Who Pays for Markups in a Global Economy?
The Unequal Impacts of International Rent-Shifting

Siying Ding
Indiana University

Ahmad Lashkaripour
Indiana University

Volodymyr Lugovskyy
Indiana University

First Version: August 2021
This version: June 2022

Abstract

Trade openness over the past decades has prompted tremendous production relocation across international borders and industries. These developments, we argue, have triggered a systematic transfer of markup rents and welfare from low- to high-income nations. We establish this claim by providing semi-parametric formulas for the welfare cost of markups in open economies. Our formulas indicate that trade can magnify the incidence of markups on individual countries by (1) increasing markup dispersion relative to autarky or (2) shifting markup rents from one country to another (rent-shifting). We map these formulas to data by estimating firm-level markups for many countries, using demand-based and cost-based estimation techniques. Our analysis reveals that trade openness has increased the welfare cost of markups among low-income countries by 21% while lowering it by 10% among high-income nations. These unequal effects are driven entirely by rent-shifting from low-income countries to high-income trading partners. We show that these rent-shifting effects are akin to *hidden* tariffs that violate reciprocity and favor high-income nations.

1 Introduction

Policy-makers have frequently raised concerns over *unfair* trade in recent years. These concerns reflect a prevalent zero-sum mentality, whereby existing trade arrangements benefit some countries while harming others. These worries are by no means confined to small emerging economies: The United States' withdrawal from the Trans-Pacific Partnership (TPP) was primarily motivated by concerns over unfair trade (Chow et al. (2018)).

From the lens of economic theory, these concerns are unwarranted when national economies are efficient and free of distortions. All countries, in such circumstances, benefit from trade without exception. But in *second-best* economies, trade can exacerbate or alleviate pre-existing distortions in ways that benefit some countries at the expense of others (Bhagwati and Ramaswami, 1963).

Among the myriad distortions that plague open economies, markups have become increasingly prominent over time (De Loecker et al., 2020). Markup distortions are also distinct from other distortions in a key aspect. While costly, markups generate rents (or profits) that are rebated to consumers and partially mitigate the efficiency cost associated with them.[1]

In a globalized world, the distortionary cost of markups is often borne by households in one location, while the resulting rents accrue to households elsewhere. So, in addition to modifying the global cost of markups, trade can shift the incidence of the cost from one country to another through *international rent-shifting*. These are zero-sum effects that create national-level winners and losers and conceivably fuel some of the global shift in attitude towards trade.

International rent-shifting is more consequential when globalization induces systematic specialization across low and high-markup industries. Countries that specialize in low-markup industries become net payers of markup rents to the rest of the world. The incidence of global markups, as a result, falls disproportionately on their households, who are entitled to lesser rents or profit rebates. A casual analysis of firm-level data indicates that this type of specialization is prevalent. Over the past two decades, low-income countries have increasingly specialized in industries with relatively thin profit margins, with high-income nations dominating high-profit margin industries.

Against this backdrop, the current literature has given little attention to international rent-shifting, focusing instead on the *pro-competitive* and *pro-productivity* effects of trade in markup-distorted economies. This gap reflects two technical difficulties. First, measuring the rent-shifting effects of trade requires firm-level markup estimation across a wide range of countries and industries. Until recently, we were short of tools and data to conduct this type of large-scale estimation. Second, isolating the rent-shifting effects of trade can be difficult without a solid theoretical foundation, even from a purely computational standpoint.

We bridge this gap by developing a simple technique to measure the welfare consequences of international rent-shifting. Our approach involves three basic steps:

First, we derive sufficient statistics formulas for the welfare cost of markups in open economies. These formulas determine how trade influences the cost of markups by (1) changing the degree of markup dispersion relative to autarky and (2) international rent-shifting. Despite their simple configuration, our formulas internalize salient features of the global economy, such as input-output linkages and rich within-industry firm heterogeneity. They, moreover, reduce the data requirements for our analysis to a set of observable shares and average markups.

Second, we capitalize on recent advancements in the literature to estimate markups across a wide range of countries and industries. Our analysis uses both the demand-based and cost-based approaches to markup estimation. We provide the first large-scale comparison of these approaches, which had proven difficult in the face of previous data limitations. We, encouragingly, detect notable similitudes between the predictions of cost- and demand-based estimation techniques.

Third, We plug the estimated markups into our formula to measure the impact of trade on the cost of markups for 65 major economies. We find that trade openness has increased the cost of markups

[1] Distortions that reflect financial or contracting frictions, for instance, may generate quasi-rents that are competed away and leave the economy as dead-weight consumption loss.

among low-income countries by 21% but has lowered it among high-income countries by 10%.[2] These asymmetric effects are driven primarily by rent-shifting from low- to high-income countries. Intuitively, this happens because low-income countries specialize in less-sophisticated, low-markup industries, while the opposite is true for high-income countries.

We then show that rent-shifting effects are akin to hidden protectionism—i.e., an excess tariff charged by high-income countries on low-income partners. Two types of policy reform can neutralize these hidden tariffs. One option is that governments mitigate markup distortions via domestic policies, but the World Trade Organization effectively prohibits this option. Alternatively, high-income countries could unilaterally lower their tariffs on low-income countries by more than 10%.

The Simple Analytics of International Rent-Shifting. Our formula for the welfare cost of markups elucidates the role of international rent-shifting. Consider a multi-industry Krugman (1980) model with restricted entry, which is the simplest model that admits international rent-shifting effects. Let $(1/\mu)$ denote the vector of firm-level inverse markups aggregated by industry. We show that the welfare cost of markups for an open economy can be decomposed as

$$\text{Cost of Markups} \approx \underbrace{\frac{1}{2}\left[\text{CV}_e\left(\frac{1}{\mu}\right)\right]^2}_{\text{markup dispersion}} + \underbrace{\mu^{avg} \times \text{Cov}_e\left(\frac{1}{\mu}, \text{RCA}\right)}_{\text{international rent-shifting}}.$$

Markup dispersion represents the expenditure-weighted coefficient of variation (CV_e) of inverse markups. This term is well-understood and the sole driver of welfare cost in a closed economy. *International rent-shifting* represents the inter-industry covariance between the index of revealed comparative advantage (RCA) and inverse markup.[3] It is lower if a country specializes in high-markup (high-profit margin) industries. A similar decomposition, we show, continues to hold in more general models featuring firm selection into export markets à la Melitz (2003), variable markups à la Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019), and input-output linkages à la Caliendo and Parro (2015).[4]

Digging deeper, international rent-shifting reflects a *decoupling* between cost bearing and rent rebates. In a closed economy, markup rents eventually accrue to households that bear the cost of markup distortions. In a globalized economy, however, the cost of markups may be borne by households in one location, with profits rebated to households elsewhere. Consider, for instance, US-based firms that export high-markup goods to international markets. Foreign households that purchase these goods bear the cost of markup distortions, but the corresponding markup rents are rebated to

[2] We observe heterogeneity not only across but also within income groups. While trade contributes to a significant reduction in the cost of markups for the modal high-income country, some like Canada experience an increase in cost.

[3] To be specific, RCA denotes the ratio of industry-level production to consumption, which is a theory-consistent index for comparative advantage in our framework. This index is one of many indexes of revealed comparative advantage used in the literature—see Ballance et al. (1987). Compared to the widely-used Balassa (1965) index, it has the advantage of using information on domestic output/expenditure in addition to export values.

[4] Trade openness can modify the cost of markups by changing the degree of *markup dispersion* and *international rent-shifting*. The dispersion term is invariant to trade in our baseline model where industry-level expenditure shares are constant Cobb-Douglas weights. Markup dispersion, however, is impacted by trade openness in the richer environments we consider, such as those with CES preferences across industries or input-output linkages.

US nationals. Extrapolating from this example, countries that specialize in high-markup industries are net receivers of markup rents from the rest of the world and, thus, experience a lower cost from markups. By contrast, countries that specialize in low-markup industries pay net rents to the rest of the world and are exposed to a relatively higher welfare cost from markups. The rent-shifting term in our simple formula accounts for these asymmetric effects.

Notably, one can measure international rent-shifting without taking a stance on the drivers of inter-industry specialization. The index of revealed comparative advantage, RCA, is a sufficient and observable statistic that conveys all the information one needs about a country's production capabilities across industries. Interacting this index with estimated firm-level markups determines the cost of international rent-shifting irrespective of the underlying production parameters and the origins of comparative advantage.

Related Literature. The existing literature on trade and (markup) distortions has primarily focused on (1) inter-firm reallocation and (2) pro-competitive effects of trade liberalization.[5] Given this focus, existing studies generally speak to how trade modifies the dispersion in markup or non-markup wedges relative to autarky. On the former issue, Epifani and Gancia (2011), Bai, Jin, and Lu (2019), Chung (2019), and Berthou, Chung, Manova, and Bragard (2020) examine how trade-induced reallocation across firms with different productivity and distortion levels impacts aggregate productivity and welfare. On the latter issue, Holmes, Hsu, and Lee (2014), De Blas and Russ (2015), Edmond, Midrigan, and Xu (2015), Feenstra and Weinstein (2017), Arkolakis et al. (2019), and Edmond et al. (2018) consider settings where firms adjust their markup in response to import competition. Edmond et al. (2015) estimate non-trivial pro-competitive effects from trade liberalization under oligopolistic competition with CES preferences à la Atkeson and Burstein (2008). Arkolakis et al. (2019), meanwhile, find that under a wide class of variable elasticity demand systems, firm-selection on the extensive margins can neutralize the conditional pro-competitive effects.[6]

Our emphasis on international rent-shifting has less precedent in the literature. The trade policy literature has examined circumstances wherein trade agreements should explicitly remedy international rent- or profit-shifting externalities (Brander and Spencer (1985); Eaton and Grossman (1986); Mrázová (2011); Ossa (2012); Bagwell and Staiger (2012); Head and Spencer (2017)).[7] We complement this literature by establishing a duality between international rent-shifting and unbalanced tariffs, uncovering a possible limitation of how reciprocity is defined under existing trade agreements.

Our paper also relates to a vibrant literature measuring market power in international settings. De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) estimate markups by applying the cost-

[5] Relatedly, Nocco et al. (2014); Dhingra and Morrow (2019); Behrens et al. (2020); Mrázová et al. (2021) examine how the distribution of markups and the degree of misallocation depends on the underlying demand and supply structure.

[6] We use this result to extend our sufficient statistics formulas to settings with variable markups. It is also worth noting that our paper is also tangentially connected to the vibrant literature on the gains from trade in distorted economies—see Atkin and Donaldson (2021) for a review. More recently, Baqaee and Farhi (2019) provide non-parametric formulas for the welfare impacts of piecemeal tariff reduction, establishing a duality between trade liberalization and productivity growth.

[7] Relatedly, Ossa (2012) and Lashkaripour (2021), among others, calibrate trade models with profits to quantify the cost of non-cooperative tariffs.

based approach to a sample of Indian firms, paying particular attention to the challenges of production function estimation when dealing with multi-product firms. They find pro-competitive effects from output tariff liberalization but anti-competitive effects from input tariff liberalization in India. De Loecker and Eeckhout (2018) and Díez, Fan, and Villegas-Sánchez (2021) estimate markup by applying the cost-based approach to firm-level data spanning multiple countries and industries, analyzing how the distribution of markups has changed over time and space. Beyond basic markup estimation, Asker, Collard-Wexler, and De Loecker (2019) estimate misallocation in global oil extraction by contrasting factual cost curves to undistorted (counterfactual) supply curves. Caliendo, Parro, and Tsyvinski (2022) develop a novel sufficient statistics approach to infer the sum of markup and non-markup wedges from the global input-output matrix.

2 Suggestive Evidence for International Rent-Shifting

Throughout this paper, we use *"rents"* to denote net profits, which correspond to total firm profits from selling at a markup over marginal cost minus total investments in overhead costs.[8] *International rent-shifting* occurs when, upon opening to trade, some countries specialize in high-profit margin industries while others move in the opposite direction. As a result, the former group of countries receives net rents from the latter. We open our analysis by presenting suggestive evidence for international rent-shifting of this sort.

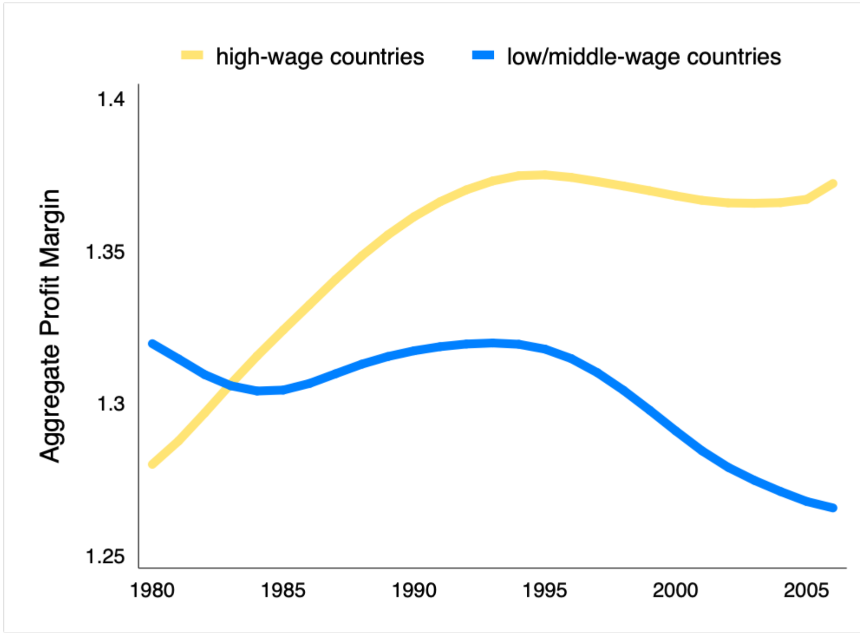
A first glance at industry-level production data unveils a divergence in net accounting profit margins (i.e., rents) between low- and high-wage economies. This pattern is displayed in Figure 1. It is based on the TRADEPROD database from CEPII, which covers 180 countries and 27 ISIC rev.2 industries. The aggregate accounting profit margin is calculated as sales divided by cost, aggregated over all 27 industries. From 1980 to 2005, aggregate profit margin in high-wage economies (i.e., countries in the top 20 percentile of the wage distribution) exhibit an upward trend, while aggregate profit margins in low- and middle-wage countries decline.

The pattern in Figure 1 may reflect developments other than trade-induced specialization across low- and high-profit margin industries. This pattern may be, for instance, driven by changes to firm-level markups in response to intensified competition in low-wage countries or cost reduction in high-wage economies. It is difficult to rule out these possibilities without a formal analysis like the one performed in Section 7 of this paper. But we can demonstrate that trade-induced specialization is a crucial part of the story in the US economy for which we possess rich firm-level financial accounts data for a long time horizon.

A casual examination of firm-level production data for the US economy reveals that trade openness coincided with significant resource reallocation from low- to high-profit margin industries. Figure 2 illustrates this pattern. The left panel indicates that the US economy has become twice more open to trade from 1960 to 2005. The right panel demonstrates that, at the same time, production

[8] This definition is akin to that adopted by Hsieh and Klenow (2009) and Aghion et al. (2019), among others.

Figure 1: International divergence in accounting profits by income group



activity in the US economy has become more concentrated in high-profit (or markup) industries.[9]

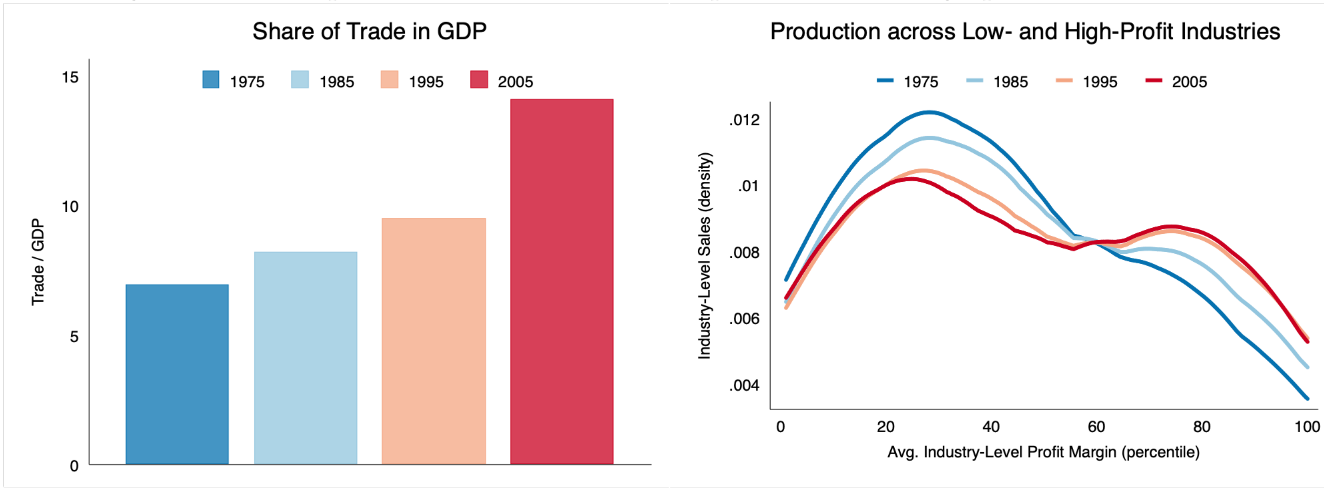
The above evidence must be viewed as merely suggestive. Other factors aside from trade openness may have contributed to the international divergence of profit margins or the change in US's pattern of production over time. But on the surface at least, trade openness seems to have internationally asymmetric effects on profit margins and the corresponding cost of markups. In what follows, we use theory and data to carefully measure the extent to which trade has multiplied or mitigated the cost of markups for different countries through international rent-shifting.

