (1), 100–112.

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

_ , _ , and _ , "How Costly Are Markups?," *Journal of Political Economy*, 2023, *131* (7), 1619–1675.

Epifani, Paolo and Gino Gancia, "Trade, Markup Heterogeneity and Misallocations," *Journal of International Economics*, 2011, *83* (1), 1–13.

Felbermayr, Gabriel, Benjamin Jung, and Mario Larch, "Optimal Tariffs, Retaliation, and the Welfare Loss from Tariff Wars in the Melitz Model," *Journal of International Economics*, 2013, *89* (1), 13–25.

Gopinath, Gita and Brent Neiman, "Trade Adjustment and Productivity in Large Crises," *American Economic Review*, 2014, *104* (3), 793–831.

Gros, Daniel, "A Note on the Optimal Tariff, Retaliation and the Welfare Loss from Tariff Wars in a Framework with Intra-Industry Trade," *Journal of International Economics*, 1987, *23* (3–4), 357–367.

Gutiérrez, Germán, Callum Jones, and Thomas Philippon, "Entry Costs and Aggregate Dynamics," *Journal of Monetary Economics*, 2021, *124* (S), S77–S91.

Haaland, Jan I. and Anthony J. Venables, "Optimal Trade Policy with Monopolistic Competition and Heterogeneous Firms," *Journal of International Economics*, 2016, *102*, 85–95.

Halpern, László, Miklós Koren, and Adam Szeidl, "Imported Inputs and Productivity," *American Economic Review*, 2015, *105* (12), 3660–3703.

Hummels, David, Jun Ishii, and Kei-Mu Yi, "The Nature and Growth of Vertical Specialization in World Trade," *Journal of International Economics*, 2001, *54* (1), 75–96.

Ignatenko, Anna, Ahmad Lashkaripour, Luca Macedoni, and Ina Simonovska, "Making America Great Again? The Economic Impacts of Liberation Day Tariffs," *Journal of International Economics*, 2025, *157.*

Itskhoki, Oleg and Dmitry Mukhin, "The Optimal Macro Tariff," Working Paper 33839, National Bureau of Economic Research 2026.

Johnson, Harry G., "Optimum Tariffs and Retaliation," *Review of Economic Studies*, 1953, *21* (2), 142–153.

Johnson, Robert C. and Guillermo Noguera, "Accounting for Intermediates: Production Sharing and Trade in Value Added," *Journal of International Economics*, 2012, *86* (2), 224–236.

Jones, Lin, Serge Shikher, and Yoto V. Yotov, "Determinants of Intermediate Goods Trade," Economics Working Paper 2025–12–A, U.S. International Trade Commission, Office of Economics December 2025.

Jung, Benjamin and Wilhelm Köhler, "Input-Output Linkages and Monopolistic Competition: Input Distortion and Optimal Policies," Discussion Paper 06-2021, University of Hohenheim 2021.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Klenow, Peter J. and Jonathan L. Willis, "Real Rigidities and Nominal Price Changes," *Economica*, 2016, *83* (331), 443–472.

Kocherlakota, Narayana R., "Optimal Tariffs When Labor Income Taxes Are Distortionary," Working Paper 33759, National Bureau of Economic Research 2025.

Lashkaripour, Ahmad and Volodymyr Lugovskyy, "Profits, Scale Economies, and the Gains from Trade and Industrial Policy," *American Economic Review*, 2023, *113* (10), 2759–2808.

Loecker, Jan De, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Melitz, Marc J. and Gianmarco I. P. Ottaviano, "Market Size, Trade, and Productivity," *Review of Economic Studies*, 2008, *75* (1), 295–316.

Ossa, Ralph and Stephen J. Redding, "The Economics of Tariffs," Working Paper 34915, National Bureau of Economic Research 2026.

Ostry, Jonathan D. and Andrew K. Rose, "Tariffs and the Macroeconomy: Some Empirical Evidence," IMF Working Paper 89/70, International Monetary Fund September 1989.

Schmitt-Grohé, Stephanie and Martín Uribe, "Closing Small Open Economy Models," *Journal of International Economics*, 2003, *61* (1), 163–185.

Simonovska, Ina and Michael E. Waugh, "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Yi, Kei-Mu, "Can Vertical Specialization Explain the Growth of World Trade?," *Journal of Political Economy*, 2003, *111* (1), 52–102.

Appendix

For Online Publication

A Aggregation

In this section, we derive the aggregation results used in Section 3.2.

Firm Problem. Conditional on importer status $j \in \{n, m\}$, a firm with efficiency z minimizes its variable cost

$$\min_{l_j(z), x_j(z)} W l_j(z) + P_j x_j(z)$$

subject to

$$y_j(z) = z l_j(z)^{\nu} x_j(z)^{1-\nu}.$$

The first-order conditions imply

$$\frac{W l_j(z)}{P_j x_j(z)} = \frac{\nu}{1-\nu}.$$

Letting $c_j(z)$ denote the firm's marginal cost, constant returns to scale imply that its total variable cost is $c_j(z) y_j(z)$. We can therefore write the optimal factor demands as

$$l_j(z) = \frac{\nu c_j(z) y_j(z)}{W} \quad \text{and} \quad x_j(z) = \frac{(1-\nu) c_j(z) y_j(z)}{P_j}. \tag{18}$$

Substituting these expressions into the production function and canceling $y_j(z)$ we can solve for the marginal cost

$$c_j(z) = \frac{1}{z} \underbrace{\left(\frac{W}{\nu}\right)^{\nu} \left(\frac{P_j}{1-\nu}\right)^{1-\nu}}_{\Omega_j}.$$

We next derive the firm's optimal price. Using the inverse demand curve $p_j(z) = D\Upsilon'(q_j(z))$ and the definition of $q_j(z) = y_j(z)/Y$ we can write operating profits as

$$\pi_j(z) = \left[D\Upsilon'(q_j(z)) - \frac{\Omega_j}{z}\right] q_j(z) Y.$$

The first-order condition with respect to $q_j(z)$ is

$$D\left[\Upsilon'(q_j(z)) + \Upsilon''(q_j(z)) q_j(z)\right] = \frac{\Omega_j}{z},$$

which can be rewritten as

$$p_j(z) \left(1 + \frac{\Upsilon''(q_j(z)) q_j(z)}{\Upsilon'(q_j(z))}\right) = \frac{\Omega_j}{z}.$$

It then follows that the firm's optimal price is a markup over marginal cost,

$$p_j(z) = \mu_j(z)\frac{\Omega_j}{z}, \qquad \mu_j(z) \equiv \left(1 + \frac{\Upsilon''(q_j(z))\, q_j(z)}{\Upsilon'(q_j(z))}\right)^{-1}. \tag{19}$$

For the Klenow and Willis (2016) aggregator that we use

$$\mu_j(z) = \frac{\sigma}{\sigma - q_j(z)^{\frac{\varepsilon}{\sigma}}},$$

so equation (19) gives the expression for the firm markup in equation (2).

Aggregate Markup. We define the aggregate markup to be the harmonic sales-weighted average of firm level markups,

$$\frac{1}{\mathcal{M}} = \int_1^{\bar{z}} \frac{p_n(z)\, y_n(z)}{Y}\frac{1}{\mu_n(z)}\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} \frac{p_m(z)\, y_m(z)}{Y}\frac{1}{\mu_m(z)}\mathrm{d}F(z). \tag{20}$$

Using the inverse demand curve $p_j(z) = D\Upsilon'(q_j(z))$ and the definition $q_j(z) = y_j(z)/Y$, the sales share of a firm with importer status j can be written as

$$\frac{p_j(z)\, y_j(z)}{Y} = D\Upsilon'(q_j(z))\, q_j(z).$$

Substituting into equation (20) gives

$$\frac{1}{\mathcal{M}} = D\left(\int_1^{\bar{z}} \frac{\Upsilon'(q_n(z))\, q_n(z)}{\mu_n(z)}\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} \frac{\Upsilon'(q_m(z))\, q_m(z)}{\mu_m(z)}\mathrm{d}F(z)\right), \tag{21}$$

which is the expression for the aggregate markup in Section 3.2.