3 Baseline Theoretical Framework

Our baseline model is a multi-industry Krugman model with restricted entry. This minimal model is later extended to account for variable markups, firm-heterogeneity with selection into export markets, input-output linkages, and capital production inputs. The world economy consists of multiple countries and industries. Countries are indexed by of $i, j = 1, ..., N$. Industries are indexed by $g, k = 1, ..., K$. Industries admit different markup levels and trade elasticities. Each country i is populated by L_i individuals who supply one unit of labor inelasticity. Labor is the sole primary factor of production. Workers are unable to relocate internationally but are perfectly mobile across national industries. Each worker in country i is, accordingly, paid a country-wide wage, w_i.

[9] The data from COMPUSTAT consolidates annual sales and costs for each firm headquartered in the US. So, the gross accounting profit margins, we calculate, include profits booked in international tax havens.

Figure 2: Trade openness coincides with US's specialization in high-profit industries



Note: The left panel is constructed with data from the PENN WORLD TABLES v10. The right panel reports the distribution of US production activity across low- and high-profit industries. This figure is constructed with data from COMPUSTAT. Accounting profit margins are measured as gross sales to cost ratio and industries are defined according to the SIC classification. SIC industries are ordered from low- to high-profit margin (i.e., percentiles 0 to 100) based on their median average profit margin during the 1960-2005 period.

Preferences. The representative consumer in country i maximizes a three-tier utility function that aggregates over goods sources from various origins and industries:

$$U_i = \prod_k \mathcal{U}_{i,k} \left(Q_{1i,k}, ..., Q_{Ni,k} \right)^{e_{i,k}}.$$

The upper tier is a Cobb-Douglas aggregator, which ensures a constant share, $e_{i,k} > 0$, of national expenditure is spent on industry k goods—with $\sum_k e_{i,k} = 1$. The lower tier aggregates over national-level composites from various origins, with $Q_{ji,k}$ denoting the national-level composite that aggregates over firm-level varieties sold by origin j to destination i in industry k. In principle, we need not to impose parametric restrictions on the cross-national utility aggregator $\mathcal{U}_{i,k}(.)$—our sufficient statistics formulas for the welfare cost of markups are independent of the parameters underlying $\mathcal{U}_{i,k}(.)$. But in the interest of exposition, we present our model using the conventional CES parametrization,

$$\mathcal{U}_{i,k} = \left(\sum_{j=1}^{N} b_{ji,k}^{\frac{1}{\sigma_k}} Q_{ji,k}^{\frac{\sigma_k - 1}{\sigma_k}} \right)^{\frac{\sigma_k}{\sigma_k - 1}},$$

where $b_{ji,k}$ is a constant demand shifter and σ_k is the constant inter-national elasticity of substitution in industry k. As we explain later in Section 4, our sufficient statistics formulas require only mild conditions on the sub-national aggregator $Q_{ji,k}$. But to streamline presentation, we temporarily focus on the case where $Q_{ji,k}$ has a CES parameterization. Namely,

$$Q_{ji,k} = \left(\int_{\omega \in \Omega_{j,k}} \xi_{ji,k}(\omega)^{\frac{1}{\gamma_k}} q_{ji,k}(\omega)^{\frac{\gamma_k - 1}{\gamma_k}} d\omega \right)^{\frac{\gamma_k}{\gamma_k - 1}},$$

where ω indexes firm-level varieties and $\Omega_{j,k}$ denotes the fixed set of firms operating from *origin j–industry k*. In this formulation, the constant demand shifter $\xi_{ji,k}$ accounts for firm-level quality and γ_k is the constant elasticity of substitution in industry k, which regulates firm-level market power in industry k. Utility maximization subject to the budget constraint,

$$\sum_{k=1}^{K} \sum_{j=1}^{N} \left[\int_{\omega \in \Omega_{j,k}} p_{ji,k}(\omega) q_{ji,k}(\omega) \right] \leq Y_i,$$

yields a familiar CES demand for firm-level and national-level varieties,

$$q_{ji,k}(\omega) = \xi_{ji,k}(\omega) \left(\frac{p_{ji,k}(\omega)}{P_{ji,k}} \right)^{-\gamma_k} Q_{ji,k}; \qquad Q_{ji,k} = b_{ji,k} \left(\frac{P_{ji,k}}{P_{i,k}} \right)^{-\sigma_k} \frac{Y_i e_{i,k}}{P_{i,k}}; \tag{1}$$

where $P_{ji,k} = \left(\sum_{\omega \in \Omega_{j,k}} \xi_{ji,k}(\omega) p_{ji,k}(\omega)^{1-\gamma_k} \right)^{\frac{1}{1-\gamma_k}}$ and $P_{i,k} = \left(\sum_{j=1}^{N} b_{ji,k} P_{ji,k}^{1-\sigma_k} \right)^{\frac{1}{1-\sigma_k}}$ are respectively the origin–industry specific and industry-level CES price indexes.

Production and Supply. Labor is the sole factor of production. Industry k in origin i is represented by a constant mass $\overline{M}_{i,k} = \mid \Omega_{i,k} \mid$ of ex-ante identical firms who compete under monopolistic competition. Firms from the same origin and industry can have heterogeneous productivity levels. More specifically, every firm $\omega \in \Omega_{i,k}$ independently draws a productivity $\varphi_{i,k}(\omega)$ from a possibly non-degenerate distribution $G_{i,k}(\varphi)$. The marginal cost of serving destination country j by firm $\omega \in \Omega_{i,k}$ is equal to $\tau_{ij,k} w_i / \varphi_{i,k}(\omega)$, where $\tau_{ij,k} \geq 1$ denotes the iceberg trade cost and w_i denotes the wage rate in origin i. The optimal price, $p_{ji,k}(\omega) = \mu_k \tau_{ij,k} w / \varphi_{i,k}(\omega)$, charged by firm ω features a constant multiplicative markup, μ_k, which is common to all firms in industry k and depends on the lower-tier CES elasticity, γ_k:

$$\mu_k \equiv \frac{\gamma_k}{\gamma_k - 1} \sim \text{optimal markup in indsutry } k.$$

Accordingly, the CES price index of the composite good sold by origin i–industry k to destination j is

$$P_{ij,k} = \mu_k \tau_{ij,k} w_i / \overline{\varphi}_{i,k}$$

where $\overline{\varphi}_{i,k} \equiv \overline{M}_{i,k}^{\frac{1}{\gamma_k - 1}} \left[\int_{\varphi} \varphi^{\gamma_k - 1} \xi(\varphi) dG_{i,k}(\varphi) \right]^{\frac{1}{\gamma_k - 1}}$ is the variety-adjusted average productivity of firm-level varieties from origin $i-$industry k . Since entry is restricted, firms in origin i–industry k collect a total net profit equal to

$$\Pi_{i,k} = \sum_{j=1}^{N} \left[\left(1 - \frac{1}{\mu_k} \right) P_{ij,k} Q_{ij,k} \right].$$

National-level rents are, accordingly, equal to the sum of industry-level net profits: $\Pi_i = \sum_{k=1}^{K} \Pi_{i,k}$. We assume that Π_i is rebated to workers in origin i as a lump-sum transfer that complements their wage income.

General Equilibrium and Welfare. Combining the expressions for optimal demand and prices, equilibrium can be defined as follows. For a given vector of parameters and exogenous variables, $\left\{\mu_k, \sigma_k, \overline{L}_i, e_{i,k}, \overline{\varphi}_{i,k}, \tau_{ji,k}\right\}_{j,i,k}$, equilibrium is a vector of national-level wages and rents, $\{w_i, \Pi_i\}_i$, that satisfy the labor market clearing (LMC) and balanced budget (BB) conditions in each country i:

$$w_i \overline{L}_i = \sum_{k=1}^{K} \sum_{j=1}^{N} \left[\frac{\frac{1}{\mu_k} \left(\tau_{ij,k} w_i / \overline{\varphi}_{ij,k}\right)^{1-\sigma_k}}{\sum_{n=1}^{N} \left(\tau_{ni,k} w_n / \overline{\varphi}_{ni,k}\right)^{1-\sigma_k}} e_{j,k} (w_j \overline{L}_j + \Pi_j) \right] \tag{LMC}$$

$$\Pi_i = \sum_{k=1}^{K} \sum_{j=1}^{N} \left[\frac{\left(1 - \frac{1}{\mu_k}\right) \left(\tau_{ij,k} w_i / \overline{\varphi}_{ij,k}\right)^{1-\sigma_k}}{\sum_{n=1}^{N} \left(\tau_{nj,k} w_n / \overline{\varphi}_{nj,k}\right)^{1-\sigma_k}} e_{j,k} (w_j \overline{L}_j + \Pi_j) \right] \tag{BB}$$

In the above equations, $w_j \overline{L}_j + \Pi_j = Y_j$ denotes the total income of the representative consumer in country j, which is the sum of wage payments and rents. Under the Cobb-Douglas-CES demand structure, country i's welfare (i.e., real income) can be expressed as[10]

$$W_i = \frac{\Pi_i + w_i \overline{L}_i}{P_i}, \quad \text{where} \quad P_i = \prod_{k=1}^{K} \left(\sum_{j=1}^{N} \left(\mu_k \tau_{ij,k} w_i / \overline{\varphi}_{ij,k} \right)^{1-\sigma_k} \right)^{\frac{e_{i,k}}{1-\sigma_k}}$$

Notation: *Mean*, *Covariance*, and *Coefficient of Variation*

To simplify the presentation of our sufficient statistics formulas, we use the mean, variance, and covariance operators. Below, we formally define each operator.

Weighted Mean. Let X_k denote a generic variable or parameter that varies by industry k. The operator,

$$\mathbb{E}_\omega [X] \equiv \sum_k [\omega_k X_k],$$

denotes the weighted mean of X, where ω_k is the weight assigned to industry k with the weights adding up to one: $\sum_k \omega_k = 1$. To give an example, $\mathbb{E}_{e_i} \left[\frac{1}{\mu}\right] = \sum_k \left[e_{i,k} \frac{1}{\mu_k}\right]$ denotes the expenditure-weighted average of the (inverse) industry-level markups.

Coefficient of Variation. Let $\text{Var}_\omega (X) \equiv \sum \left[\omega_k (X_k - \mathbb{E}_\omega [X])^2\right]$ denote the weighted variance of a generic variable X. The coefficient of variation of X is defined as

$$\text{CV}_\omega (X) = \frac{\sqrt{\text{Var}_\omega (X)}}{\mathbb{E}_\omega [X]}.$$

[10] Recall that $e_{i,k} > 0 \quad \forall i, k$ is the constant share of expenditure on industry , with $\sum_k e_{i,k} = 1$.

For example, consider markups, μ_k, that vary by industry. $\text{CV}_{e_i}\left(\frac{1}{\mu}\right)$ denotes the coefficient of variation of inverse markups in country i, weighted by the country-specific expenditure shares.

Covariance. Consider two generic variables or parameters, Z_k, and X_k. The operator,

$$\text{Cov}_\omega\left(Z, X\right) \equiv \sum_k \left[\omega_k\left(X_k - \mathbb{E}_\omega\left[X\right]\right)\left(Z_k - \mathbb{E}_\omega\left[Z\right]\right)\right],$$

denotes the covariance between Z, and X, across industries. We should note that our operators can apply to variables that have industry and country subscripts. For instance, let $X_{i,k}$ be a generic variable that is country i and industry k specific. $\mathbb{E}_\omega\left[X_i\right]$ and $\text{CV}_\omega\left(X\right)$, in that case, denote the mean and coefficient of variation of X across industries within country i.

4 Inferring the Welfare Cost of Markups from Observables

In this section, we provide sufficient statistics formulas for the welfare cost of markups in open economies. To set the stage for our main result, we first characterize the cost of markups in a closed economy insulated from foreign trade.

4.1 The Welfare Cost of Markups in a Closed Economy

We denote the closed economy equilibrium with superscript A, which indicates that variables are being evaluated under *autarky*. To determine the cost of markups we must identify the efficient allocation, which we distinguish by superscript $\star$. It is straightforward to verify that the efficient allocation is implemented when prices exhibit no markup or a common markup that is constant *and* uniform across all goods (see Appendix A for details). For simplicity and without loss of generality, we focus on the no-markup case where efficient prices equal the marginal cost:[11]

$$P^\star_{ii,k} = \tau_{ii,k} w_i / \overline{\varphi}_{i,k}$$

Notice that the price of domestically-produced and consumed varieties,$\{P_{ii,k}\}_{i,k}$, is the only relevant price variable under autarky, with the value assigned to international price variable, $P_{ji,k}$, where $j \neq i$, being irrelevant. Noting our focus on autarky, let W_i^A denote welfare under the market equilibrium and $W_i^{\star A}$ denote welfare under the Pareto-efficient (i.e., marginal-cost-pricing) equilibrium. As is standard, we define the aggregate cost of markups in country i as the welfare distance between the market equilibrium and the Pareto-efficient equilibrium:

$$\mathcal{D}_i^A \equiv \log W_i^{\star A} - \log W_i^A > 0, \qquad \text{(welfare cost of markup distortions)}$$

[11] To be more specific, an price schedule that satisfies $P^*_{ii,k} = \bar{\mu}\tau_{ii,k} w_i / \overline{\varphi}_{i,k}$ will also yield the efficient outcome. Without loss of generality, we normalize $\bar{\mu} = 1$ hereafter.

Taking note of above definition, we present our first intermediate result, which characterizes the cost markups in a closed economy as a function of markups and observable shares.

Lemma 1. *The welfare cost of markups to economy i under autarky, can be inferred from markups and observable expenditure shares,* $\mathbf{X} = \{e_{i,k}, \mu_k\}_k$, *as*

$$\mathcal{D}_i^A(\mathbf{X}) = \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right].$$

A formal proof for the above lemma is presented in appendix. To gain intuition about Lemma 1 it is helpful to link $\mathcal{D}_i^A(\mathbf{X})$ to inter-sectoral markup heterogeneity. For this purpose, we can use Taylor's Theorem to derive the following second-order approximation for the aggregate cost of markups in a closed economy:

$$\mathcal{D}_i^A(\mathbf{X}) \approx \frac{1}{2}\left[\text{CV}_{e_i}\left(\frac{1}{\mu}\right)\right]^2 \tag{2}$$

Stated verbally, the cost of markup distortions is approximated by the coefficient of variation of inverse markups, underscoring the critical and well-understood role of markup dispersion. There is, accordingly, no welfare cost associated with markups when they are positive but uniform across industries. Finally, Lemma 1 invokes the Cobb-Douglas assumption that industry-level expenditure shares remain the same if we move from the status quo to autarky, i.e., $e_{i,k} = e_{i,k}^A$. The constancy of industry-level expenditure shares no longer holds in richer environments with CES preferences across industries or input-output linkages—a point we come back to in Section 6.[12]

Restricted *vs.* Free Entry. Before moving forward, a discussion about the role of restricted entry is in order. Our goal in this paper is to quantify international rent-shifting—hence, our focus on restricted entry. We should emphasize, nonetheless, that restricted entry is not what makes markups costly. If anything, restricted entry deflates the welfare cost of markups. Under *free entry*, the welfare cost of markups for closed economy *i* is

$$\mathcal{D}_i^A = \mathbb{E}_{e_i}[\mu \log \mu] - \mathbb{E}_{e_i}[\mu] \log \mathbb{E}_e[\mu] \approx \frac{1}{2}\left[\text{CV}_{e_i}\left(\frac{1}{\mu}\right)\right]^2 + \text{Cov}_{e_i}(\mu, \log \mu),$$

which is strictly greater than the cost under restricted entry—see Appendix B for derivation details. To provide some intuition, under free entry, markups distort entry decisions across industries. They lead to inefficiently low entry in high-markup industries and excessive entry in low-markup industries. The welfare consequences of these distortions are amplified by the fact that high-markups industries exhibit a higher degree of love-for-variety.[13] The term, $\text{Cov}_{e_i}(\mu, \log \mu)$, reflects these amplification effects, as it corresponds to the covariance between the industry-level markup and degree

[12] Note that autarky expenditure shares exactly coincide with autarky revenue shares. As such, the averages specified by Lemma 1 can be alternatively presented as revenue-weighted (rather than expenditure-weighted) averages.

[13] These amplification effects are not an artifact of CES preferences, and hold in more general environments (Baqaee and Farhi (2020)). Also, see Costinot and Rodríguez-Clare (2014) and Kucheryavyy et al. (2016) for a characterization of the gains from trade in multi-sector trade models with free entry.

of love-for-variety. This distinction aside, the free entry case is qualitatively similar to the restricted entry case. More specifically, if trade openness magnifies the cost of markups under restricted entry, it must have the same effect under free entry albeit with a different magnitude.

4.2 The Welfare Cost of Markups in an Open Economy

We now extend Lemma 1 to derive a sufficient statistics formula for the welfare cost of markups in open economies. To set the stage for our derivation, we define two additional statistics that gain relevance in an open economy—namely, industry-level revenue shares and revealed comparative advantage. We use $r_{i,k}$ to denote industry k's gross revenue share in economy i:

$$r_{i,k} \equiv \frac{\sum_j P_{ij,k} Q_{ij,k}}{\sum_{k'} \sum_j P_{ij,k'} Q_{ij,k'}} \sim \text{revenue share}$$

Without trade, the revenue share on industry k exactly coincides with its expenditure share in each country. Opening to trade, however, leads to a decoupling between revenue and expenditure shares—the pattern of which reflects a countries patten of specialization. If $r_{i,k} > e_{i,k}$, one can deduce that country i is a net exporter of industry k goods to the rest of the world and vice versa. Accordingly, we track a country's pattern of specialization using the following index for revealed of comparative advantage (RCA):

$$\text{RCA}_{i,k} \equiv \frac{r_{i,k}}{e_{i,k}} \sim \text{revealed comprative advantage}$$

The above index is one of the multiple RCA indexes surveyed by Ballance et al. (1987), and has been previously adopted by Gustavsson et al. (1999) among others. Aside from being compatible with our theoretical framework, it has a basic advantages over the widely-used Balassa index. It uses information on domestic output and expenditure levels in addition to export values. Under autarky, $\text{RCA}^A_{i,k} = 1$ for all k.[14] Under trade, $\text{RCA}_{i,k} > 1$ indicates that country i has a revealed comparative advantage in industry k, while $\text{RCA}_{i,k} < 1$ implies a revealed comparative *disadvantage*.

Dissecting the Consequences of Markup Correction. Analogous to the closed economy case, the *globally* efficient allocation (which is denoted by superscript $\star$) can be obtained if each good is priced at marginal cost, i.e., $P^\star_{ij,k} = \tau_{ij,k} w_i / \bar{\varphi}_{i,k}$ for all i, j, k. One may naturally speculate that the *national-level* cost of markups for an open economy is, accordingly, measured by distance to the *globally* efficient frontier, i.e., $\mathcal{D}_i \equiv \log W^\star_i - \log W_i$. But this measure can be flawed on paper, because $\mathcal{D}_i$ is possibly contaminated with terms-of-trade effects unrelated to markup distortions.

In light of this issue, we first show that restoring marginal cost pricing preserves each country's terms-of-trade, ensuring that $\mathcal{D}_i \equiv \log W^\star_i - \log W_i$ merely reflects the gains from correcting markup

[14] To elaborate, under autarky, country i's expenditure on industry k should equal its revenues from industry k sales. As such, the revenue and expenditure shares exactly coincide under autarky:$r^A_{i,k} = e^A_{i,k}$. Moreover, given that preferences are Cobb-Douglas, industry-level expenditure shares are invariant to trade (i.e., $e^A_{i,k} = e_{i,k}$), which implies that $r^A_{i,k} = e_{i,k}$.

distortions. To articulate this claim, let $W_i = \mathcal{W}_i(\boldsymbol{\mu};\mathbf{w})$ denote national welfare as an explicit function of industry-level markups, $\boldsymbol{\mu} = \{\mu_k\}_k$, and national-level wages, $\mathbf{w} = \{w_i\}_i$. Based on this functional representation, the welfare effects of eliminating markups can be decomposed as

$$\log W_i^\star - \log W_i = \int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \boldsymbol{\mu}} \cdot \mathrm{d}\log \boldsymbol{\mu} + \underbrace{\int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \mathbf{w}} \cdot \mathrm{d}\log \mathbf{w}}_{\text{terms-of-trade effects}}.$$

The fact that terms-of-trade effects channel exclusively through changes in $\mathbf{w}$ is a basic manifestation of Theorem 1 in Lashkaripour and Lugovskyy (2021). More specifically, a reduction in markups can alter the terms-of-trade directly and also indirectly through general equilibrium wage adjustments. The direct effect on the terms-of-trade occurs only when markups are reduced asymmetrically across countries. Here, markups are reduced uniformly (or symmetrically) for all countries, implying that any terms-of-trade effects channel exclusively through general equilibrium wage adjustments. Appendix C, meanwhile, establishes that $\mathbf{w}$ is invariant to restoring marginal cost pricing. That is,

$$\int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \mathbf{w}} \cdot \mathrm{d}\log \mathbf{w} = 0 \qquad \Longrightarrow \qquad \log W_i = \log W_i^\star - \int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \boldsymbol{\mu}} \cdot \mathrm{d}\log \boldsymbol{\mu}, \qquad (\forall i).$$

Hence, $\mathcal{D}_i \equiv \log W_i - \log W_i^\star$ can be interpreted as the change in welfare holding the terms-of-trade effects at their status quo level. Following the logic of Lemma 1 we can, then, derive the following relationship for the welfare cost of markups in open economy i:

$$\mathcal{D}_i \equiv \log W_i - \log W_i^\star = \log \mathbb{E}_{r_i}\left(\frac{1}{\mu}\right) - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right] \tag{3}$$

The derivation leading to the above formula is provided in Appendix E. Also, note the analogy between the above expression for $\mathcal{D}_i$ and the closed economy expression characterized under Lemma 1. To see the relationship, observe that $r_{i,k}^A = e_{i,k}$ when country i operates as a closed economy. Appealing to this observation, the expression for $\mathcal{D}_i$ (as specified by Equation 3) converges in the closed economy limit to the expression for $\mathcal{D}_i^A$ specified by Lemma 1: $\lim_{r_i \to r_i^A} \mathcal{D}_i = \mathcal{D}_i^A$.

5 How Trade Alters the Cost of Markups via Rent-Shifting

Subtracting the autarky cost of markups (as specified by Lemma 1) from the actual cost (as specified by Equation 3) delivers a sufficient statistics formula for how trade influences the cost of markups for different economies. The following proposition summarizes this result.

Proposition 1. *For an open economy i, the trade-induced change in the cost of markups, $\Delta\mathcal{D}_i = \mathcal{D}_i - \mathcal{D}_i^A$, can be inferred from markup values and observable shares, $\mathbf{X} = \{\mu_k, r_{i,k}, e_{i,k}\}_{i,k}$, as*

$$\Delta\mathcal{D}_i(\mathbf{X}) = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right].$$

Using Taylor's Theorem, we can relate $\Delta\mathcal{D}_i(\mathbf{X})$ to country i's pattern of specialization for small departures from autarky. Recall that country i's pattern of specialization is summarized by vector $\{\text{RCA}_{i,k}\}_k$ that represents its index of revealed comparative advantage across industries. Applying Taylor's Theorem to the formula specified by Proposition 1, we can derive the following approximation for the effect of trade on the welfare cost of markups in country i,

$$\Delta\mathcal{D}_i(\mathbf{X}) \approx \text{Cov}_{e_i}\left(\frac{1}{\mu}, \text{RCA}_i\right)\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{-1}. \tag{4}$$

Since a higher $\text{RCA}_{i,k}$ coincides with increased specialization in industry k, $\text{Cov}\left(\text{RCA}_i, \frac{1}{\mu}\right)$ describes country i's pattern of specialization across low- and high-markup industries. In particular,

$$\Delta\mathcal{D}_i(\mathbf{X}) \approx \begin{cases} \text{Cov}_{e_i}\left(\frac{1}{\mu}, \text{RCA}_i\right) < 0 & \text{if country } i \text{ specializes in high-markup industries} \\ \text{Cov}_{e_i}\left(\frac{1}{\mu}, \text{RCA}_i\right) > 0 & \text{if country } i \text{ specializes in low-markup industries} \end{cases}$$

The above approximation, therefore, indicates that trade raises the cost of markups for countries that have a comparative advantage in low-markup industries but lowers it for countries that have a comparative advantage in high-markup industries. As we explain shortly, these asymmetric reflect international *rent-shifting,* whereby trade-induced specialization shifts markup rents from some countries to others.

The Zero-Sum Nature of International Rent-Shifting. As noted in our introduction, the traditional literature on trade and distortions has typically emphasized the *pro-competitive* or *pro-allocative efficiency* effects of trade. The international rent-shifting emphasized by our framework is distinct from these effects in its zero-sum nature. One can immediately verify that the global weighted average of $\Delta\mathcal{D}_i$ is zero. Namely,

$$\sum_{i=1}^{N} y_i \Delta\mathcal{D}_i(\mathbf{X}) = 0,$$

where y_i denotes the share of country i's GDP from world GDP. The above expression indicates that when trade impacts markup distortions only through international rent-shifting, it prompts a transfer of welfare from one set of countries to another without delivering efficiency gains at the global level. As we explain shortly, this need not be the case in richer environments where trade yields *pro-competitive* or *pro-allocative efficiency* effects.

The Gains from Trade in Markup-Distorted Economies. With the aid of Proposition 1, we can decompose the welfare gains from trade into (1) the standard efficiency gains, and (2) the effect of trade on markup distortions. For this, we invoke the well-known result that for an efficient economy, $\log W_i^\star = \log W_i^{\star A} + \sum_k \left[\frac{e_{i,k}}{1-\sigma_k}\log\lambda_{ii,k}\right]$, where $W_i^\star$ and $W_i^{\star A}$ respectively denote country i's real income under the open and closed economy *efficient* equilibria. Define the gains from trade as $\text{GT}_i = \log\left(W_i/W_i^A\right)$. Noting that, by definition, $\log\left(W_i/W_i^A\right) = \log\left(W_i^\star/W_i^{\star A}\right) + \Delta\mathcal{D}_i$, we can use the expression for $\Delta\mathcal{D}_i$ from Proposition 1 to obtain the following expression for the gains from trade in country i:

$$\mathrm{GT}_i = \underbrace{\mathbb{E}_{e_i}\left[\frac{1}{1-\sigma}\log\lambda_{ii}\right]}_{\text{efficiency gains (ACR)}} - \underbrace{\left(\log\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]\right)}_{\Delta\mathcal{D}_i}. \tag{5}$$

To elaborate, the first term on the right-hand side corresponds to the traditional gains from trade in efficient economies, as outlined by Arkolakis et al. (2012). The second term measures how trade affects the welfare cost of markups in country i though international rent-shifting (Proposition 1).[15]

5.1 Unpacking the Cost of Markups in Open Economies

Using our previous results, we can decompose the cost markups in open economies into a *markup dispersion* cost and an *international rent-shifting* cost. Combining Equations 2 and 4, the welfare cost of markups for economy i is approximated as

$$\mathcal{D}_i \approx \underbrace{\frac{1}{2}\left[\mathrm{CV}_{e_i}\left(\frac{1}{\mu}\right)\right]^2}_{\text{markup dispersion}} + \underbrace{\mathrm{Cov}_{e_i}\left(\frac{1}{\mu}, \mathrm{RCA}_i\right)\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{-1}}_{\text{international rent-shifting}}, \tag{6}$$

The dispersion term is familiar from the closed economy literature. It represents the cost of markup distortions when the corresponding rents remain in country i. In that case, the welfare cost of markups is solely determined by the cross-industry dispersion in markups. To be clear, this is the only source of welfare cost in a closed economy. In our baseline framework, this term is not impacted by trade openness because both industry-level markups and expenditure shares are invariant to trade. As we argue shortly, the constancy of industry-level markups is compatible with variable markups at the firm level. The constancy of industry-level expenditure shares, moreover, is relaxed in Section 6 by replacing Cobb-Douglas preferences with CES preferences across industries or introducing input-output linkages.[16]

The international *rent-shifting* term has less precedents in the literature and attains relevance only when country i is an open economy. Markups, in a closed economy, distort consumer prices but also generate rents/profits that are rebated to consumers. In an open economy, the rents associated with markups are only partially rebated to domestic consumers. The remaining rents are collected by foreign firms and rebated to foreign households. As a result, markup distortions can become more or less costly depending on whether country i is a net receiver or a net payer of rents—which is captured

[15] Absent markup distortions, $\mathcal{D}_i = \mathcal{D}_i^A = 0$ and the second term on the right-hand side collapses to zero leaving us with the plain gains from trade formula popularized by Arkolakis et al. (2012).

[16] That $\mathrm{CV}_{e_i}\left(\frac{1}{\mu}\right)$ changes in response to trade openness under CES preferences is detailed in Section 6 . But let us offer brief explanation here: Under CES preference, the industry-level expenditure shares, $e_{i,k}$, are no longer constant but response to trade openness. The dispersion term under autarky will, accordingly, differ from its status quo level as it is weighted by a different vector of expenditure shares. A basic implication of this is that trade will not only impact misallocation through rent-shifting but also through changes in markup dispersion.

by the *rent-shifting* term in Equation 6.

5.2 Applicability of Proposition 1 to Settings with *Variable Markups*

While our baseline model assumes a constant and uniform markup for all firms in industry k, Proposition 1 is more general. We show that with slight reinterpretation, Proposition 1 continues to hold in richer environments with variable and heterogeneous markups within industries. The intuition is that, under some reasonable assumptions, trade reallocates resources across (low- and high-markup) industries but not across (low- and high-markup) firms within industries. We demonstrate this point by adopting the *homothetic* class of preferences analyzed by Arkolakis et al. (2019). The demand for a firm-level variety ω (corresponding to origin j–destination i–industry k) under this more flexible demand system is given by

$$q_{ji,k}(\omega) = \frac{Q_k\left(p_{ji,k}(\omega)/\mathcal{P}_k\right)}{\sum_n \int_{\omega' \in \Omega_{n,k}} p_{ni,k}(\omega') Q_k\left(p_{ni,k}(\omega')/\mathcal{P}_k\right) d\omega'} e_{i,k} Y_{i,k},$$

where $Q_k(.)$ is a strictly decreasing function; while the price index, $\mathcal{P}_k \equiv \mathcal{P}_k(\mathbf{p}_{i,k})$, implicitly solves $\sum_n \left[\int_{\omega \in \Omega_{n,k}} H_k(p_{ni,k}(\omega)/\mathcal{P}_k) d\omega\right] = 1$, where $H(.)$ is strictly increasing and strictly concave. Following Arkolakis et al. (2019), assume that $\mathcal{P}_k$ is also the choke price, wherein $Q_k\left(p/\mathcal{P}_k\right) = 0$ if $p > \mathcal{P}_k$. Firms draw their productivity, φ, independently from a Pareto distribution $G_i(\varphi) = 1 - (b_i/\varphi)^\theta$. After the realization of productivity draw, firm ω operates with marginal cost $c_{ij,k}(\omega) = \tau_{ij,k} w_i/\varphi(\omega)$, when selling from origin i to destination j.

Our goal is to show that the firm-level markup distribution is independent of market i and invariant to trade. To this end, it is useful to specify firms in each market by their marginal cost relative to the industry-level choke price, $\nu_k \equiv c/\mathcal{P}_k$. Firm's profit maximization implies the following formula for the optimal markup as a function of ν in each market:

$$\mu \equiv m_k(\nu) = \frac{\varepsilon_k(m_k/\nu_k)}{\varepsilon_k(m_k/\nu_k) - 1},$$

where $\varepsilon_k(x) \equiv \mid Q_k'(x) \mid$. It is straightforward to check that $m_k(.)$ is a strictly decreasing function provided that Marshall's Second Law of Demand is satisfied (i.e., $\varepsilon_k'(x) < 0$. Since $m_k(.)$ is monotone and admits an inverse function, the markup distribution in economy i can be described by the following formula which is common for all markets and independent of the trade cost vector, $\boldsymbol{\tau} \equiv \left\{\tau_{ij,k}\right\}$ (see Appendix C.4):

$$\mathcal{M}_k(\mu) = 1 - \left(m_k^{-1}(\mu)\right)^{\theta_k}.$$

The markup distribution, moreover, has spread $\mu \in [1, \infty)$ in each market irrespective of $\boldsymbol{\tau}$. As a result, consumers in each market pay a common sales-weighted average markup on industry k

goods,

$$\overline{\mu}_k = \left[\int_{\mu=1}^{\infty} \mu \frac{\frac{\mu}{m_k^{-1}(\mu)} Q_k(\mu/m_k^{-1}(\mu))}{\int_{\mu'=1}^{\infty} \frac{\mu'}{m_k^{-1}(\mu')} Q_k(\mu'/m_k^{-1}(\mu')) d\mathcal{M}_k(\mu')} d\mathcal{M}_k(\mu)\right],$$

which is invariant to international trade costs, $\boldsymbol{\tau}$.[17] These equations indicate that trade openness (which is measured by $\boldsymbol{\tau}$) has no effect on the allocation of resources across low- and high-markup firms *within* industries. Trade influences the cost of markups, $\mathcal{D}_i$, entirely through resource reallocation across low- and high-$\overline{\mu}_k$ industries, where $\overline{\mu}_k$ is the average markup in industry k.