Labor and Intermediate Input Shares. Combining equations (18) and (19) yields

$$Wl_j(z) = \nu\frac{p_j(z)\, y_j(z)}{\mu_j(z)} \qquad \text{and} \qquad P_j x_j(z) = (1-\nu)\frac{p_j(z)\, y_j(z)}{\mu_j(z)}. \tag{22}$$

Aggregating across firms' labor demand and using equation (21) gives

$$WL = \nu\int_1^{\bar{z}} \frac{p_n(z)\, y_n(z)}{\mu_n(z)}\mathrm{d}F(z) + \nu\int_{\bar{z}}^{\infty} \frac{p_m(z)\, y_m(z)}{\mu_m(z)}\mathrm{d}F(z) = \frac{\nu}{\mathcal{M}}Y.$$

Thus, the aggregate labor share is

$$\frac{WL}{Y} = \frac{\nu}{\mathcal{M}}. \tag{23}$$

We next derive the intermediate input share. The total spending by Home firms on intermediate inputs, inclusive of tariffs, is

$$\int_1^{\bar{z}} x_n(z)\,\mathrm{d}F(z) + \int_{\bar{z}}^{\infty} \left[x_m^h(z) + \tau x_m^f(z)\right]\mathrm{d}F(z) = \frac{1-\nu}{\mathcal{M}}Y$$

where the equality follows from aggregating firms' optimal intermediate input choices. Using the trade balance condition (6), the left-hand side is equal to $X + T$, i.e., the amount of the Home final good used as an intermediate input plus the tariff revenue. It follows that

$$X = \frac{1-\nu}{\mathcal{M}} Y - T.$$

Defining the tariff-adjusted aggregate wedge $\bar{\mathcal{M}}$ by

$$\frac{1}{\bar{\mathcal{M}}} = \frac{1}{\mathcal{M}} - \frac{T}{(1-\nu)Y}$$

then gives

$$\frac{X}{Y} = \frac{1-\nu}{\bar{\mathcal{M}}}.$$

Aggregate Productivity. Finally, define aggregate labor productivity as

$$Z \equiv \frac{Y}{L},$$

so that the aggregate production function is $Y = ZL$. Using the expression for the aggregate labor share in equation (23), we have

$$Z = \frac{Y}{L} = \frac{\mathcal{M}W}{\nu}.$$

Moreover, since $P_n = 1$, the definition of $\Omega_n = \left(\frac{W}{\nu}\right)^{\nu} \left(\frac{P_n}{1-\nu}\right)^{1-\nu}$ implies

$$W = \nu(1-\nu)^{\frac{1-\nu}{\nu}} \Omega_n^{\frac{1}{\nu}}.$$

It follows that

$$Z = \mathcal{M}(1-\nu)^{\frac{1-\nu}{\nu}} \Omega_n^{\frac{1}{\nu}}. \tag{24}$$

To express Ω_n in terms of the allocation of production across firms, recall that the optimal labor demand of a firm with importer status j can be written as

$$l_j(z) = \nu \frac{\Omega_j}{W} \frac{q_j(z)}{z} Y.$$

Aggregating labor demand across firms and using that $\Omega_m/\Omega_n = P_m^{1-\nu}$ therefore gives

$$L = \nu \frac{\Omega_n}{W} \underbrace{\left[\int_1^{\bar{z}} \frac{q_n(z)}{z} \mathrm{d}F(z) + P_m^{1-\nu} \int_{\bar{z}}^{\infty} \frac{q_m(z)}{z} \mathrm{d}F(z) \right]}_{\equiv \mathcal{Q}} Y. \tag{25}$$

Multiplying equation (25) by W/Y and using the expression for the aggregate labor share in equation (23) yields

$$\Omega_n = \frac{1}{\mathcal{M}\mathcal{Q}}.$$

Substituting this expression into equation (24) gives

$$Z = \left(\frac{1-\nu}{\mathcal{M}}\right)^{\frac{1-\nu}{\nu}} \mathcal{Q}^{-\frac{1}{\nu}} = \left(\frac{1-\nu}{\mathcal{M}}\right)^{\frac{1-\nu}{\nu}} \left(\int_1^{\bar{z}} \frac{q_n\left(z\right)}{z} \mathrm{d}F\left(z\right) + P_m^{1-\nu} \int_{\bar{z}}^{\infty} \frac{q_m\left(z\right)}{z} \mathrm{d}F\left(z\right)\right)^{-\frac{1}{\nu}},$$

which is the expression for aggregate productivity in equation (9).

B Trade Elasticity

This section derives the trade elasticity, decomposes it into intensive and extensive margins, and uses the decomposition to explain why the Kimball and CES calibrations require different values of the Home-Foreign elasticity γ.

Let ϑ denote the Foreign share of an importer's intermediate spending and let ω denote the share of aggregate intermediate spending accounted for by importers. Optimal sourcing and factor demands imply

$$\frac{\vartheta}{1-\vartheta} = \tau^{1-\gamma} \qquad \text{and} \qquad \frac{\omega}{1-\omega} = \frac{L_m}{L_n},$$

where

$$L_m = \int_{\bar{z}}^{\infty} l_m(z)\,\mathrm{d}F(z) \qquad \text{and} \qquad L_n = \int_1^{\bar{z}} l_n(z)\,\mathrm{d}F(z)$$

are total employment of importers and non-importers, respectively. Let λ denote the Home share of aggregate intermediate spending. Since only importers purchase Foreign inputs,

$$1 - \lambda = \vartheta\omega.$$

Thus, $(1-\lambda)/\lambda$ is aggregate spending on Foreign inputs relative to Home inputs. We define the trade elasticity as

$$\mathcal{A} \equiv -\frac{\mathrm{d}\log[(1-\lambda)/\lambda]}{\mathrm{d}\log\tau} = \underbrace{-\left.\frac{\partial\log[(1-\lambda)/\lambda]}{\partial\log\tau}\right|_{\bar{z}}}_{\mathcal{A}^{\text{int}}} + \underbrace{-\left.\frac{\partial\log[(1-\lambda)/\lambda]}{\partial\log\bar{z}}\right|_{\tau}\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}}_{\mathcal{A}^{\text{ext}}}.$$

The intensive margin holds importer status fixed and captures both substitution toward Home inputs within continuing importers and the contraction of continuing importers relative to non-importers. The extensive margin captures the decline in spending on Foreign inputs induced by firms switching out of importing.

Intensive Margin. Holding the importing cutoff fixed, differentiating $1-\lambda = \vartheta\omega$ gives

$$\mathcal{A}^{\text{int}} = -\frac{1}{\lambda}\left[\left.\frac{\partial \log \vartheta}{\partial \log \tau}\right|_{\bar{z}} + \left.\frac{\partial \log \omega}{\partial \log \tau}\right|_{\bar{z}}\right].$$

Using $\vartheta/(1-\vartheta) = \tau^{1-\gamma}$ and $\omega/(1-\omega) = L_m/L_n$, this becomes

$$\mathcal{A}^{\text{int}} = \frac{1-\vartheta}{\lambda}(\gamma-1) - \frac{1-\omega}{\lambda}\left.\frac{\partial \log(L_m/L_n)}{\partial \log \tau}\right|_{\bar{z}}.$$

Let λ_i denote firm i's Home input spending share,

$$\lambda_i = \begin{cases} 1-\vartheta, & \text{if } i \text{ is an importer,} \\ 1, & \text{otherwise,} \end{cases}$$

and define normalized import-share dispersion as

$$\mathcal{V} \equiv \frac{\text{Var}(\lambda_i)}{\lambda(1-\lambda)} = \frac{\vartheta(1-\omega)}{\lambda},$$

where the variance is weighted by intermediate spending. It follows that

$$1-\mathcal{V} = \frac{1-\vartheta}{\lambda} \qquad \text{and} \qquad \frac{\mathcal{V}}{\vartheta} = \frac{1-\omega}{\lambda}.$$

We can therefore write

$$\mathcal{A}^{\text{int}} = (1-\mathcal{V})(\gamma-1) + \mathcal{V}\left[-\frac{1}{\vartheta}\left.\frac{\partial \log(L_m/L_n)}{\partial \log \tau}\right|_{\bar{z}}\right]. \tag{26}$$

As in Edmond et al. (2015), $\mathcal{V}$ measures normalized dispersion in import shares and determines the weights in the intensive-margin trade elasticity. Equation (26) is a convex combination of two elasticities. The first, $\gamma-1$, captures substitution toward Home inputs within continuing importers. The second captures the contraction of continuing importers relative to non-importers. When import shares are more dispersed across firms, $\mathcal{V}$ is larger and this relative-contraction margin receives more weight. We note that those weights are directly pinned down by the data on import share, importer employment share and the share of spending on intermediate inputs, which implies that $\mathcal{V} = 0.24$. Thus, the intensive margin elasticity is primarily determined by γ.

We next characterize the contraction of continuing importers relative to non-importers. Define

$$\chi_j(z) \equiv -\frac{\partial \log q_j(z)}{\partial \log p_j(z)}\frac{\partial \log p_j(z)}{\partial \log\left(\Omega_j/(zD)\right)} = \frac{\mu_j(z)}{\mu_j(z)-1}\frac{1}{1+\frac{\varepsilon}{\sigma}\mu_j(z)},$$

and let $\bar{\chi}_m$ and $\bar{\chi}_n$ denote its employment-weighted averages among importers and non-importers. The object $\chi_j(z)$ is the firm's demand elasticity times its cost pass-through and therefore governs its quantity response to a change in normalized marginal cost.

The relative contraction of continuing importers can be written as

$$-\frac{1}{\vartheta}\left.\frac{\partial\log(L_m/L_n)}{\partial\log\tau}\right|_{\bar z}=(1-\nu)(\bar\chi_m-1)+\frac{\bar\chi_m-\bar\chi_n}{\vartheta}\left.\frac{\partial\log(\Omega_n/D)}{\partial\log\tau}\right|_{\bar z}.$$

The first term captures the direct response of continuing importers to higher trade costs. A one percent increase in τ raises importer marginal cost by $(1-\nu)\,\vartheta$ percent. Under Kimball demand, this cost increase is only partially passed through to prices, and quantities therefore fall by less than they would under constant markups. The response is governed by $\bar\chi_m$, which combines the demand elasticity with cost pass-through. The subtraction of one reflects that the object relevant for the trade elasticity is expenditure rather than quantity alone.

The second term captures reallocation among continuing firms. The importer cost shock changes the common cost Ω_n/D. This adjustment affects importers and non-importers differently whenever their demand elasticities differ. In our Kimball economy, $\bar\chi_n>\bar\chi_m$, so the decline in Ω_n/D expands non-importers more than importers and further reduces L_m/L_n.

Under CES demand, $\bar\chi_n=\bar\chi_m=\sigma$, so the reallocation term vanishes and the intensive margin trade elasticity simplifies to

$$\mathcal{A}^{\text{int}}_{\text{CES}}=(1-\mathcal{V})(\gamma-1)+\mathcal{V}(1-\nu)(\sigma-1).$$

Finally, if all firms import, there is no dispersion in firm-level import shares and $\mathcal{V}=0$. The intensive margin then reduces to the standard Armington elasticity $\gamma-1$.

Extensive Margin. The extensive margin is

$$\mathcal{A}^{\text{ext}}=-\left.\frac{\partial\log[(1-\lambda)/\lambda]}{\partial\log\bar z}\right|_{\tau}\frac{\mathrm{d}\log\bar z}{\mathrm{d}\log\tau}.$$

The size of the extensive margin therefore depends on how strongly higher trade costs shift the importing cutoff and how strongly relative spending on Foreign inputs responds to a given shift in that cutoff.

To characterize the responsiveness of Foreign spending, note that at a fixed τ, ϑ is constant. Using $1-\lambda=\vartheta\,\omega$ and $\omega/(1-\omega)=L_m/L_n$,

$$\left.\frac{\partial\log[(1-\lambda)/\lambda]}{\partial\log\bar z}\right|_{\tau}=\frac{1-\omega}{\lambda}\left.\frac{\partial\log(L_m/L_n)}{\partial\log\bar z}\right|_{\tau}.$$

A change in the cutoff affects L_m/L_n through two channels. First, firms at the cutoff switch importing status. For a change $\mathrm{d}\lambda_m$ in the mass of importers,

$$\mathrm{d}L_m=l_m(\bar z)\,\mathrm{d}\lambda_m\qquad\text{and}\qquad\mathrm{d}L_n=-l_n(\bar z)\,\mathrm{d}\lambda_m,$$

so the direct effect on relative employment is

$$\mathrm{d}\log\left(\frac{L_m}{L_n}\right) = \left[\frac{l_m(\bar{z})}{L_m} + \frac{l_n(\bar{z})}{L_n}\right]\mathrm{d}\lambda_m.$$

The two terms add because a decline in the mass of importers both reduces L_m and increases L_n.

Second, under Kimball demand, a change in the cutoff changes costs and thereby reallocates activity among continuing firms. Combining the two effects,

$$\left.\frac{\partial\log[(1-\lambda)/\lambda]}{\partial\log\bar{z}}\right|_{\tau} = \frac{1-\omega}{\lambda}\left\{\left[\frac{l_m(\bar{z})}{L_m} + \frac{l_n(\bar{z})}{L_n}\right]\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\bar{z}} + (\bar{\chi}_n - \bar{\chi}_m)\left.\frac{\partial\log(\Omega_n/D)}{\partial\log\bar{z}}\right|_{\tau}\right\}. \tag{27}$$

The first term is larger when firms around the importing cutoff are larger. The second captures the additional reallocation among continuing firms induced by a change in the cutoff. When $\bar{\chi}_n > \bar{\chi}_m$, as in the calibrated Kimball economy, non-importers expand more than importers, further reducing L_m/L_n. Under CES demand, $\bar{\chi}_n = \bar{\chi}_m = \sigma$, so this additional reallocation is absent.

We next characterize the response of the cutoff itself. The marginal importer satisfies

$$f_m = \pi_m(\bar{z}) - \pi_n(\bar{z}).$$

Differentiating this indifference condition gives

$$\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau} = (1-\nu)\vartheta\frac{l_m(\bar{z})}{l_m(\bar{z}) - l_n(\bar{z})} + \frac{\mathrm{d}\log(\Omega_n/D)}{\mathrm{d}\log\tau} - \frac{\nu f_m}{W[l_m(\bar{z}) - l_n(\bar{z})]}\frac{\mathrm{d}\log(DY)}{\mathrm{d}\log\tau}.$$

For a given sourcing technology, the cutoff is more responsive under Kimball demand. Importing lowers marginal cost, but part of this cost advantage is absorbed by a higher markup, so the gain from importing rises less steeply with productivity. At the same time, when higher trade costs cause an importer to contract, its markup and profit margin fall as well, making the gain from importing more sensitive to trade costs. Both forces make the importing cutoff respond more strongly.