Based on these observations, we can draw two immediate conclusions. First, for a given vector of average industry-level markups and expenditure shares, the model with variable markups predicts a greater misallocation cost associated with markups. Intuitively, in the presence of variable markups, misallocation derives from both within-industry and cross-industry markup dispersions. Second, despite this difference, the model with variable markups predicts the same impact from trade on the cost of markups—when calibrated to the same vector of average industry-level markups and expenditure/revenue shares. More specifically, under variable markups,

$$\Delta\mathcal{D}_i(\mathbf{X}) = \log \mathbb{E}_{r_i}\left[\frac{1}{\overline{\mu}}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\overline{\mu}}\right],$$

which is akin to the formula presented under Proposition 1 with one subtle difference: The above equation is based on $\overline{\mu}_k$, which is the sales-weighted average markup in industry k, while our baseline formula is based on μ_k, which is the constant and uniform CES markup in industry k. This equivalence result is especially useful when taking our model to data, as we observe systematic heterogeneity in markups within industries.

5.3 Applicability of Proposition 1 to Settings with *Non-Markup Distortions*

In our baseline model, markups are the only source of misallocation. Proposition 1, however, continues to hold even when the economy is plagued with additional (non-markup) distortions, which reflect inefficient transaction costs, financial frictions and contracting frictions à la Liu (2019). To demonstrate this, consider an economy featuring additional wedges between price and marginal cost, denoting them with $\psi_k \geq 1$. With these additional wedges, the price of variety ij,k becomes,[18]

$$P_{ij,k} = \psi_k \times \mu_k \tau_{ij,k} w_i / \overline{\varphi}_{i,k}.$$

The distortionary wedges ψ_k are different from iceberg costs in that they inflate cost without raising labor input demand.[19] They differ from markup wedges in that they create quasi-rents, which are

[17] The Pareto assumption (with regards to the firm-level productivity distribution) ensures that international expenditure shares assume a constant elasticity formulation as in the baseline model (see Arkolakis et al. (2019)).

[18] The implicit assumption here is that non-markup wedges, ψ_k, are constant and industry-specific.

[19] Regarding welfare implication: an economy featuring iceberg costs is constrained-efficient, whereas an economy featuring non-markup wedges can be inefficient even if we treat wedges as technical constraints—see Liu (2019).

competed aways and leave the economy a deadweight loss. These wedges are, therefore, inefficient and do not *explicitly* contribute to real profits, which are still given by

$$\Pi_i = \sum_j \sum_k \left[\left(1 - \frac{1}{\mu_k}\right) P_{ij,k} Q_{ij,k} \right].$$

Without loss of generality, we can henceforth think of ψ_k as reflecting the degree of financial frictions. Appendix C.5 illustrates that financial frictions amplify the degree of sectoral misallocation. The autarky distance to the efficient frontier is, in particular, the sum of the welfare cost associated with markups and the cost associated with financial frictions:[20]

$$\mathcal{D}_i^A(\mathbf{X}) = \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right] + \underbrace{\mathbb{E}_{e_i}\left[\log \psi\right]}_{\text{non-markup distortions}}.$$

Intuitively, the cost of markups, $\log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right] \approx \frac{1}{2}\left[\mathrm{CV}_{e_i}\left(\frac{1}{\mu}\right)\right]^2$ is tied to cross-industry dispersion because markups generate rents; while the cost of financial frictions, $\mathbb{E}_{e_i}[\log \psi]$, is tied to the average wedges size because they act as a deadweight loss. Despite the economy being plagued with greater misallocation, the effect of trade on misallocation channels entirely through its effect on the cost of markups. That is, the impact of trade on the welfare cost markups is still described by Proposition 1, namely,

$$\Delta \mathcal{D}_i(\mathbf{X}) = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right],$$

with the welfare gains from trade still given by Equation 5. Intuitively, even though financial frictions amplify sectoral misallocation, the amount of misallocation they create is invariant to trade. This result, however, comes with some limitations, as it is sensitive to the assumption that preferences across industries are Cobb-Douglas and financial wedges are industry rather than country-specific.

6 Measuring International Rent-Shifting in Richer Environments

In this section we move beyond our baseline model and derive sufficient statics formulas for how trade affects the cost of markups in more general environments. First, we consider an extension where preference are CES rather than Cobb-Douglas across industries. Second, we introduce input-output linkages into our baseline model à la Caliendo and Parro (2015). Third, we account for firm-selection into export markets à la Melitz (2003). Lastly, we provide a brief description of extended formulas that accommodate origin-specific markups and elastic labor supply.

[20] To be clear, $\mathcal{D}_i^A$ denotes distance to the efficient equilibrium in which (a) rents do not leave the economy and (b) allocative is restored subject to the preservation of rents. To elaborate, welfare in in an economy with non-markup distortions, ψ, can be specified as $W_i = \frac{w_i L_i + \Pi_i + \tilde{\Pi}_i}{P_i} - \delta_i \frac{\tilde{\Pi}_i}{P_i}$, where $\tilde{\Pi}_i$ denotes the rents associated with ψ and δ_i regulates whether these rents leave the economy as a dead weight loss ($\delta_i = 1$) or not ($\delta_i = 0$). One can show that the market equilibrium is constrained-efficient subject to $\delta_i = 1$. But policy interventions (say in the form of financial reforms) can prevent the dissipation of rents, lowering δ_i to zero. Correspondingly, $\mathcal{D}_i^A$ represents the welfare gains from preventing rent dissipation *and* restoring marginal cost pricing. See Appendix C.5 for more details.

6.1 CES Preferences Across Industries

Suppose preferences across industries have a CES rather than Cobb-Douglas parameterization. Namely,

$$U_i = \left[\sum_k b_{i,k}^{\frac{1}{\eta}} Q_{i,k}^{\frac{\eta-1}{\eta}}\right]^{\frac{1-\eta}{\eta}}, \qquad \text{where} \qquad Q_{i,k} = \left(\sum_j b_{ji,k}^{\frac{1}{\sigma_k}} Q_{ji,k}^{\frac{\sigma_k-1}{\sigma_k}}\right)^{\frac{\sigma_k}{\sigma_k-1}}.$$

Under this formulation, $\eta \geq 1$ denotes the elasticity of substitution across industries, with the special case $\eta = 1$ coinciding with our baseline Cobb-Douglas model. In the CES model, industry-level expenditure shares are endogenous and respond to trade openness or markup correction. Considering this, $e_{i,k}$ throughout this section denotes the *endogenous* expenditure share on industry k rather than a constant Cobb-Douglas weight. Despite this added layer of richness, we can still infer the autarky cost markup for economy i from observable shares, markups, and substitution elasticities. The following lemma presents this result with a formal proof provided in Appendix D.

Lemma 2. *Suppose preferences across industries are CES with substitution elasticity, η. The welfare cost of markups for country i (under autarky) can be inferred from industry-level markups, expenditure shares, and substitution elasticities, $\boldsymbol{X} = \{\lambda_{ii,k}, e_{i,k}, \mu_k, \sigma_k, \eta\}_{i,k}$, as*

$$\mathcal{D}_i^A(\boldsymbol{X}) = \log \mathbb{E}_{e_i^A}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\right] - \frac{1}{1-\eta}\log \mathbb{E}_{e_i^A}\left[\left(\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1}{1-\sigma}}\right)^{1-\eta}\right],$$

where $\tilde{\lambda}_{ii,k}^{\frac{1-\eta}{1-\sigma_k}} = \lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_k}} / \mathbb{E}_{e_i}\left[\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right]$ denotes the normalized the domestic expenditure share.

Evaluating the cost of markups under CES preferences requires three additional statistics, domestic expenditure shares, $\{\lambda_{ii,k}\}_{i,k}$, trade elasticities, $\sigma_k - 1$, and the cross-industry substitutability parameter, η. These additional statistics enable us to infer the change in industry-level expenditure shares after efficiency is restored in economy i. As in the baseline model, we can apply Taylor's Theorem to exact formula presented under Lemma 2 to derive the following approximation for the autarky cost of markups:

$$\mathcal{D}_i^A \approx \frac{\eta}{2}\left[\mathrm{CV}_{e_i}\left(\frac{1}{\mu}\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right)\right]^2.$$

Notice that the above formula reduces to our baseline formula in the Cobb-Douglas limit where $\eta = 1$. Capitalizing on the expression for $\mathcal{D}_i^A$, we can derive a revised formula for $\Delta\mathcal{D}_i = \mathcal{D}_i - \mathcal{D}_i^A$ that is compatible with CES preferences across industries. The next proposition outlines this result with a formal proof presented in Appendix D.

Proposition 2. *Suppose preferences across industries are CES with substitution elasticity, η. The trade-induced change in the cost of markups, $\Delta\mathcal{D}_i$, can be inferred from industry-level markups, expenditure shares,* *and substitution elasticities,* $\boldsymbol{X} = \{\lambda_{ii,k}, e_{i,k}, \mu_k, \sigma_k, \eta\}_{i,k}$, *as*

$$\Delta\mathcal{D}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\right] - \frac{1}{1-\eta}\log\left(\frac{\mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]}{\mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\right]}\right).$$

The CES-compatible expression for $\Delta\mathcal{D}_i$ exhibits an additional term that accounts for the impact of trade on *markup dispersion*. Specifically, as elaborated under Equation 6, the degree of misallocation in an open economy is composed of a *rent-shifting* term and a *markup dispersion* term. Under Cobb-Douglas preferences, the extent of *markup dispersion* is invariant to trade because industry-level expenditure shares are constant. Under CES preferences, however, industry-level expenditure shares react to trade, which translates into a change in the (expenditure-weighted) markup dispersion.

6.2 Global Input-Output Networks

Consider an extension of our baseline model where production in economy i employs labor and tradable intermediate inputs. The demand side of the economy has the same specification as the baseline model introduced in Section 3. The supply side is richer and modeled à la Caliendo and Parro (2015). That is, production in origin i–industry k combines labor with intermediate inputs. Let $\nu_{i,k}$ denote the constant share of labor in production (or the value added share) in industry k. Assuming constant-returns to scale, $1-\nu_{i,k}$ corresponds the overall share of intermediate inputs in production. Intermediate inputs used by industry k are sourced from various industries, with $(1-\nu_{i,k})\alpha_{i,gk}$ denoting the share of industry g inputs in industry k's production, with $\sum_g \alpha_{i,gk} = 1$. The composite bundle of inputs sourced from industry g (namely, $I_{i,gk}$) is an Armington aggregator of inputs from various origin countries. In particular,

$$I_{i,gk} = \left(I_{1i,gk}^{\frac{\sigma_k-1}{\sigma_k}} + ... + I_{Ni,gk}^{\frac{\sigma_k-1}{\sigma_k}}\right)^{\frac{\sigma_k}{\sigma_k-1}},$$

where $I_{ji,gk}$ denotes the quantity of industry g inputs sourced from origin j. The above production structure assumes that the CES input aggregator has the same parameterization as the CES consumption aggregator in industry k. Cost minimization subject to this production structure yields the following price for composite variety ij,k (origin i–destination j–industry k),

$$P_{ij,k} = \mu_k\left(\tau_{ij,k}/\overline{\varphi}_{i,k}\right) w_i^{\nu_{i,k}} \prod_{g=1}^{K} P_{i,g}^{(1-\nu_{i,k})\alpha_{i,gk}}, \tag{7}$$

where $P_{i,g} = \left(\sum_j P_{ji,g}^{1-\sigma_g}\right)^{\frac{1}{1-\sigma_g}}$ is a CES price index associated with inputs sourced worldwide from industry g. Note that $P_{i,g}$ also described the industry-level consumer price index in this setup, because all goods can be used for either input or final use with the same CES aggregator. Country i's total expenditure on industry k goods is, accordingly, the sum of consumption spending and input spending. Given that preferences for the final consumption good are Cobb-Douglas-CES, country i's total expenditure on industry k goods is given by

$$E_{i,k} = e_{i,k}\left(w_i L_i + \Pi_i\right) + \left(1 - \nu_{i,k}\right)\sum_g \left[\alpha_{i,kg} C_{i,g}\right], \tag{8}$$

where $C_{i,g}$ is the total input cost bill (which includes payments to labor and intermediate inputs) in origin i–industry g. By definition, origin i–industry g's total input cost is equal to total gross sales by net of the markup,

$$C_{i,g} = \sum_{j=1}^{N}\left[\frac{1}{\mu_g}\lambda_{ij,g}E_{j,g}\right], \qquad \text{where} \qquad \lambda_{ij,g} = \frac{P_{ij,g}^{1-\sigma_g}}{\sum_n P_{nj,g}^{1-\sigma_g}}. \tag{9}$$

General Equilibrium under IO Linkages. For a given vector of parameters and exogenous variables, $\left\{\mu_k, \sigma_k, L_i, e_{i,k}, \tau_{ij,k}, \overline{\varphi}_{i,k}, \nu_{i,k}, \alpha_{i,kg}\right\}_{i,k,g}$, equilibrium is a vector of wages, aggregate rents, industry-level price indexes, gross expenditure, and input cost levels, $\{w_i, \Pi_i, P_{i,k}, E_{i,k}, C_{i,k}\}_{i,k}$, that satisfy Equations 7–9 as well as the market clearing conditions in each market i,

$$w_i\overline{L}_i = \sum_k \nu_{i,k}C_{i,k}; \qquad \Pi_i = \sum_k (\mu_k - 1)C_{i,k}.$$

Gross Expenditure Shares. With IO linkages, the gross expenditure share on industry k typically differs from the net (or final good) expenditure share, $e_{i,k}$. Gross expenditure shares reflect both intermediate and final good expenditure. This difference plays a prominent role in our analysis, so we use

$$\tilde{e}_{i,k} \equiv \frac{E_{i,k}}{\sum_g E_{i,g}} \sim \text{gross expenditure share,}$$

to denote the gross expenditure share on industry k, where the gross expenditure level, $E_{i,k}$, is described by Equation 8. Generally speaking, $\tilde{e}_{i,k}$ will be greater than the net expenditure share, $e_{i,k}$, for upstream industries but lower for downstream industries. Moreover, unlike the net expenditure share, the gross expenditure is not *invariant* to trade. That is we cannot readily determine $\tilde{e}_{i,k}^A$ based on its observed value, $\tilde{e}_{i,k}$. We can, nonetheless, infer country i's *autarky* gross expenditure shares from constant net expenditure shares, $\mathbf{e}_i = [e_{i,k}]_k$, and the markup-adjusted input-output matrix, $\mathbf{\Phi}_i \equiv \left[\frac{1}{\mu_g}(1-\nu_{i,g})\alpha_{i,kg}\right]_{k,g}$. In particular, as shown in Appendix E,

$$\tilde{\mathbf{e}}_i^A = \left(\mathbf{I} - \mathbf{\Phi}_i\left(\mathbf{I} - \mathbf{e}_i \otimes \mathbf{1}\right)\right)^{-1}\mathbf{e}_i. \tag{10}$$

where $\mathbf{I}$ is an $K \times K$ identity matrix and $\mathbf{1}$ is a column vector of ones. Considering this point, we hereafter treat $\tilde{e}_i^A$ as an observable statistic.