Combining this with equation (27) gives

$$\mathcal{A}^{\text{ext}} = \frac{1-\omega}{\lambda}\left\{\left[\frac{l_m(\bar{z})}{L_m} + \frac{l_n(\bar{z})}{L_n}\right]\eta\lambda_m - (\bar{\chi}_n - \bar{\chi}_m)\left.\frac{\partial\log(\Omega_n/D)}{\partial\log\bar{z}}\right|_{\tau}\right\}\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}$$

where we used that $-\mathrm{d}\lambda_m = \eta\,\lambda_m\mathrm{d}\log\bar{z}$ under Pareto productivity. Thus, the extensive margin is larger when trade costs move the importing cutoff more strongly, when firms around the cutoff are larger, and, under Kimball demand, when the cutoff movement induces stronger reallocation among continuing firms.

Quantitative Decomposition. Table 11 reports the decomposition for the calibrated Kimball and CES economies. Both match an aggregate trade elasticity of four, but through very different margins.

Table 11: Trade Elasticity: Baseline Decomposition

	Kimball	CES
$\mathcal{A}$	4.000	4.000
$\mathcal{A}^{\text{int}}$	1.559	2.923
substitution within importers	1.024	2.151
contraction of continuing importers	0.536	0.772
$\mathcal{A}^{\text{ext}}$	2.441	1.077
direct switch in importer status	1.886	1.077
induced reallocation of continuing firms	0.554	0.000
Selection		
$\mathrm{d}\log\bar{z}/\mathrm{d}\log\tau$	0.264	0.309
$-\mathrm{d}\lambda_m/\mathrm{d}\log\tau$	0.147	0.176
$l_m(\bar{z})/L$	6.979	3.146
$l_n(\bar{z})/L$	3.383	1.803

Notes: The table reports local derivatives in the calibration experiment. The mass of firms is normalized to one, so employment at the cutoff is expressed relative to average employment per firm.

Under Kimball demand, much less of the trade response comes from the intensive margin (1.56 vs. 2.92) and much more from the extensive margin (2.44 vs. 1.08). The larger extensive margin does not arise because more firms stop importing. A one percent increase in trade costs raises the cutoff by 0.26% under Kimball and 0.31% under CES, so the mass of firms that stops importing is actually smaller under Kimball. Instead, each switch is substantially more consequential. Marginal importers are much larger (7.0 vs. 3.1 times average employment), and under Kimball a switch also induces additional reallocation toward continuing non-importers.

To understand why the Kimball calibration requires a lower γ, it is useful to hold γ fixed across the two economies. Table 12 recalibrates the remaining parameters to the baseline targets except for the trade elasticity.

Suppose γ were the same in the two economies. Because the import share and importer employment share are common targets, $\mathcal{V}$ is also the same. The substitution component $(1-\mathcal{V})(\gamma-1)$ would therefore be identical under Kimball and CES. This is the component of the intensive margin that changes substantially as γ varies. By contrast, the contraction of continuing importers is nearly invariant to γ and is systematically smaller under Kimball

Table 12: Trade Elasticity Holding γ Fixed

	Kimball			CES		
	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
$\mathcal{A}$	3.696	4.856	6.351	2.216	3.144	4.179
$\mathcal{A}^{\text{int}}$	1.292	2.061	2.826	1.532	2.292	3.053
substitution within importers	0.760	1.521	2.281	0.760	1.521	2.281
contraction of continuing importers	0.531	0.540	0.545	0.772	0.772	0.772
$\mathcal{A}^{\text{ext}}$	2.404	2.795	3.526	0.684	0.852	1.126
$\mathrm{d}\log\bar{z}/\mathrm{d}\log\tau$	0.203	0.377	0.549	0.088	0.211	0.329

Notes: For each value of γ, the remaining parameters are recalibrated to the baseline targets except for the trade elasticity.

because markup adjustment dampens the quantity response.

At a common γ, however, the extensive margin is much larger under Kimball. The cutoff itself also moves more. At $\gamma = 3$, for example, a one percent increase in trade costs raises $\bar{z}$ by 0.38% under Kimball and 0.21% under CES. Variable markups flatten the gain from importing as a function of efficiency, and an importer that contracts after an increase in trade costs also loses markup and profit margin. Moreover, marginal importers are larger under Kimball and their switch induces additional reallocation among continuing firms. The Kimball economy therefore generates a larger aggregate trade elasticity at any common γ (4.86 vs. 3.14 at $\gamma = 3$).

Matching the same trade elasticity of four consequently requires a lower γ under Kimball. Lowering γ primarily reduces the within-importer substitution component, while leaving the contraction of continuing importers nearly unchanged. It also increases the cost advantage from importing, which makes the gain from importing steeper in productivity and dampens the cutoff response. This is why the cutoff ultimately moves by similar amounts in the two calibrated economies and, in fact, slightly less under Kimball (0.26 vs. 0.31).

C Planner's Problem

In this section, we derive the efficient allocations described in Section 3.4. Since the planner places equal weights on Home and Foreign consumption, we focus on a symmetric allocation and write the problem in terms of consumption in one country.

Efficient Sourcing of Intermediate Inputs. We first derive the resource cost of the intermediate input bundle used by an importer. For each importer with efficiency z, the planner solves

$$\min_{x_m^h(z),\, x_m^f(z)} x_m^h(z) + (1+\kappa) x_m^f(z)$$

subject to

$$x_m(z) = \left[x_m^h(z)^{\frac{\gamma-1}{\gamma}} + x_m^f(z)^{\frac{\gamma-1}{\gamma}} \right]^{\frac{\gamma}{\gamma-1}}.$$

Letting P_m denote the multiplier on the constraint, the first-order conditions are

$$1 = P_m \left(\frac{x_m^h(z)}{x_m(z)} \right)^{-\frac{1}{\gamma}}, \qquad 1+\kappa = P_m \left(\frac{x_m^f(z)}{x_m(z)} \right)^{-\frac{1}{\gamma}}.$$

Using these conditions and the intermediate input aggregator gives

$$P_m = \left[1 + (1+\kappa)^{1-\gamma} \right]^{\frac{1}{1-\gamma}} \tag{28}$$

and total resource use equal to $x_m^h(z) + (1+\kappa) x_m^f(z) = P_m x_m(z)$. Thus, P_m measures the resource cost of one unit of the composite intermediate input for an importer.

Planner's Problem. The planner chooses the importer cutoff $\bar{z}$ and the allocation of labor and intermediate inputs across firms to maximize

$$C = Y - X - f_m \left[1 - F(\bar{z}) \right] \tag{29}$$

subject to the Kimball aggregator

$$1 = \int_1^{\bar{z}} \Upsilon(q_n(z))\, \mathrm{d}F(z) + \int_{\bar{z}}^{\infty} \Upsilon(q_m(z))\, \mathrm{d}F(z),$$

the production functions

$$q_j(z)\, Y = z l_j(z)^{\nu} x_j(z)^{1-\nu}, \qquad j \in \{n, m\},$$

and the labor and intermediate input resource constraints

$$L = \int_1^{\bar{z}} l_n(z)\, \mathrm{d}F(z) + \int_{\bar{z}}^{\infty} l_m(z)\, \mathrm{d}F(z),$$
$$X = \int_1^{\bar{z}} x_n(z)\, \mathrm{d}F(z) + P_m \int_{\bar{z}}^{\infty} x_m(z)\, \mathrm{d}F(z).$$

Notice that this is the same definition of X as in the main text. Under symmetry, $\bar{z}^* = \bar{z}$ and $x_m^{*h}(z) = x_m^f(z)$. Using the optimal sourcing condition $x_m^h(z) + (1+\kappa) x_m^f(z) = P_m x_m(z)$ then gives the expression above.

Allocation of Production. We solve the planner's problem in stages. First, fix $\bar{z}$, L and X and choose the allocation of labor and intermediate inputs across firms to maximize Y. Let YD, ω_l and ω_x denote the multipliers on the Kimball aggregator, the labor constraint and the intermediate input constraint, respectively. The first-order conditions for labor are

$$\nu YD\Upsilon'\left(q_j\left(z\right)\right)\frac{q_j\left(z\right)}{l_j\left(z\right)} = \omega_l, \qquad j \in \{n, m\},$$

and those for intermediate inputs are

$$\begin{aligned}(1-\nu)YD\Upsilon'\left(q_n\left(z\right)\right)\frac{q_n\left(z\right)}{x_n\left(z\right)} &= \omega_x,\\ (1-\nu)YD\Upsilon'\left(q_m\left(z\right)\right)\frac{q_m\left(z\right)}{x_m\left(z\right)} &= P_m\omega_x.\end{aligned}$$

Combining the labor and intermediate input conditions and using the firm production functions gives

$$\Upsilon'\left(q_j\left(z\right)\right) = \frac{\zeta}{z}P_j^{1-\nu}, \qquad j \in \{n, m\}, \tag{30}$$

where $P_n = 1$ and

$$\zeta \equiv \frac{1}{D}\left(\frac{\omega_l}{\nu}\right)^{\nu}\left(\frac{\omega_x}{1-\nu}\right)^{1-\nu}.$$

Equation (30) is the efficient allocation condition reported in the main text. Unlike the decentralized condition (3), it does not contain firm-level markups.

Aggregate Intermediate Input Use. Substituting equation (30) into the first-order conditions for labor and intermediate inputs above and aggregating across firms gives

$$L = \frac{\nu}{\omega_l}YD\zeta\left(\int_1^{\bar{z}}\frac{q_n\left(z\right)}{z}\mathrm{d}F\left(z\right) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m\left(z\right)}{z}\mathrm{d}F\left(z\right)\right)$$

and

$$X = \frac{1-\nu}{\omega_x}YD\zeta\left(\int_1^{\bar{z}}\frac{q_n\left(z\right)}{z}\mathrm{d}F\left(z\right) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m\left(z\right)}{z}\mathrm{d}F\left(z\right)\right).$$

Combining these expressions with the definition of ζ yields the aggregate production function

$$Y = AL^{\nu}X^{1-\nu}, \qquad A \equiv \left(\int_1^{\bar{z}}\frac{q_n\left(z\right)}{z}\mathrm{d}F\left(z\right) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m\left(z\right)}{z}\mathrm{d}F\left(z\right)\right)^{-1}.$$

For a given cutoff, the planner chooses X to maximize $Y - X$. The first-order condition is therefore

$$(1-\nu)AL^{\nu}X^{-\nu} = 1,$$

which implies

$$\frac{X}{Y} = 1 - \nu. \tag{31}$$

Thus, the planner chooses a larger intermediate input share than the decentralized economy, where the aggregate markup reduces the share to $(1-\nu)/\mathcal{M}$ in the absence of tariffs.

Efficient Importer Cutoff. Finally, the planner chooses which firms import by choosing the cutoff $\bar{z}$. Since the allocation of production is optimal conditional on $\bar{z}$, the envelope theorem implies that we only need to account for the direct effect of changing the cutoff on fixed costs and on the limits of integration in the three constraints. The first-order condition with respect to $\bar{z}$ is

$$F'(\bar{z})\left(f_m + DY\left(\Upsilon\left(q_n(\bar{z})\right) - \Upsilon\left(q_m(\bar{z})\right)\right) + \omega_l\left(l_m(\bar{z}) - l_n(\bar{z})\right) + \omega_x\left(P_m x_m(\bar{z}) - x_n(\bar{z})\right)\right) = 0.$$

Since $F'(\bar{z}) > 0$, dividing by $F'(\bar{z})$ and rearranging gives

$$f_m = DY\left(\Upsilon\left(q_m(\bar{z})\right) - \Upsilon\left(q_n(\bar{z})\right)\right) - \omega_l\left(l_m(\bar{z}) - l_n(\bar{z})\right) - \omega_x\left(P_m x_m(\bar{z}) - x_n(\bar{z})\right).$$

The first-order conditions for labor and intermediate inputs above imply, for $j \in \{n, m\}$,

$$\omega_l l_j(\bar{z}) = \nu DY \Upsilon'\left(q_j(\bar{z})\right) q_j(\bar{z}),$$
$$\omega_x P_j x_j(\bar{z}) = (1-\nu) DY \Upsilon'\left(q_j(\bar{z})\right) q_j(\bar{z}).$$

Therefore,

$$\omega_l\left(l_m(\bar{z}) - l_n(\bar{z})\right) + \omega_x\left(P_m x_m(\bar{z}) - x_n(\bar{z})\right) = DY\left(\Upsilon'\left(q_m(\bar{z})\right) q_m(\bar{z}) - \Upsilon'\left(q_n(\bar{z})\right) q_n(\bar{z})\right).$$

Substituting this expression into the first-order condition for $\bar{z}$ above and dividing by DY gives

$$\frac{f_m}{DY} = \Upsilon\left(q_m(\bar{z})\right) - \Upsilon\left(q_n(\bar{z})\right) - \left(\Upsilon'\left(q_m(\bar{z})\right) q_m(\bar{z}) - \Upsilon'\left(q_n(\bar{z})\right) q_n(\bar{z})\right).$$

Defining

$$\epsilon_j(z) \equiv \frac{\Upsilon\left(q_j(z)\right)}{\Upsilon'\left(q_j(z)\right) q_j(z)},$$

the efficient cutoff condition becomes

$$\frac{f_m}{DY} = \left(\epsilon_m(\bar{z}) - 1\right) \Upsilon'\left(q_m(\bar{z})\right) q_m(\bar{z}) - \left(\epsilon_n(\bar{z}) - 1\right) \Upsilon'\left(q_n(\bar{z})\right) q_n(\bar{z}),$$

as reported in the main text.

D Analytical Characterization of the Impact of Tariffs

In this section, we provide the derivations underlying the analytical characterization of the impact of tariffs in the main text. We first characterize the response of consumption to changes in trade costs and the amplification due to distortions. We then specialize this characterization to our model, derive the response of aggregate productivity and the components of Δ, characterize the response of trade costs to changes in tariffs, and finally derive the response of consumption and the optimal tariff.

D.1 Response of Consumption to Changes in Trade Costs

As in the main text, write consumption net of tariff revenue as

$$C - T = \bar{C}(a, \tau).$$

Totally differentiating gives

$$\frac{\mathrm{d}\bar{C}(a,\tau)}{\mathrm{d}\tau} = \frac{\partial \bar{C}(a,\tau)}{\partial \tau} + \nabla_a \bar{C}(a,\tau)' \frac{\mathrm{d}a}{\mathrm{d}\tau}.$$

Holding allocations fixed, a change in the trade cost affects consumption only through the direct cost of imports, so

$$\frac{\partial \bar{C}(a,\tau)}{\partial \tau} = -\mathrm{IM}.$$

It follows that, as in equation (10),

$$\frac{\mathrm{d}(C-T)}{\mathrm{d}\tau} = -(1+\Delta)\mathrm{IM}, \tag{32}$$

where

$$\Delta \equiv -\frac{1}{\mathrm{IM}} \nabla_a \bar{C}(a,\tau)' \frac{\mathrm{d}a}{\mathrm{d}\tau}. \tag{33}$$

Thus, Δ captures the effect of the change in allocations induced by higher trade costs on consumption, relative to the direct resource cost of imports. At the efficient allocation, $\nabla_a \bar{C}(a,\tau) = 0$ and therefore $\Delta = 0$.