Welfare Cost of Markups in a Closed Economy. As in the baseline model, the welfare cost of markups is measured as the the welfare distance between the factual equilibrium and the efficient marginal cost-pricing equilibrium, $\mathcal{D}_i = \log W_i^\star - \log W_i$. It is well known that IO linkages typically amplify the cost associated with a given vector of markup wedges. When industries are connected via IO linkages, the markup in industry k distorts production in all industries that use inputs from industry k. To account for these ripple effects in a closed economy, suppose markups are eliminated in every industry. Let $P_{i,k}^\star$ denote the price index after all markups are eliminated and $\hat{P}_{i,k} = P_{i,k}^\star / P_{i,k}$ denote the resulting change in the price index. Normalizing w_i to one by choice of numeraire, the change in industry k's price index is the product of industry k's own markup reduction $(1/\mu_k)$ and the compounded reduction in input markups. In particular,

$$\hat{P}_{i,k} = \frac{1}{\mu_k} \times \prod_g \hat{P}_{i,g}^{(1-\nu_{i,k})\alpha_{i,gk}}.$$

We can invert the above system to obtain $\hat{P}_{i,k} = \prod_g \left(\frac{1}{\mu_g}\right)^{a_{i,kg}}$, where $a_{i,gk}$ denotes the (g,k) entry of economy i's inverse Leontief matrix.[21] The change in the consumer price index, $\hat{P}_i = \prod_k P_{i,k}^{e_{i,k}}$, can thus be written as a compounded reduction in industry-level markups:[22]

$$\log \hat{P}_i = \sum_k \left[\beta_{i,k} \log \frac{1}{\mu_k}\right], \quad \text{where} \quad \beta_{i,k} \equiv \sum_g \left[e_{i,g} a_{i,gk}\right]. \tag{11}$$

Weight $\beta_{i,k}$ can be interpreted as industry k's compounded weight in the consumer price index (CPI)—it reflects how a reduction in industry k's markup translates to a reduction in CPI in the presence of ripple effects. Accordingly, for a strictly downstream industry, $\beta_{i,k}$ simply equals $e_{i,k}$ which is the Cobb-Douglas share of industry k in the consumption basket. Using the above observation and extrapolating the logic outlined in Section 4, we can produce an IO-adjusted sufficient statistics formula for the cost of markups in a closed economy.

Lemma 3. *Suppose production in each industry uses traded intermediate inputs. The welfare cost of markups for country i (under autarky) can be inferred from markups and observable shares,* $\mathbf{X} = \{\mu_k, \tilde{e}_{i,k}, \nu_{i,k}, \alpha_{i,gk}\}$*, as*

$$\mathcal{D}_i^A(\mathbf{X}) = \log\left(\frac{\mathbb{E}_{\tilde{e}_i^A}\left[\frac{\nu_i}{\mu}\right]}{1-\mathbb{E}_{\tilde{e}_i^A}\left[\frac{1-\nu_i}{\mu}\right]}\right) - \mathbb{E}_{\beta_i}\left[\log\frac{1}{\mu}\right],$$

where $\tilde{e}_{i,k}^A$ *and* $\beta_{i,k}$ *are respectively given by 10 and 11.*

Let us connect the above lemma to our baseline Lemma 1. Absent input output linkages, which corresponds to $\nu_{i,k} = 1$ and $\beta_{i,k} = e_{i,k}$ for all (i,k), the IO-adjusted formula for $\mathcal{D}_i$ collapses to the baseline formula presented under Lemma 1. Beyond this special case, the cost of markups depends crucially on the economy-wide input-output table, $\mathbf{A}_i \equiv \left[(1-\nu_{i,g})\alpha_{i,kg}\right]_{k,g}$, which is directly observ-

[21] More specifically, $\left[a_{i,gk}\right]_{g,k} = (\mathbf{I}-\mathbf{A}_i)^{-1}$, where $\mathbf{A}_i = \left[(1-\nu_{i,k})\alpha_{i,gk}\right]_{k,g}$ denotes country i's input-output matrix.

[22] It is easy to check that $\sum_k \beta_{i,k} = 1$, so $\beta_{i,k}$ satisfies the condition to serve as a weight in the mean operator, $\mathbb{E}[.]$.

able. Also worth highlighting is that the IO-adjusted formula for $\mathcal{D}_i$ uses information on both gross and net expenditure shares, $\tilde{e}_{i,k}$, and $e_{i,k}$.

Trade-Induced Change in Markup Distortions under IO Linkages. Next, we build on Lemma 2 to derive an IO-adjusted sufficient statistics formula for how trade affects the cost of markups. Derivation details are presented in Appendix E and follow the same logic as our baseline Propositions 1. The resulting formula is presented below.

Proposition 3. *Suppose production employs traded intermediate inputs. The trade-induced change in the cost of markups, $\Delta\mathcal{D}_i$, can be inferred from industry-level markup observable shares, $\mathbf{X} = \{\mu_k, \nu_{i,k}, r_{i,k}, \tilde{e}_{i,k}, \alpha_{i,gk}\}$, as*

$$\Delta\mathcal{D}_i(\mathbf{X}) = \log\left(\frac{\mathbb{E}_{r_i}\left[\frac{\nu_i}{\mu}\right]}{1-\mathbb{E}_{r_i}\left[\frac{1-\nu_i}{\mu}\right]}\right) - \log\left(\frac{\mathbb{E}_{\tilde{e}_i^A}\left[\frac{\nu_i}{\mu}\right]}{1-\mathbb{E}_{\tilde{e}_i^A}\left[\frac{1-\nu_i}{\mu}\right]}\right),$$

where $\tilde{e}_{i,k}^A$ is given, in closed form, by 10, as a function of the observable shares in set $\mathbf{X}$.

When interpreting, the above proposition observe that gross expenditure and revenue shares coincide under autarky: $\tilde{e}_i^A = r_i^A$. The above expression for $\Delta\mathcal{D}_i$, therefore, contrasts the allocation of resources across low- and high-markup industries under trade and autarky. In other words, the impact of trade on the cost of markups still channel primarily through rent-shifting effect, which are now adjusted for input-output linkages.

6.3 Firm Selection under Fixed Overhead Costs

Our baseline model abstracted from firm-selection into export market, by assuming away fixed marketing costs. Here, we extend our baseline formula for $\Delta\mathcal{D}_i$ to environments where firms have to incur a fixed marketing cost to serve international markets. More formally, firm $\omega \in \Omega_{i,k}$ from origin–industry (i, k) has to pay a fixed marketing cost, $w_j f_{j,k}$, to serve destination j. The fixed cost is, by assumption, paid in terms of labor in the destination market. As is standard, we assume that firms in industry k independently draw their productivity, φ, from a Pareto distribution that has an industry-specific shape parameter $\theta_k > \gamma_k - 1$. Under these assumptions, one can show that fixed marketing costs exhaust a constant fraction, ρ_k, of origin i's sales to destination . Namely,

$$M_{ij,k} w_j f_{j,k} = \rho_k P_{ij,k} Q_{ij,k}, \qquad \text{where} \qquad \rho_k \equiv 1 - \frac{1+\theta_k}{\mu_k\theta_k}.$$

To be clear, $M_{ij,k}$ denotes the mass of firms that can profitably serve destination j from origin i–industry k—which is a fraction of the exogenously-fixed number of firms, $M_{i,k}$. Appendix F uses the above equation to derive a *firm-selection-adjusted* sufficient statistics formula for the impact of trade on the cost of markups. Here, we outline two considerations that distinguish this setup from our baseline model. First, a fraction of the markup is now paid to cover the fixed marketing cost. Hence, misallocation drives from heterogeneity in excess markups—that is markups in excess of what is needed to pay the fixed cost. Second, since the fixed cost is paid in terms of labor in the destination market, host economies claim a fraction of the profit raised by foreign firms. The extent of rent-shifting, as a result, depends crucially on whether countries are net importers in industries with high or low fixed marketing costs. Our previously-defined index of revealed comparative advantage, $\text{RCA}_{i,k}$ tracks these patterns, revealing how much country i pays to and receives from the rest of the world in terms of fixed marketing costs. Accordingly, $\Delta\mathcal{D}_i$ depends not only on the pattern of specialization between low- and high-markup industries but also the pattern of specialization across low- and high-ρ industries. The following proposition formalizes this point.

Proposition 4. *The firm selection-adjusted effect of trade on sectoral misallocation in country i is given by*

$$\Delta\mathcal{D}_i\left(\mathbf{X}\right) \;=\; \log\left(\mathbb{E}_{r_i}\left[\frac{1}{\mu}+\rho\right]-\left(1+\frac{\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right]}{1-\mathbb{E}_{e_i}\left[\rho\right]}\right)\text{Cov}_{e_i}\left(\rho,\text{RCA}_i\right)\right)\;-\;\log\mathbb{E}_{e_i}\left[\frac{1}{\mu}+\rho\right]$$

as a function of industry-level markups, shape of the firm-size distribution, and observable revenue and expenditure shares, $\mathbf{X}=\{\mu_k,\theta_k,r_{i,k},e_{i,k}\}$.

To give some intuition $\frac{1}{\mu_k}+\rho_k$ can be interpreted as the inverse of the excess markup in industry k. As note earlier, the heterogeneity in excess markups determines misallocation in the present setup. Accordingly, when $\rho_k=0$, which is the limit with zero payments to fixed costs, the above formula reduces to the baseline formula specified under Proposition 1.

The adjustment, $\text{Cov}_{e_i}\left(\rho,\text{RCA}_i\right)$, accounts for fixed cost payment which transfer profits from one country to another. It balances fixed cost payments paid to foreign workers against the payments received from foreign firms. The following example may help understand the crucial role of fixed marketing costs. Proposition 3 indicates that—unlike the baseline model—trade can worsen misallocation in country i even if markups are uniform across industries. In particular, one can easily check from Proposition 3 that if $\mu_k=\mu\ \forall k$, then $\Delta\mathcal{D}_i\neq 0$ unless ρ_k is also θ_k uniform across industries. Intuitively, if the degree of firm heterogeneity, θ_k, varies across industries, the *excess* markup collected from industry-level sales may vary across industries despite the gross firm-level markup being uniform. As such trade can worsen or improve misallocation depending on whether resources are relocated to high- or low-ρ industries.

6.4 Other Extensions

Appendix G provides sufficient statistics formulas for the cost of markups, $\mathcal{D}_i$, and the impacts of international rent-shifting, $\Delta\mathcal{D}_i$, in richer environments with capital inputs, elastic labor supply, and origin- and destination-specific markups. Below, we briefly discuss these extensions, which, unlike the previous extension, are difficult to quantify given data limitations.

Capital as a Primary Input in Production. Appendix G.1 considers an economy where production employs both labor and capital as primary inputs. Labor and industry-specific capital are supplied inelastically. The Production function has a Cobb-Douglas parameterization with ς_k and $1-\varsigma_k$, respectively, denoting the share of labor and capital inputs in industry k's production. The impact of international rent-shifting, in this case, is given by

$$\Delta\mathcal{D}_i = \log \mathbb{E}_{\varsigma r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right],$$

which differs from our baseline formula in that the mean operators, $\mathbb{E}$, are weighted by the *labor-share-adjusted* expenditure and revenue shares. The logic is that labor is the only mobile factor across industries, so labor misallocation is the only driver of markup-driven inefficiency. Put differently, in the limit, $\varsigma_k \to 0$, the economy becomes constrained-efficient. Accordingly, the cost of markups, $\mathcal{D}_i$, and its response to trade via profit-shifting, $\Delta\mathcal{D}_i$, approaches zero when $\varsigma_k \to 0$.

Origin- and Destination-Specific Markups. Appendix G.2 provides formulas for $\mathcal{D}_i$ and $\Delta\mathcal{D}_i$ when average markups vary across both industries and origin and destination countries. In both cases, international rent-shifting is the sum of two terms: A term representing international rent-shifting if foreign markup levels were the same as those paid or charged by location i, and another representing the spread between the domestic and foreign markups paid or charged by location i. Mapping the formulas for $\mathcal{D}_i$ and $\Delta\mathcal{D}_i$, in these cases, is complicated by that fact that markups must be estimated independently across different origins and destinations. Appendix G.2 explains why this type of estimation is difficult to conduct at scale with existing techniques and datasets.[23]

Elastic labor Supply. Appendix G.3 examines economies with elastic labor supply. Evaluating the formulas for $\mathcal{D}_i$ and $\Delta\mathcal{D}_i$, in this case, requires information on the labor supply elasticity. The formulas, moreover, exhibit additional terms that account for distorted labor supply choices. The cost of markups in a closed economy, for instance, depends on the dispersion in consumer goods markups and their average value. The average markup regulates the extent to which labor supply choices are distorted by the markup on consumer goods. Accordingly, the impact of trade on the cost markups, $\Delta\mathcal{D}_i$, is composed of two terms: a term representing international rent-shifting and another representing the extent to which trade aggravates or mitigates the inefficient labor supply problem.

7 Quantitative Implementation

To evaluate formulas specified by Propositions 1-3, we need estimates for average industry-level markups as well as industry and country-level data on net expenditure, gross expenditure, gross revenue shares, value-added, and input-output shares. More specifically, we need the following set

[23] A possible way to circumvent this challenge is to infer firm-level markups from aggregate trade data (e.g., Ossa (2012); Firooz and Heins (2020)). For this, one has to either assume that the national-level and firm-level degrees of market power coincide or interpret the national-level trade data as representing a single firm per origin. The framework in Firooz and Heins (2020) is particularly elegant, here, as it allows for markups to diverge from trade elasticities and for the distribution of export markups to be origin- and destination-specific.

Table 1: List of countries/regions in the ICIO data

High Income				Low/Middle Income		
Australia	Austria	Belgium	Canada	China	Mexico	Turkey
Chile	Czech Republic	Denmark	Estonia	Argentina	Brazil	Brunei Darussalam
Finland	France	Germany	Greece	Bulgaria	Hungary	Cambodia
Iceland	Ireland	Israel	Italy	Colombia	Costa Rica	India
Japan	Korea	Latvia	Lithuania	Indonesia	Kazakhstan	Malaysia
Luxembourg	Netherlands	New Zealand	Norway	Morocco	Peru	Philippines
Poland	Portugal	Slovak Republic	Slovenia	Romania	Russian Federation	South Africa
Spain	Sweden	Switzerland	United Kingdom	Thailand	Tunisia	Vietnam
United States	Croatia	Cyprus	Hong Kong			
Malta	Saudi Arabia	Singapore	Taiwan			

Note: The classification by income group is based on the UNITED NATIONS COUNTRY CLASSIFICATION.

of sufficient statistics:

$$\mathbf{X} = \left\{\mu_k, e_{i,k}, \tilde{e}_{i,k}, r_{i,k}, \nu_{i,k}, \alpha_{i,kg}\right\}.$$

Of the sufficient statistics in **X**, markups need to be estimated, but the remaining elements are directly observable shares. We take these shares from the OECD Inter-Country Input-Output (ICIO) Tables. The ICIO includes a sample of 64 major countries and covers 33 sectors that span the entire economy. We use data from the entire time-span of the ICIO datasets which is 2005 to 2015. When reporting results for a specific year, we use 2010 as the benchmark.

Classifying Countries Based on Income Level. One of our goals is to determine if trade has asymmetric effects on high- and low-income countries. For this, we classify countries as *low/middle income* or *high-income* based on the UNITED NATIONS COUNTRY CLASSIFICATION. Table 1 lists all 64 countries in our sample alongside their income status. Note that our sample also includes an aggregate of the rest of the world, which we treat as unclassified.

7.1 Estimating Firm-Level Markups

We estimate markups using two different approaches: The cost-based approach (à la De Loecker and Warzynski (2012)) and the demand-based approach (à la Lashkaripour and Lugovskyy (2021)). While both approaches are well-understood, their macro-level implications have been rarely contrasted. In part, because the demand-based approach has proven difficult to implement across a wide range of countries and industries. To the best of our knowledge, this is the first attempts to compare the macro-level implication of markups estimated using the demand- and cost-based approaches. Next, we briefly describe the estimation strategy under each approach.

Cost-Based Markup Estimation. In this section, we follow De Loecker and Warzynski (2012) and estimate firm-level markups. we proceed in two stages. In the first stage, we use firm-level financial accounts data from COMPUSTAT to estimate the elasticity of output with respect to variable inputs.

For this, we adopt the Olley and Pakes (1996) approach with the Ackerberg et al. (2015) correction to estimate the industry and year-specific output elasticity, ε_{kt}, for each of the 36 ICIO industries in 2005-2015. This dataset covers only US firms. So, to apply the estimates to global data, we follow De Loecker and Eeckhout (2018) and assume that the output elasticity is the same across countries. We focus on ICIO industries because our international production and expenditure data is reported at this level of aggregation, so our goal is to estimate (average) markups at the level of ICIO industries.