Second-Order Approximation for Δ. A second-order approximation of $\bar{C}(a,\tau)$ around a^p gives

$$\bar{C}(a,\tau) \simeq \bar{C}(a^p,\tau) - \frac{1}{2}\hat{a}' H \hat{a},$$

where we used that $\nabla_a \bar{C}(a^p, \tau) = 0$ and where H is defined as in the main text. Therefore, we have that

$$\mathcal{L} \equiv \bar{C}(a^p, \tau) - \bar{C}(a, \tau) \simeq \frac{1}{2}\hat{a}' H \hat{a}. \tag{34}$$

Next, a first-order approximation of $\nabla_a \bar{C}(a, \tau)$ around a^p gives

$$\nabla_a \bar{C}(a, \tau) \simeq \nabla_a \bar{C}(a^p, \tau) - H\hat{a} = -H\hat{a}.$$

Using the definition of Δ and plugging in the expansion of the gradient above gives

$$\Delta = -\frac{1}{\mathrm{IM}} \nabla_a \bar{C}(a, \tau)' \frac{\mathrm{d}a}{\mathrm{d}\tau} \simeq \frac{1}{\mathrm{IM}} \hat{a}' H \frac{\mathrm{d}a}{\mathrm{d}\tau}.$$

Dividing and multiplying the right-hand side by $\frac{1}{2}\hat{a}' H \hat{a}$, letting

$$\hat{\beta} \equiv \frac{\hat{a}' H \frac{\mathrm{d}a}{\mathrm{d}\tau}}{\hat{a}' H \hat{a}}$$

and using equation (34) gives

$$\Delta \simeq \frac{2\mathcal{L}}{\mathrm{IM}} \hat{\beta}.$$

D.2 Determinants of Δ

Recall from equation (11) that in our model consumption net of tariff revenue is

$$C - T = WL + \Pi = \left(1 - \frac{1-\nu}{\mathcal{M}}\right) Z - \lambda_m f_m.$$

We first derive how aggregate productivity responds to trade costs and then use this result to derive the components of Δ.

Response of Aggregate Productivity to Changes in Trade Costs. Recall from Appendix A that

$$Z = \left(\frac{1-\nu}{\mathcal{M}}\right)^{\frac{1-\nu}{\nu}} \mathcal{Q}^{-\frac{1}{\nu}},$$

where, from equation (25),

$$\mathcal{Q} = \int_1^{\bar{z}} \frac{q_n(z)}{z} \mathrm{d}F(z) + P_m^{1-\nu} \int_{\bar{z}}^{\infty} \frac{q_m(z)}{z} \mathrm{d}F(z).$$

Taking logs and totally differentiating with respect to the trade cost τ gives

$$\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau} = -\frac{1-\nu}{\nu} \frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau} - \frac{1}{\nu} \frac{\mathrm{d}\log\mathcal{Q}}{\mathrm{d}\log\tau}. \tag{35}$$

It therefore remains to derive the response of $\mathcal{Q}$. Applying Leibniz's rule and using $\mathrm{d}\bar{z} = \bar{z}\,\mathrm{d}\log\bar{z}$ gives

$$\frac{\mathrm{d}\mathcal{Q}}{\mathrm{d}\log\tau} = P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\mathrm{d}F(z)\,\frac{\mathrm{d}\log\left(P_m^{1-\nu}\right)}{\mathrm{d}\log\tau}$$
$$+\int_1^{\bar{z}}\frac{q_n(z)}{z}\frac{\mathrm{d}\log q_n(z)}{\mathrm{d}\log\tau}\mathrm{d}F(z) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\frac{\mathrm{d}\log q_m(z)}{\mathrm{d}\log\tau}\mathrm{d}F(z)$$
$$+F'(\bar{z})\left[q_n(\bar{z}) - P_m^{1-\nu}q_m(\bar{z})\right]\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}. \tag{36}$$

The last term reflects the fact that the importing cutoff changes with τ. Add and subtract $q_m(\bar{z})$ to write

$$q_n(\bar{z}) - P_m^{1-\nu}q_m(\bar{z}) = \left[q_n(\bar{z}) - q_m(\bar{z})\right] + \left(1 - P_m^{1-\nu}\right)q_m(\bar{z}).$$

Substituting this expression into equation (36) and using $\mathrm{d}\log\left(P_m^{1-\nu}\right) = (1-\nu)\mathrm{d}\log P_m$ gives

$$\frac{\mathrm{d}\log\mathcal{Q}}{\mathrm{d}\log\tau} = \frac{1}{\mathcal{Q}}\Bigg[(1-\nu)P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\mathrm{d}F(z)\,\frac{\mathrm{d}\log P_m}{\mathrm{d}\log\tau}$$
$$+\int_1^{\bar{z}}\frac{q_n(z)}{z}\frac{\mathrm{d}\log q_n(z)}{\mathrm{d}\log\tau}\mathrm{d}F(z) + P_m^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_m(z)}{z}\frac{\mathrm{d}\log q_m(z)}{\mathrm{d}\log\tau}\mathrm{d}F(z)$$
$$+F'(\bar{z})\left[q_n(\bar{z}) - q_m(\bar{z})\right]\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}$$
$$+F'(\bar{z})\left(1 - P_m^{1-\nu}\right)q_m(\bar{z})\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}\Bigg]. \tag{37}$$

We next rewrite each component in the form used in the main text.

Direct Effect of Importing Costs. Recall that

$$P_m = \left(1 + \tau^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$$

and that

$$\vartheta \equiv \frac{\tau x_m^f(z)}{P_m x_m(z)} = \frac{\tau^{1-\gamma}}{1+\tau^{1-\gamma}}$$

is the share of imported intermediate inputs in an importer's total spending on intermediates. Differentiating the expression for P_m gives

$$\frac{\mathrm{d}\log P_m}{\mathrm{d}\log\tau} = \vartheta. \tag{38}$$

From the firm's optimal intermediate input demand,

$$P_m x_m(z) = (1-\nu)\frac{\Omega_m}{z}y_m(z) = (1-\nu)\Omega_n P_m^{1-\nu}\frac{q_m(z)}{z}Y.$$

Therefore, aggregate spending on imported intermediate inputs is

$$\tau\,\mathrm{IM} = \vartheta(1-\nu)\Omega_n P_m^{1-\nu} Y \int_{\bar{z}}^{\infty} \frac{q_m(z)}{z} \mathrm{d}F(z)\,.$$

Total spending by Home firms on intermediate inputs, inclusive of tariffs, is

$$\frac{1-\nu}{\mathcal{M}} Y = (1-\nu)\Omega_n \mathcal{Q} Y,$$

where the equality uses $\Omega_n = 1/(\mathcal{M}\mathcal{Q})$ from Appendix A. Since λ denotes the domestic intermediate spending share, it follows that

$$1-\lambda = \vartheta \frac{P_m^{1-\nu} \displaystyle\int_{\bar{z}}^{\infty} \frac{q_m(z)}{z} \mathrm{d}F(z)}{\mathcal{Q}}. \tag{39}$$

Combining equations (38) and (39), the first term in equation (37) becomes

$$\frac{1}{\mathcal{Q}}(1-\nu) P_m^{1-\nu} \int_{\bar{z}}^{\infty} \frac{q_m(z)}{z} \mathrm{d}F(z) \frac{\mathrm{d}\log P_m}{\mathrm{d}\log \tau} = (1-\nu)(1-\lambda).$$

It therefore contributes

$$-\frac{1-\nu}{\nu}(1-\lambda)$$

to the response of aggregate productivity in equation (35). The main text also writes this term as

$$-\frac{1-\nu}{\nu} \frac{1}{\mathcal{A}} \frac{\mathrm{d}\log \lambda}{\mathrm{d}\log \tau},$$

where $\mathcal{A}$ is the trade elasticity such that

$$\mathcal{A} = -\frac{\mathrm{d}\log\left((1-\lambda)/\lambda\right)}{\mathrm{d}\log \tau}.$$

Using

$$\mathrm{d}\log \frac{1-\lambda}{\lambda} = -\frac{1}{1-\lambda} \mathrm{d}\log \lambda,$$

gives

$$1-\lambda = \frac{1}{\mathcal{A}} \frac{\mathrm{d}\log \lambda}{\mathrm{d}\log \tau}.$$

Hence

$$-\frac{1-\nu}{\nu}(1-\lambda) = -\frac{1-\nu}{\nu} \frac{1}{\mathcal{A}} \frac{\mathrm{d}\log \lambda}{\mathrm{d}\log \tau}. \tag{40}$$

Aggregate Markup. The aggregate-markup contribution follows directly from equation (35) and is

$$-\frac{1-\nu}{\nu} \frac{\mathrm{d}\log \mathcal{M}}{\mathrm{d}\log \tau}. \tag{41}$$

Quantity Reallocation. Using the definition of $\mathcal{R}$ in the main text,

$$\mathcal{R}=\int_{1}^{\bar{z}}\frac{q_{n}\left(z\right)}{z}\frac{\mathrm{d}\log q_{n}\left(z\right)}{\mathrm{d}\log\tau}\mathrm{d}F\left(z\right)+P_{m}^{1-\nu}\int_{\bar{z}}^{\infty}\frac{q_{m}\left(z\right)}{z}\frac{\mathrm{d}\log q_{m}\left(z\right)}{\mathrm{d}\log\tau}\mathrm{d}F\left(z\right)+F'\left(\bar{z}\right)\left[q_{n}\left(\bar{z}\right)-q_{m}\left(\bar{z}\right)\right]\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau},$$

the quantity-reallocation terms contribute

$$-\frac{1}{\nu\mathcal{Q}}\mathcal{R}\tag{42}$$

to the response of aggregate productivity.

Loss of Importing Technology. The remaining term in equation (37) is

$$\frac{F'\left(\bar{z}\right)\left(1-P_{m}^{1-\nu}\right)q_{m}\left(\bar{z}\right)}{\mathcal{Q}}\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}.$$

Since

$$\lambda_{m}=1-F\left(\bar{z}\right),$$

we have

$$-\frac{\mathrm{d}\lambda_{m}}{\mathrm{d}\log\tau}=\bar{z}F'\left(\bar{z}\right)\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}.\tag{43}$$

Moreover, using

$$l_{m}(z)=\nu\frac{\Omega_{n}}{W}P_{m}^{1-\nu}\frac{q_{m}(z)}{z}Y$$

and equation (25),

$$L=\nu\frac{\Omega_{n}}{W}\mathcal{Q}Y,$$

we obtain, after taking the ratio and evaluating at $z=\bar{z}$,

$$\frac{l_{m}\left(\bar{z}\right)}{L}=\frac{P_{m}^{1-\nu}q_{m}\left(\bar{z}\right)}{\bar{z}\mathcal{Q}}.\tag{44}$$

Using equations (43) and (44), we can write

$$\begin{aligned}&\frac{F'\left(\bar{z}\right)\left(1-P_{m}^{1-\nu}\right)q_{m}\left(\bar{z}\right)}{\mathcal{Q}}\frac{\mathrm{d}\log\bar{z}}{\mathrm{d}\log\tau}\\&=\left(-\frac{\mathrm{d}\lambda_{m}}{\mathrm{d}\log\tau}\right)\frac{l_{m}\left(\bar{z}\right)}{L}\left(\frac{1}{P_{m}^{1-\nu}}-1\right).\end{aligned}$$

It therefore contributes

$$-\frac{1}{\nu}\left(-\frac{\mathrm{d}\lambda_{m}}{\mathrm{d}\log\tau}\right)\frac{l_{m}\left(\bar{z}\right)}{L}\left(\frac{1}{P_{m}^{1-\nu}}-1\right)\tag{45}$$

to the response of aggregate productivity. Combining these results gives

$$\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau} = -\underbrace{\frac{1-\nu}{\nu}(1-\lambda)}_{\text{ACR}} - \underbrace{\frac{1-\nu}{\nu}\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}}_{\text{markup response}} - \underbrace{\frac{1}{\nu\mathcal{Q}}\mathcal{R}}_{\text{quantity reallocation}}$$

$$-\underbrace{\frac{1}{\nu}\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\frac{l_m(\bar{z})}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)}_{\text{loss of importing technology}}, \tag{46}$$

which is equation (12) in the main text.

Components of Δ. Starting from equation (11) and using $L = 1$, so that $Y = Z$, totally differentiating with respect to $\log\tau$ gives

$$\frac{\mathrm{d}(C-T)}{\mathrm{d}\log\tau} = \left(1-\frac{1-\nu}{\mathcal{M}}\right)Z\frac{\mathrm{d}\log Z}{\mathrm{d}\log\tau} + \frac{1-\nu}{\mathcal{M}}Z\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau} - f_m\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}. \tag{47}$$

By equation (10),

$$\frac{\mathrm{d}(C-T)}{\mathrm{d}\log\tau} = -(1+\Delta)\tau\,\mathrm{IM}. \tag{48}$$

To rewrite the first term in equation (46), note from equation (39) that

$$\tau\,\mathrm{IM} = \frac{1-\nu}{\mathcal{M}}(1-\lambda)Z. \tag{49}$$

Substituting equation (46) into equation (47), the contribution of the ACR term is

$$\begin{aligned}
&-\left(1-\frac{1-\nu}{\mathcal{M}}\right)Z\frac{1-\nu}{\nu}(1-\lambda)\\
=\ &-\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{\mathcal{M}}{\nu}\tau\,\mathrm{IM} = -\left(1+\frac{\mathcal{M}-1}{\nu}\right)\tau\,\mathrm{IM}.
\end{aligned}$$

The first part, $-\tau\,\mathrm{IM}$, is the direct resource cost already included in equation (48). The remaining part therefore contributes

$$\frac{\mathcal{M}-1}{\nu} \tag{50}$$

to Δ. The two terms involving the response of the aggregate markup combine to

$$\begin{aligned}
&-\left(1-\frac{1-\nu}{\mathcal{M}}\right)Z\frac{1-\nu}{\nu}\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau} + \frac{1-\nu}{\mathcal{M}}Z\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}\\
=\ &-Z\frac{1-\nu}{\nu}\left(1-\frac{1}{\mathcal{M}}\right)\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}.
\end{aligned}$$

Using equation (48), this contributes

$$\frac{Z}{\tau\,\mathrm{IM}}\frac{1-\nu}{\nu}\left(1-\frac{1}{\mathcal{M}}\right)\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau} \tag{51}$$

to Δ. The quantity-reallocation term contributes

$$\frac{Z}{\tau\,\mathrm{IM}}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{1}{\nu\mathcal{Q}}\mathcal{R} \tag{52}$$

to Δ. Finally, importer exit affects consumption through the loss of importing technology and through the saving of the fixed importing cost. Combining these two terms gives

$$-\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\left[\frac{Z}{\nu}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{l_m\left(\bar{z}\right)}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)-f_m\right].$$