In the second stage, we obtain firm-level data for variable input costs $C_t(\omega)$ and sales $R_t(\omega)$ from the WORLDSCOPE GLOBAL DATABASE, which covers 71,546 firms in 134 countries in 2005-2015. Some firms in this database operate in more than one industry. In these cases, we assign firm ω to each of the industries it operates in. Since we do not observe the breakdown of firm-level sales and costs by industry, we assume that sales and costs are equally spread across different industries.[24] Following De Loecker and Warzynski (2012), the markup charged by cost-minimizing firm ω in industry k is equal to $\mu_{kt}(\omega) = \varepsilon_{kt}[C_{kt}(\omega)/R_{kt}(\omega)]^{-1}$, where $C_{kt}(\omega)$ and $R_{kt}(\omega)$ are costs and sales attributable to industry k. We can thus calculate the sales-weighted average markup in industry k as

$$\overline{\mu}_{kt} = \sum_{\omega \in \Omega_k} \left[\frac{\varepsilon_{kt}}{C_t(\omega)/R_t(\omega)} \times \frac{R_{kt}(\omega)}{\sum_{\omega' \in \Omega_k} R_{kt}(\omega')} \right].$$

Recall from Section 5.2 that the sales-weighted average markup is the sufficient markup statistic for calculating $\Delta \mathcal{D}_i$ in the presence of markup heterogeneity within industries. Figure 11 reports our cost-based markups estimates for each of the 36 ICIO industries in our analysis.

Demand-Based Markup Estimation. To implement our demand-based markup estimation, we fit a structural demand equation to data on firm-level sales and prices. For this, we use the universe of import transactions for Colombia from 2007 to 2016. Our data covers 93,966 firms from 251 different countries. It reports transaction-level sales and quantities for individual firms exporting to Colombia at the Harmonized System 10-digit product level. Our goal is to estimate the CES demand function specified by Equation 1 with this data. Written in log-linear form, the demand function facing firm ω (associated with *origin j–destination i–industry k*) is

$$\ln q_{ji,kt}(\omega) = -\gamma_k \ln p_{ji,kt}(\omega) + D_{ji,kt} + \varepsilon_{ji,kt}(\omega), \tag{12}$$

where the year subscript, t, is added to account for the time dimension of our data. The term $D_{ji,kt} \equiv \ln \left(P_{ji,kt}^{\gamma_k - \sigma_k} P_{i,kt}^{\sigma_k} Q_{i,kt} \right)$ can be treated as an *origin j–destination i–industry k* fixed effect and $\varepsilon_{ji,kt}(\omega) \equiv \ln \xi_{ji,kt}(\omega)$ accounts for variety-specific demand shifters. To take full advantage of the richness in our data, we define industries in the most disaggregate way possible. That is, we let k denote a 10-digit product category and assume that γ_k is uniform across all 10-digit products pertaining to the same ICIO industry. This way, we can estimate an industry-level markup while controlling for an

[24] Specially, suppose firm ω is classified as operating in $n_t(\omega)$ different ICIO industries in year t, one of which is k. We assume that firm ω collects a revenue equal to $R_{kt}(\omega) = R_t(\omega)/n_t(\omega)$ from sales in industry k and pays $C_{kt}(\omega) = C_t(\omega)/n_t(\omega)$ in variable input costs.

exhaustive set of fixed effects.

A typical concern when estimating demand functions, like the one specified above, is that we observe aggregate sales and prices but individual consumers have possibly heterogeneous demands for variety ω. The standard way to deal with such heterogeneity—in small scale estimations—is to use the random coefficient demand estimation technique developed by Berry et al. (1995). This technique, however, is costly to implement and practically infeasible at the scale we plan to conduct our estimation. To overcome this challenge, we appeal to Theorem 1 in Salanié and Wolak (2019), to derive the following linear approximation of the random coefficient CES demand system:

$$\ln q_{ji,kt}(\omega) \approx -\gamma_{kt} \ln p_{ji,kt}(\omega) + \sigma_{\gamma t}^2 K_{ji,kt}(\omega) + D_{ji,kt} + \varepsilon_{ji,kt}(\omega). \tag{13}$$

The above equation differs from Equation 12 in an additional term $\sigma_{\gamma t}^2 K_{ji,kt}(\omega)$, where $\sigma_{\gamma t}$ is a standard deviation parameter that governs the heterogeneity in individual-level demand slopes; and $K_{ji,kt}(\omega) \equiv \left(\frac{1}{2}\ln p_{ji,kt}(\omega) - \ln \overline{p}_{i,kt}\right) \ln p_{ji,kt}(\omega)$, with $\ln \overline{p}_{i,kt}$ denoting the sales-weighted average of (log) price paid by destination i for industry k goods. To put it simply, the additional control, $K_{ji,kt}(\omega)$, accounts (to a first-order approximation) for individual-level heterogeneity in demand slopes. Absent individual-level demand heterogeneity (i.e., $\sigma_\gamma \to 0$), Equation 13 reduces to 12.

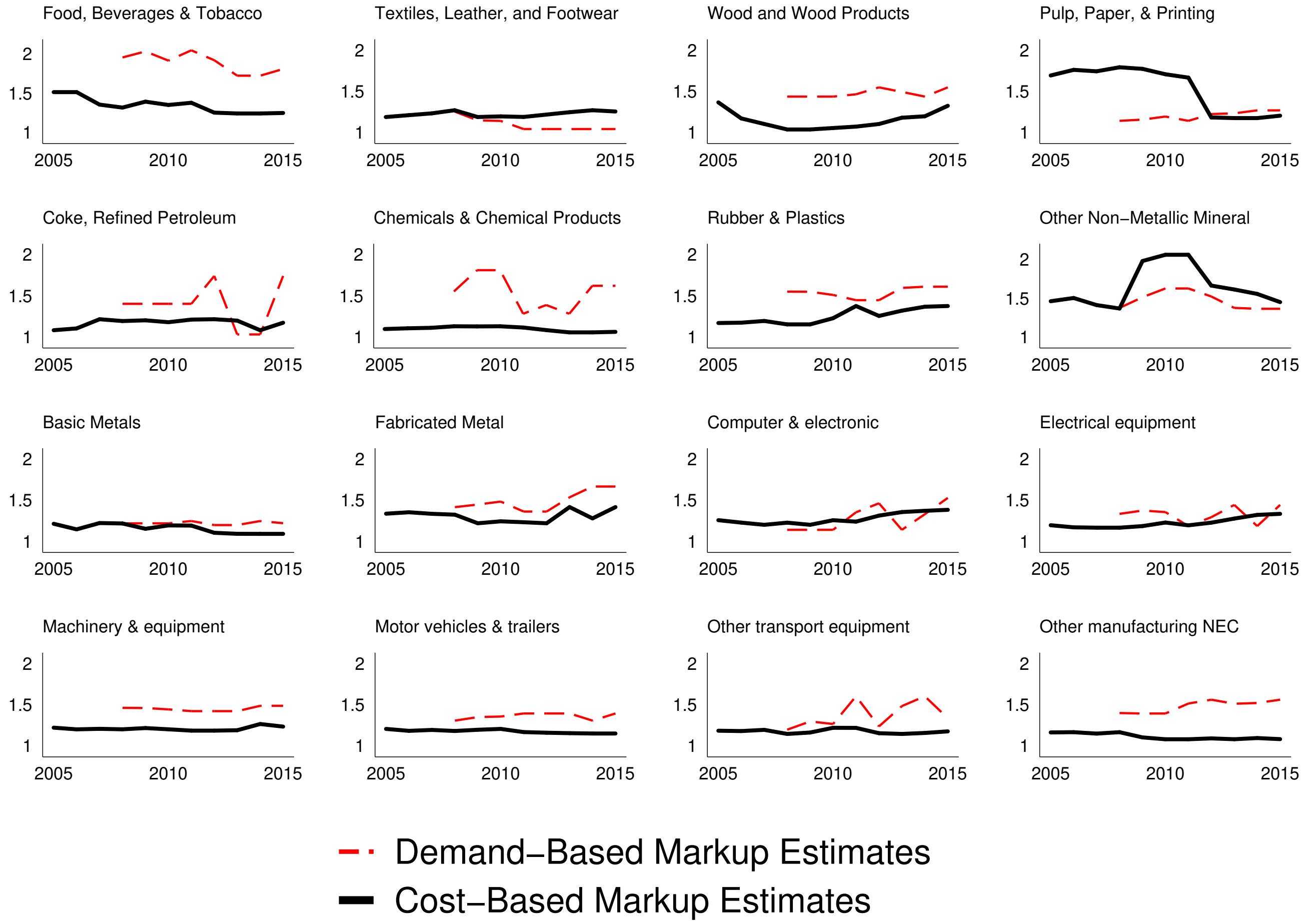
To estimate Equation 13 we must deal with the usual identification challenges facing demand estimation. For this, we use the identification strategy in Lashkaripour and Lugovskyy (2021). That is, we assume that the residual demand shifter, $\varepsilon_{ji,kt}(\omega) = \epsilon_{ji}(\omega) + \tilde{\varepsilon}_{\omega jikt}$, is composed of a time-invariant and firm-specific component plus an idiosyncratic component. We eliminate $\epsilon_{ji}(\omega)$ by taking first-differences from Equation 13. We then interact the variation in national-level exchange rates with prior monthly export shares to construct a shifter-share instrument (for price) that captures firm-level exposure to exchange rate shocks. We also include additional instruments, which are somewhat standard in the literature on import demand estimation (see Appendix H).

Figure 3 displays the estimated markups for manufacturing industries during 2005-2015. The cost-based markup estimates reported in Figure 3 are the sales-weighted average markups in a given year. Given that our transaction-level import data begins in 2007, our demand-based markup estimates (which are obtained from a first-difference estimator) cover years after 2008. Not surprisingly, one can detect the occasional discrepancy between the demand- and cost-based markup values. In many industries, however, demand- and cost-based markup estimates closely track one another over time. As we will see next, the demand- and cost-based markup estimates yield starkly similar macro-level predictions about the degree of sectoral misallocation and its response to trade openness.

7.2 The Welfare Cost of Markups

As a starting point, we report the welfare cost of markups which is the sum of the cost associate with markup dispersion and international rent-shifting. We do so by plugging our estimated markup values and data points into the formulas specified by Lemmas 1-3. Here, we report numbers based

Figure 3: The variation in estimated markups over time: *manufacturing industries*

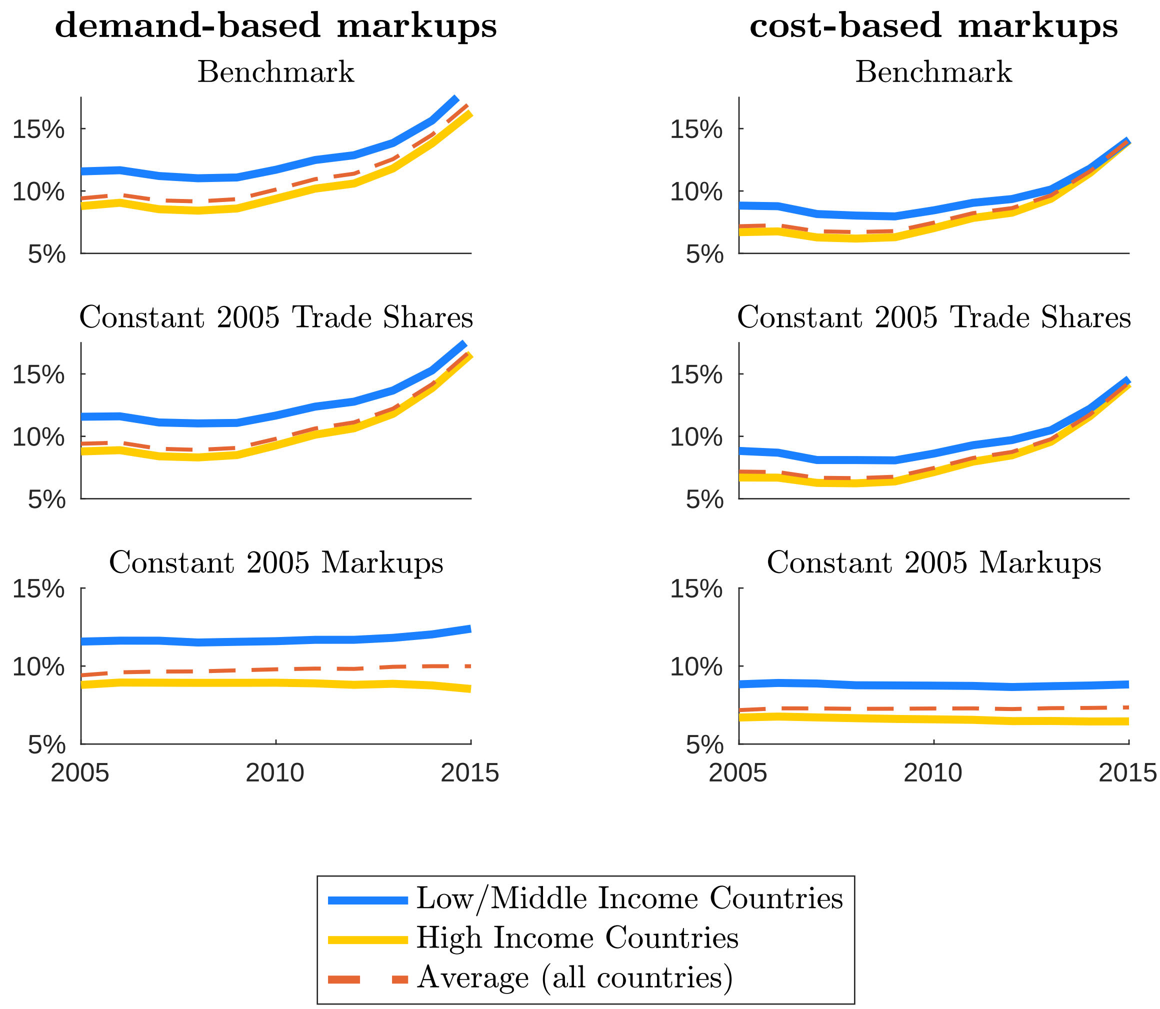


Note: Cost-Based markups are estimated using the WORLDSCOPE and COMPUSTAT datasets. Industry classifications are based on the ICIO database. Demand-based markups are estimated with data on Colombian import transactions from PANJIVA. This data covers the period from 2007 to 2016, allowing us to estimate markups for 2008 onwards.

on Lemma 3, which account for input-output linkages. The results are displayed in the top panel of Figure 4 for all eleven years in our sample. The y-axis in Figure 4, is $\mathcal{D}_i$, which denotes the percent increase in real income in response to markup correction. Instead of reporting values for individual countries, each panel in Figure 4 displays average values for high-income and low/middle-income countries—classified under Table 1. The left panel reports numbers implied by our cost-based markup estimates, while the right panel corresponds to our demand-based markup estimates.

Figure 4 highlights two systematic patterns. First, the cost of markups has risen over time, which echos the past literature. Figure 4, meanwhile, uncovers two additional patterns. It shows that the the cost of markups has risen not only in the US but universally—all while accounting for input-output linkages. This pattern, moreover, holds irrespective of which methods markups are estimated with. The previous literature has mostly documented the rise in markup distortions using cost-based markup estimates, while Figure 4 reveals that the same pattern is borne out of demand-based markup

Figure 4: The welfare cost of markups and its drivers over time



Note: The above graph reports the welfare cost of markups for national economies. For example, a 5% figure implies that markups lower national welfare by 5% relative to the efficient level. The figures in the top panel are computed using Lemma 3 (which accounts for input-output linkages) using annual data on markups and trade/expenditure shares. The figures in the middle panel are computed by assuming that trade/expenditure shares remain constant at their 2005 level. The figures in the bottom panel are computed by assuming that markups remain constant at their 2005 level. Data on industry-level expenditure, trade, production, and input-output shares are from the ICIO.

estimates.[25]

Second, Figure 4 indicates that markups are markedly more costly for low-income countries. Again, this result arises irrespective of how markups are estimated. It also emerges despite our

[25]The global rise in the welfare cost of markups is not driven by a few outliers but is a common occurrence in our sample of countries. To elucidate this claim, Figure 12 in the appendix reports the welfare cost of markups and its evolution for select countries. The cost of markups has increased in each country in a way that mirrors the global trend. Figure 12 also brings clarity to the role of cross-substitutability between industries. The implied cost of markups is noticeably higher when we assign a higher value to η, the cross-industry elasticity of substitution. The cost of markups for the US economy, for example, more than doubles when we elevate η from the baseline value of 1 to 2. The intuition is that the greater the substitutability between industries, the greater the extent to which markups distort the cross-industry allocation of resources.

conservative approach, which computes average industry-level markups based on the presumption that these averages are uniform across low- and high-income countries. In that regards, Figure 4 indicates that for a common vector of markups, low-income countries exhibit a patten of specialization and expenditure that lends itself to greater markup-driven misallocation.

The lower panels in Figure 4 indicate that the rising cost of markups is driven primarily by the change in markup levels over time. This is done by decomposing the change in markup distortions into changes driven by (1) adjustment to trade shares over time, and (2) adjustments to markup levels over time. The middle panel fixes trade and expenditure shares to their 2005 level and tracks the change in $\mathcal{D}_i$ in response to changing markup levels. The bottom panel fixes markups to their 2005 level and tracks the change in $\mathcal{D}_i$ in response to changing trade shares. Evidently, the change in markup levels explains pretty much the entire rise in $\mathcal{D}_i$ over time.

7.3 The Unequal Impacts of International Rent-shifting

Next, we use Propositions 1-4 to compute the extent to which trade openness has impacted the cost of markups via international rent-shifting. As explained in Section 4.2, trade openness induces sectoral specialization, reallocating resources to comparative advantage industries in each country. These adjustments trigger international rent-shifting, dampening the welfare cost of markups for countries that specialize in high-markup industries while amplifying it for others.

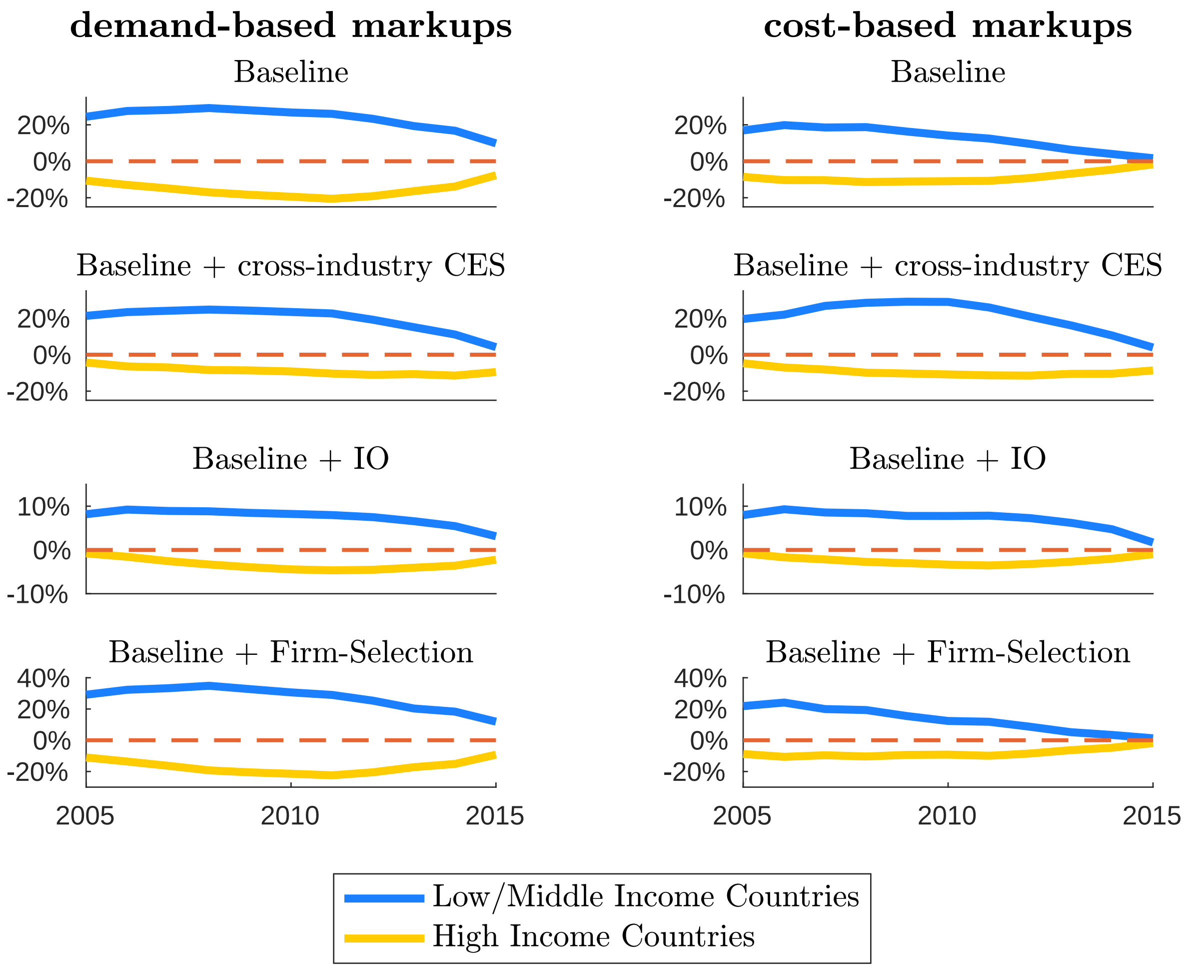
Figure 5 reports the change in the cost of markups due to trade and rent-shifting. We report results by income group to contrast impacts across low- and high-income nations. A stark takeaway from this figure is that –through international rent-shifting– trade has systematically amplified the cost of markups for low-income countries, while lowering it for high-income nations. This pattern prevails irrespective of which method markups are estimated with. It is also robust to the presence of CES demand across industries, input-output (IO) networks, and firm-selection into export markets.

In terms of magnitudes, our baseline model indicates that international rent-shifting has amplified the welfare cost of markups for low-income countries by up to 21% in given fiscal year. These effects are even larger when we control for firm-selection into export markets, but smaller in the presence of IO linkages. To understand this, note that the cost associated with markup dispersion is generally larger under IO linkages. The degree of rent-shifting is, by contrast, smaller, due to integration in the global IO network. As a result, one expects that international rent-shifting constitutes a smaller fraction of $\mathcal{D}_i$ under IO linkages.

Digging deeper, Figure 13 (in the appendix) maps out the impacts of rent-shifting on a country-by-country basis using our baseline model and cost-based markup estimates. The map highlights the heterogenous impacts of international rent-shifting even within low and middle-income countries. Rent-shifting, for instance, dampens the cost of markups for the Indian economy by around 20%, while exacerbating the cost for Indonesia and Brazil by around 15%. The same level of heterogeneity emerges under various modeling specifications—as reported in Table 3 of the appendix.[26]

[26]For most countries, the impacts of international rent-shifting have a stable sign across specifications. In a few instances,

Figure 5: The change in the cost of markups due to trade and rent-shifting



Note: The above graph reports the percent change in the cost of markups attributable to trade (and rent-shifting in particular). For example, a 5% change corresponds to a 5% increase in the cost of markups due to rent-shifting. The figures in the top, middle, and bottom panels are computed using the formulas under Lemmas 1, 2, and 3 respectively. Data on industry-level expenditure, production and input-output shares are from the ICIO.

7.4 Discussions and Limitations

The Anatomy of International Rent-Shifting. Figure 5 revealed systematic rent-shifting from low- to high-income countries;[27] but it concealed the within-income group heterogeneity in exposure to rent-shifting. Figure 5 is also mute on the rent transfers associated with each bilateral trade relation. Figure 6 reveals this information—tracking bilateral rent transfers from low-income countries to individual high-income trading partners. Low-income countries are visibly net payers of markup rents

however, accounting for input-output linkages or firm-selection reverses our baseline predictions. To give an example, the baseline model predicts that international rent-shifting exacerbates the cost of markups for Singapore, but accounting for input-output linkages and firm-selection reverses this prediction. At the end, the country-specific results in Table 3 indicate that non-income-related factors matter for exposure to international rent-shifting—a point we elaborate on in the following section.

[27] Outside of the baseline specification, trade can also exacerbate "markup dispersion." Though, comparing across specification, one can infer that dispersion effects are not the root cause of the high- versus low-income asymmetry.

to high-income countries, but the rents are not allocated equally across high-income partners. Japan, Korea, and major European economies are the primary recipients of these rents.

The Deep Origins of International Rent-Shifting. Exposure to international rent-shifting is determined by a country's pattern of specialization across low- and high-markup industries. Accordingly, Figure 5 suggests that high-income countries tend to have a comparative advantage in high-markup industries. What explains this pattern of comparative advantage? Is it national institutions, resource endowments, or geography? We explore these questions in Appendix I. Figure 8 in this appendix plots the relationship between the national-level cost of rent-shifting ($\Delta\mathcal{D}_i$) and national institutions, resource endowment, and geography. It indicates that better labor and credit market institutions are associated with specialization in high-markup industries and positive exposure to rent-shifting. By contrast, natural resource abundance is associated with greater adverse exposure to rent-shifting—resembling a *resource curse*.

What Explains the Evolution of Rent-Shifting Patterns Over Time? Figure 5 indicates that rent-shifting from low- to high-income nations has dampened over time. One explanation is that middle-income nations have become more specialized in high-markup industries. Another explanation is that markup levels have evolved in a way that dampens rent-shifting from low-to-high-income nations. We explore these two possibilities in Appendix J. The former effect (i.e., changes in cross-industry specialization patterns) explains almost all the dampening in high-to-low income rent-shifting over time—as illustrated by Figure 10 in Appendix J.

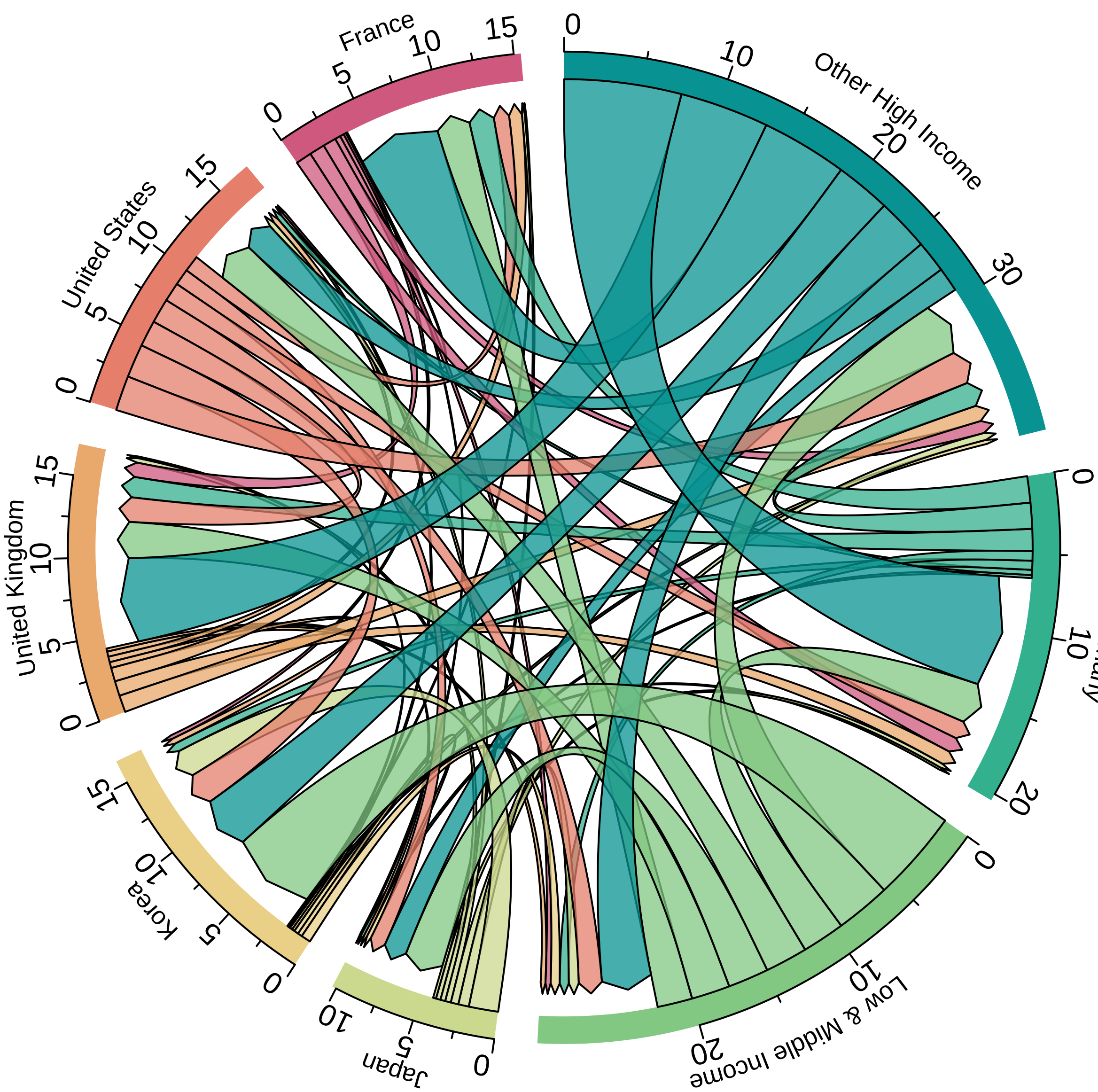
Data Limitation. Our results –while informative– are plagued by a notable data limitation. We assume that profits are fully rebated to households in a firm's country of origin. This assumption is partially relaxed when we introduce fixed marketing costs into our analysis—in which case gross profits are partially rebated to workers in the location of sales. Ideally, however, we would like to track profit payments to international shareholders.[28] From a theoretical standpoint, we can easily amend our sufficient statistics formulas to account for such payments. From a data standpoint, however, we lack the data to measure these payment flows and their implications for rent-shifting. Future work is needed to compile such data, aiding with improved measurement of international rent-shifting.

8 Implications for International Policy

The previous section established that trade has prompted systematic rent-shifting from low- to high-income countries. As a result, low-income countries have come to bear the brunt of the welfare cost

[28]This limitation may help explain why for some developing countries where a large share of exports are generated by foreign multi-national corporations (e.g., China, Hong Kong, Thailand) trade and rent-shifting remedies the cost of markups.

Figure 6: International patterns of rent-shifting



Note: An arrow from country i to j corresponds to the profits or rents collected by firms in country j from consumers in country i as a percentage of the profits collected by domestic firms. Profits or rents are calculated using sales data from ICIO and markups estimated by applying the cost-based methodology to WORLDSCOPE data—both in year 2010.

associated with markups. This observation raise a natural question: What policy interventions can neutralize international rent-shifting and level the playing field for low-income nations. One obvious answer is that governments around the globe correct markup distortions using domestic policies. This solution is perhaps controversial, as it violates WTO rules. We identify another solution, which is compatible with existing trade agreements. Our solution builds on a new duality result that asserts equivalence between international rent-shifting and imbalanced import tariffs.

To present our duality result clearly, we introduce additional notation. Let $\mathbf{t} = \left\{t_{ji,k}\right\}_{j,i,k}$ denote the vector of applied tariffs where element $t_{ji,k}$ denotes the tariffs collected by destination i from goods sourced from origin j–industry k.[29] Tariffs generate revenue and create an additional wedge between the price and marginal cost of internationally traded goods. In the case of market i, the tariff-inclusive prices can be specified as

$$P_{ji,k} = (1 + t_{ji,k})\, \mu_k \tau_{ji,k} w_j / \overline{\varphi}_{j,k}; \qquad\qquad P_{ii,k} = \mu_k \tau_{ii,k} w_i / \overline{\varphi}_{i,k}.$$

Accounting for tariffs, total income in country i equals the sum of wage income, markup rents, and tariff revenues. In particular,

$$Y_i = w_i L_i + \sum_k \sum_j \left[\left(1 - \frac{1}{\mu_k}\right) \frac{1}{1 + t_{ij,k}} \lambda_{ij,k} e_{j,k} Y_j \right] + \sum_k \sum_j \left[\frac{t_{ji,k}}{1 + t_{ji,k}} \lambda_{ij,k} e_{j,k} Y_i \right].$$

National welfare in country i is, $W_i = Y_i / P_i$, where national income, Y_i, and the consumer price index, P_i, are functions of tariffs and markups wedges. According, we can express welfare the as an explicit function of globally applied tariffs and markups, $\boldsymbol{\mu} = \{\mu_1, ..., \mu_K\}$. Namely, $W_i \equiv \mathcal{W}_i\,(\,\mathbf{t}\,,\,\boldsymbol{\mu}\,)$.

Now, perform the following thought experiment. Suppose applied tariffs are sufficiently small and we eliminate markups globally, lowering them from $\boldsymbol{\mu}$ to $\mathbf{1}$. This change impacts welfare by extinguishing *both* markup dispersion and international rent-shifting. We show that an accompanying adjustment to import tariffs can neutralize both effects and preserve welfare worldwide. More specifically, a globally-uniform tariff hike can reproduce the welfare loss associated with markup dispersion. An increase in the tariff collected by countries who benefit from rent-shifting paired with a reduction in others' tariffs can replicate international rent-shifting effects. In other words, while tariffs and markups create seemingly different price distortions, there is a strict welfare duality between them. The following proposition outlines this claim.