It follows that importer exit contributes

$$\frac{1}{\tau\,\mathrm{IM}}\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\left[\frac{Z}{\nu}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{l_m\left(\bar{z}\right)}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)-f_m\right] \tag{53}$$

to Δ. Combining the four terms yields

$$\begin{aligned}\Delta = {} & \underbrace{\frac{\mathcal{M}-1}{\nu}}_{\text{markup level}} + \underbrace{\frac{Z}{\tau\,\mathrm{IM}}\frac{1-\nu}{\nu}\left(1-\frac{1}{\mathcal{M}}\right)\frac{\mathrm{d}\log\mathcal{M}}{\mathrm{d}\log\tau}}_{\text{markup response}} \\ & + \underbrace{\frac{Z}{\tau\,\mathrm{IM}}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{1}{\nu\mathcal{Q}}\mathcal{R}}_{\text{quantity reallocation}} \\ & + \underbrace{\frac{1}{\tau\,\mathrm{IM}}\left(-\frac{\mathrm{d}\lambda_m}{\mathrm{d}\log\tau}\right)\left[\frac{Z}{\nu}\left(1-\frac{1-\nu}{\mathcal{M}}\right)\frac{l_m\left(\bar{z}\right)}{L}\left(\frac{1}{P_m^{1-\nu}}-1\right)-f_m\right]}_{\text{importer exit}}.\end{aligned} \tag{54}$$

D.3 Response of Trade Costs to Changes in Tariffs

Recall that

$$\tau = (1+\xi)(1+\kappa)E.$$

Since the Foreign tariff is $\xi^* = 0$,

$$\tau^* = \frac{1+\kappa}{E}.$$

Therefore,

$$\mathrm{d}\log\tau = \mathrm{d}\log(1+\xi) + \mathrm{d}\log E, \qquad \mathrm{d}\log\tau^* = -\mathrm{d}\log E. \tag{55}$$

From the trade balance condition (6),

$$\mathrm{IM} = \frac{\mathrm{EX}}{E}.$$

Using the definitions

$$\mathcal{D} = -\frac{\mathrm{d}\log \mathrm{IM}}{\mathrm{d}\log \tau}, \qquad \mathcal{E} = -\frac{\mathrm{d}\log(\mathrm{EX}/E)}{\mathrm{d}\log \tau^*},$$

we have

$$-\mathcal{D}\,\mathrm{d}\log \tau = -\mathcal{E}\,\mathrm{d}\log \tau^*.$$

Using equation (55) therefore gives

$$-\mathcal{D}\,\mathrm{d}\log \tau = \mathcal{E}\,\mathrm{d}\log E. \tag{56}$$

Substituting $\mathrm{d}\log \tau = \mathrm{d}\log(1+\xi) + \mathrm{d}\log E$ into equation (56) gives

$$(\mathcal{D} + \mathcal{E})\,\mathrm{d}\log E = -\mathcal{D}\,\mathrm{d}\log(1+\xi).$$

Hence

$$\frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)} = -\frac{\mathcal{D}}{\mathcal{D}+\mathcal{E}}, \tag{57}$$

which is equation (14) in the main text. It then follows that

$$\begin{aligned}\frac{\mathrm{d}\log \tau}{\mathrm{d}\log(1+\xi)} &= 1 + \frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)} \\ &= 1 - \frac{\mathcal{D}}{\mathcal{D}+\mathcal{E}} = \frac{\mathcal{E}}{\mathcal{D}+\mathcal{E}}.\end{aligned}$$

Thus,

$$\frac{\mathrm{d}\log \tau}{\mathrm{d}\log(1+\xi)} = \frac{\mathcal{E}}{\mathcal{D}+\mathcal{E}}, \tag{58}$$

which is equation (15) in the main text.

D.4 Response of Consumption to Changes in Tariffs

Tariff revenue is

$$T = \xi(1+\kappa)E\,\mathrm{IM}.$$

Using $\tau = (1+\xi)(1+\kappa)E$, we can write

$$T = \frac{\xi}{1+\xi}\tau\,\mathrm{IM} = [\tau - (1+\kappa)E]\,\mathrm{IM}.$$

Consumption is therefore

$$C = (C - T) + [\tau - (1+\kappa)E]\,\mathrm{IM}. \tag{59}$$

Totally differentiating gives

$$\mathrm{d}C = \mathrm{d}(C - T) + \mathrm{IM}\,\mathrm{d}\tau - (1+\kappa)\mathrm{IM}\,\mathrm{d}E + [\tau - (1+\kappa)E]\,\mathrm{dIM}. \tag{60}$$

Equation (10) implies

$$\mathrm{d}(C - T) = -(1+\Delta)\mathrm{IM}\,\mathrm{d}\tau.$$

Substituting this expression into equation (60), the direct $-\mathrm{IM}\,\mathrm{d}\tau$ term cancels the $+\mathrm{IM}\,\mathrm{d}\tau$ term from tariff revenue. We obtain

$$\mathrm{d}C = -(1+\kappa)\mathrm{IM}\,\mathrm{d}E - \Delta\,\mathrm{IM}\,\mathrm{d}\tau + [\tau - (1+\kappa)E]\,\mathrm{dIM}. \tag{61}$$

Dividing by C and differentiating with respect to $\log(1+\xi)$, and using

$$\tau = (1+\xi)(1+\kappa)E, \qquad \tau - (1+\kappa)E = \xi(1+\kappa)E,$$

gives

$$\frac{\mathrm{d}\log C}{\mathrm{d}\log(1+\xi)} = \frac{(1+\kappa)E\,\mathrm{IM}}{C}\left(-\frac{\mathrm{d}\log E}{\mathrm{d}\log(1+\xi)} + \xi\frac{\mathrm{d}\log \mathrm{IM}}{\mathrm{d}\log(1+\xi)} - \Delta(1+\xi)\frac{\mathrm{d}\log\tau}{\mathrm{d}\log(1+\xi)}\right) \tag{62}$$

which is equation (16) in the main text.

D.5 Optimal Tariff

Finally, the definition of $\mathcal{D}$ and equation (58) imply

$$\frac{\mathrm{d}\log \mathrm{IM}}{\mathrm{d}\log(1+\xi)} = -\mathcal{D}\frac{\mathcal{E}}{\mathcal{D}+\mathcal{E}}. \tag{63}$$

Substituting equations (57), (58), and (63) into equation (62) gives

$$\frac{\mathrm{d}\log C}{\mathrm{d}\log(1+\xi)} = \frac{(1+\kappa)E\,\mathrm{IM}}{C(\mathcal{D}+\mathcal{E})}\left[\mathcal{D} - \xi\mathcal{D}\mathcal{E} - (1+\xi)\mathcal{E}\Delta\right]. \tag{64}$$

At an interior optimum, the derivative of consumption with respect to the tariff is equal to zero. The first-order condition is therefore

$$\mathcal{D} - \xi\mathcal{D}\mathcal{E} - (1+\xi)\mathcal{E}\Delta = 0. \tag{65}$$

Rearranging,

$$\mathcal{D} - \mathcal{E}\Delta = \xi\mathcal{E}\left(\mathcal{D}+\Delta\right),$$

so that

$$\xi = \frac{1}{\mathcal{E}}\frac{\mathcal{D} - \mathcal{E}\Delta}{\mathcal{D}+\Delta}. \tag{66}$$

This is equation (17) in the main text. Absent distortions, $\Delta = 0$, and the expression reduces to

$$\xi = \frac{1}{\mathcal{E}}.$$

More generally, equation (66) is an implicit expression because $\mathcal{D}$, $\mathcal{E}$, and Δ are equilibrium objects evaluated at the candidate tariff.