Proposition 5. [Duality between tariffs and markups] *Suppose applied tariffs, $\boldsymbol{t}$, are sufficiently small and trade elasticities, $\boldsymbol{\sigma}$, are sufficiently homogeneous across industries. The welfare effects (including rent-shifting) associated with markups, $\boldsymbol{\mu}$, are observationally equivalent to a hidden tariff, $\tilde{\boldsymbol{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, where element $\tilde{t}_n$ is a uniform tariff applied to all goods imported by country n. Stated formally,*

$$\mathcal{W}_i\,(\,\boldsymbol{t} + \tilde{\boldsymbol{t}}\,,\,\mathbf{1}\,) = \mathcal{W}_i\,(\,\boldsymbol{t}\,,\,\boldsymbol{\mu}\,) \qquad \forall i = 1,..,N$$

[29] To be clear, our baseline model assumed that countries apply near zero tariffs, which corresponds to $\mathbf{t} \approx \mathbf{0}$.

where $\tilde{t}_n$ is increasing in the net rents collected by country n from the rest of the world;[30] *and $\mathcal{W}_i(\boldsymbol{t}, \boldsymbol{\mu})$ denotes national welfare under the status quo.*

The above proposition is proven in Appendix K by invoking the uniformity of optimal tariffs in efficient (markup-free) economies. Converting markup distortions into tariff-equivalent distortions is advantageous because it unveils policy interventions compatible with the existing nexus of shallow trade agreements. Under the WTO, for instance, tariffs should satisfy the principle of reciprocity (Bagwell and Staiger (1999)). Proposition 5 indicates that when the economy is plagued with markup distortions, market access concessions under the WTO may violate reciprocity even when applied tariffs appear symmetric and balanced. To ensure reciprocity, countries must adjust their tariffs depending on whether they are net payers or receivers of markup rents from the rest of the world. The patterns of rent-shifting documented in Section 7 suggest that these adjustments would amount to a one-sided tariff reduction by high-income countries on low-income partners.

Next, we present a procedure to recover the hidden tariff, $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, from observables and our estimated markup values. We use the hat-algebra notation whereby $\hat{x}$ denotes the change in a generic variable, x, when moving from the factual equilibrium, associated with $(\mathbf{t}, \boldsymbol{\mu})$, to the counterfactual equilibrium, associated with $(\mathbf{t} + \tilde{\mathbf{t}}, \mathbf{1})$. With the aid of hat-algebra notation, we express the equilibrium and welfare-neutrality conditions in changes. For simplicity, suppose applied tariffs are near zero under the status quo, i.e., $\mathbf{t} \approx \mathbf{0}$, which is broadly-consistent with real-world data. The *labor market clearing condition* equalizes wage payments in country i with net sales, which exclude of tariff and markup-related payments. Namely,

$$\hat{w}_i w_i \bar{L}_i = \sum_k \sum_n \left[\frac{1}{\mu_k \left(1 + \tilde{t}_n\right)} \times \frac{\lambda_{in,k} \left(1 + \tilde{t}_n\right)^{1-\sigma_k} \hat{w}_i^{1-\sigma_k}}{\lambda_{nn,k} \hat{w}_n^{1-\sigma_k} + \sum_{j \neq n} \lambda_{jn,k} (1 + \tilde{t}_n)^{1-\sigma_k} \hat{w}_j^{1-\sigma_k}} e_{n,k} \hat{Y}_n Y_n \right]. \tag{14}$$

The above formulation appeals to the fact that for any variable like the wage rate, $w_i' = \hat{w}_i w_i$. Also, given the CES import demand structure, $\hat{\lambda}_{in,k} = (1 + t_n^*)^{1-\sigma_k} \hat{w}_i^{1-\sigma_k} \hat{P}_{n,k}^{1-\sigma_k}$, where the change in the CES price index is $\hat{P}_{n,k}^{\sigma_k - 1} = \left(\lambda_{nn,k} \hat{w}_n^{1-\sigma_k} + \sum_{j \neq n} \lambda_{jn,k} (1 + t_n^*)^{1-\sigma_k} \hat{w}_j^{1-\sigma_k} \right)$. Next, we specify the national budget constraint, which imposes that total income be equal to wage payments plus tariff revenues and markup rents. Namely,

$$\hat{Y}_i Y_i = \frac{\hat{w}_i w_i L_i + \sum_k \sum_n \left[\frac{\left(1 - \frac{1}{\mu_k}\right) \lambda_{in,k} (1 + \tilde{t}_n)^{1-\sigma_k} \hat{w}_i^{1-\sigma_k}}{\lambda_{nn,k} \hat{w}_n^{1-\sigma_k} + \sum_{j \neq n} \lambda_{jn,k} (1 + \tilde{t}_n)^{1-\sigma_k} \hat{w}_j^{1-\sigma_k}} e_{n,k} \hat{Y}_n Y_n \right]}{1 - \sum_k \sum_{n \neq i} \left[\frac{\left(1 - \frac{1}{1 + \tilde{t}_i}\right) \lambda_{ni,k} (1 + \tilde{t}_i)^{1-\sigma_k} \hat{w}_n^{1-\sigma_k}}{\lambda_{ii,k} \hat{w}_i^{1-\sigma_k} + \sum_{j \neq i} \lambda_{ji,k} (1 + \tilde{t}_i)^{1-\sigma_k} \hat{w}_j^{1-\sigma_k}} e_{i,k} \right]}. \tag{15}$$

Finally, we present the welfare-neutrality condition, which ensures that replacing markups, $\boldsymbol{\mu}$, with

[30] More formally, $\tilde{t}_n$ is higher the larger $\text{Cov}_{e_i}\left(\frac{1}{\mu}, \text{RCA}_i\right)$.

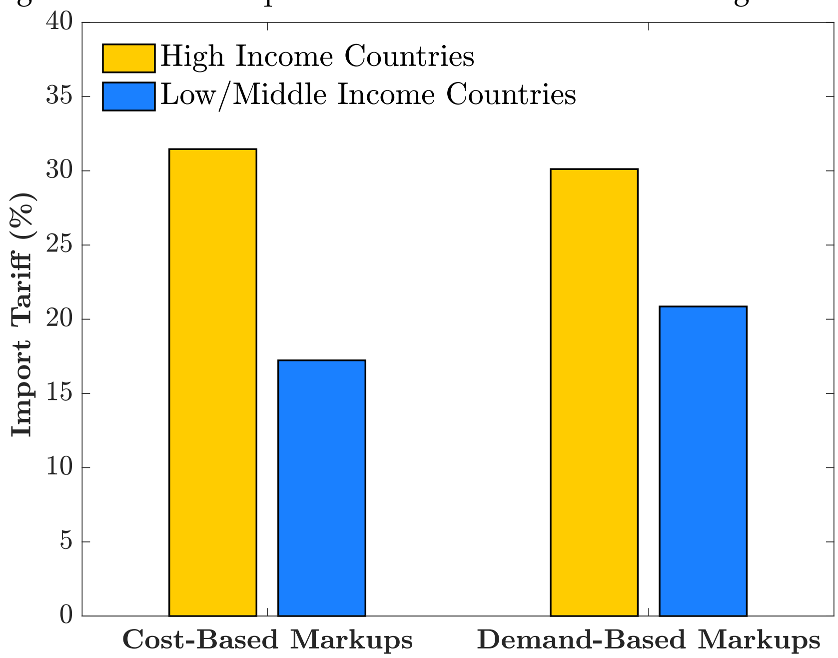
hidden tariffs, $\tilde{\mathbf{t}}$, preserves national welfare in every country. This condition can be expressed as

$$\hat{W}_i = \frac{\hat{Y}_i}{\hat{P}_i} = 1, \qquad \text{where} \qquad \hat{P}_i = \prod_k \left[\lambda_{ii,k} \hat{w}_i^{1-\sigma_k} + \sum_{n \neq i} \lambda_{ni,k} (1 + t_i^*)^{1-\sigma_k} \hat{w}_n^{1-\sigma_k} \right]^{\frac{e_{i,k}}{1-\sigma_k}}. \tag{16}$$

To take stock, Equations (14)-(16) specify a system of $3N$ independent equations and $3N$ independent unknowns, which are $\{\hat{w}_i, \hat{Y}_i, \tilde{t}_i\}_i$. Solving this system recovers the vector of hidden tariffs, $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, that mimics the international rent-shifting effects associated with, $\boldsymbol{\mu}$. The following proposition outlines this result.

Proposition 6. *The vector of hidden tariffs, $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, that mimics international rent-shifting effects can be recovered by simultaneously solving Equations (14)-(16). This task, moreover, requires information on only observables, markups, and trade elasticities, $\mathbf{X} = \{\lambda_{ni,k}, e_{i,k}, Y_i, w_i L_i, \mu_k, \sigma_k\}$.*

Figure 7: The tariff-equivalent of international rent-shifting effects



Note: Each bar represents the GDP-weighted average hidden markup-equivalent tariff among low- and high-income countries. The tariff rates for each country are calculated using Proposition 6. Our calculations uses sales and production data from ICIO and the markups estimated in Section (7)—both in year 2010. Industry-level trade elasticity values are borrowed from Caliendo and Parro (2015).

Guided by Proposition 6, we calculate the markup-equivalent (hidden) tariffs using data from ICIO, our estimated markups, and trade elasticities from Caliendo and Parro (2015). Figure 7 summarizes these hidden tariff values, reporting average rates for low- and high-income countries. Based on our calculations, international rent-shifting is akin to an imbalanced hidden tariff structure, wherein high-income countries collected an excess 10% tariff from low-income partners. That is, they are net receivers of tariff revenues from low-income partners in this dual interpretation. These results add new light to the balance of tariff concession under the WTO. Tariffs applied by WTO members tend to satisfy *nominal* reciprocity, as defined under the WTO principles. But our results indicate that they may not satisfy *real* reciprocity. As a result of rent-shifting, low-income countries are effectively granting more market access concessions than they receive under WTO principles. Accordingly, to restore *real* reciprocity, high-income countries must unilaterally lower their currently-applied tariffs by more than 10%.

9 Conclusion

The global rise in market power and trade openness are two hallmarks of the current economic era. We show that these developments have prompted significant welfare transfers from low-income to high-income countries via international rent-shifting. These effects are isomorphic to an excess hidden tariff collected by high-income nations—one that eludes standard calculations of the market access but disrupts the balance of concession under existing trade agreements. Two policy reforms can mitigate the burden of international rent-shifting on low-income countries. The first reform involves markup correction via domestic policies, which, among other reservations, violates WTO rules. The second reform requires that high-income countries unilaterally lower their tariffs on low-income partners by 7%, which aligns with WTO's principle of reciprocity when interpreted correctly.

Our calculation of international rent-shifting capitalizes on recent data availability and advancements in markup estimation. We provide one of the first large-scale estimations of markups using two conventional techniques. Yet we make many simplifying assumptions to handle residual data and methodological limitations. With the aid of better data, future work can reevaluate international rent-shifting by tracking profit payments across international shareholders and accounting for markup heterogeneity within narrowly-defined industries—both of which are beyond the reach of our elementary analysis.

References

Acemoglu, D., P. Antràs, and E. Helpman (2007). Contracts and technology adoption. *American Economic Review 97*(3), 916–943.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Aghion, P., A. Bergeaud, T. Boppart, P. J. Klenow, and H. Li (2019). A theory of falling growth and rising rents. Technical report, National Bureau of Economic Research.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2019). The elusive pro-competitive effects of trade. *The Review of Economic Studies 86*(1), 46–80.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Asker, J., A. Collard-Wexler, and J. De Loecker (2019). (mis) allocation, market power, and global oil extraction. *American Economic Review 109*(4), 1568–1615.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Atkin, D. and D. Donaldson (2021). The role of trade in economic development. Technical report, National Bureau of Economic Research.

Bagwell, K. and R. W. Staiger (1999). An economic theory of gatt. *American Economic Review 89*(1), 215–248.

Bagwell, K. and R. W. Staiger (2012). Profit shifting and trade agreements in imperfectly competitive markets. *International Economic Review 53*(4), 1067–1104.

Bai, Y., K. Jin, and D. Lu (2019). Misallocation under trade liberalization. Technical report, National Bureau of Economic Research.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage 1. *The manchester school 33*(2), 99–123.

Ballance, R. H., H. Forstner, and T. Murray (1987). Consistency tests of alternative measures of comparative advantage. *The review of economics and statistics*, 157–161.

Baqaee, D. and E. Farhi (2019). Networks, barriers, and trade. Technical report, National Bureau of Economic Research.

Baqaee, D. and E. Farhi (2020). Entry vs. rents. Technical report, National Bureau of Economic Research.

Beck, T. (2002). Financial development and international trade: Is there a link? *Journal of international Economics 57*(1), 107–131.

Behrens, K., G. Mion, Y. Murata, and J. Suedekum (2020). Quantifying the gap between equilibrium and optimum under monopolistic competition. *The Quarterly Journal of Economics 135*(4), 2299–2360.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berthou, A., J. J.-H. Chung, K. Manova, and C. S. D. Bragard (2020). Cep discussion paper no 1668 january 2020 trade, productivity and (mis) allocation.

Bhagwati, J. and V. K. Ramaswami (1963). Domestic distortions, tariffs and the theory of optimum subsidy. *Journal of Political economy 71*(1), 44–50.

Botero, J. C., S. Djankov, R. L. Porta, F. Lopez-de Silanes, and A. Shleifer (2004). The regulation of labor. *The Quarterly Journal of Economics 119*(4), 1339–1382.

Brander, J. A. and B. J. Spencer (1985). Export subsidies and international market share rivalry. *Journal of international Economics 18*(1-2), 83–100.

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

Caliendo, L., F. Parro, and A. Tsyvinski (2022). Distortions and the structure of the world economy. *American Economic Journal: Macroeconomics forthcoming*.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics 82*(2), 152–167.

Chow, D. C., I. Sheldon, and W. McGuire (2018). How the united states withdrawal from the trans-pacific partnership benefits china. *U. Pa. JL & Pub. Aff. 4*, 37.

Chung, J. H. (2019). *Firm heterogeneity, misallocation, and trade*. Auburn University, Department of Economics.

Costinot, A. (2009). On the origins of comparative advantage. *Journal of International Economics 77*(2), 255–264.

Costinot, A. and A. Rodríguez-Clare (2014). Trade theory with numbers: Quantifying the consequences of globalization. In *Handbook of international economics*, Volume 4, pp. 197–261. Elsevier.

Costinot, A. and I. Werning (2019). Lerner symmetry: A modern treatment. *American Economic Review: Insights 1*(1), 13–26.

Cuñat, A. and M. J. Melitz (2012). Volatility, labor market flexibility, and the pattern of comparative advantage. *Journal of the European Economic Association 10*(2), 225–254.

De Blas, B. and K. N. Russ (2015). Understanding markups in the open economy. *American Economic Journal: Macroeconomics 7*(2), 157–80.

De Loecker, J. and J. Eeckhout (2018). Global market power. Technical report, National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American economic review 102*(6), 2437–71.

Dhingra, S. and J. Morrow (2019). Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy 127*(1), 196–232.

Díez, F. J., J. Fan, and C. Villegas-Sánchez (2021). Global declining competition? *Journal of International Economics 132*, 103492.

Eaton, J. and G. M. Grossman (1986). Optimal trade and industrial policy under oligopoly. *The Quarterly Journal of Economics 101*(2), 383–406.

Edmond, C., V. Midrigan, and D. Y. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221.

Edmond, C., V. Midrigan, and D. Y. Xu (2018). How costly are markups? Technical report, National Bureau of Economic Research.

Epifani, P. and G. Gancia (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics 83*(1), 1–13.

Feenstra, R. C. and D. E. Weinstein (2017). Globalization, markups, and us welfare. *Journal of Political Economy 125*(4), 1040–1074.

Firooz, H. and G. Heins (2020). Profit shifting, import and export markups, and the gains from trade. *Import and Export Markups, and the Gains from Trade (November 25, 2020)*.

Gustavsson, P., P. Hansson, and L. Lundberg (1999). Technology, resource endowments and international competitiveness. *European economic review 43*(8), 1501–1530.

Hall, R. E. and C. I. Jones (1999). Why do some countries produce so much more output per worker than others? *The Quarterly Journal of Economics 114*(1), 83–116.

Head, K. and B. J. Spencer (2017). Oligopoly in international trade: Rise, fall and resurgence. *Canadian Journal of Economics/Revue canadienne d'économique 50*(5), 1414–1444.

Hodler, R. (2006). The curse of natural resources in fractionalized countries. *European Economic Reviews 50*(6), 1367–1386.

Holmes, T. J., W.-T. Hsu, and S. Lee (2014). Allocative efficiency, mark-ups, and the welfare gains from trade. *Journal of International Economics 94*(2), 195–206.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010). The worldwide governance indicators: Methodology and analytical issues. *World Bank Policy Research Working 5430*.

King, R. G. and R. Levine (1993). Finance and growth: Schumpeter might be right. *The Quarterly Journal of Economics 108*(3), 717–737.

Kletzer, K. and P. Bardhan (1987). Credit markets and patterns of international trade. *Journal of Development Economics 27*(1-2), 57–70.

Krugman, P. (1980, December). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review 70*(5), 950–959.

Krugman, P. (1987). The narrow moving band, the dutch disease, and the competitive consequences of mrs. thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of Development Economics 27*(1-2), 41–55.

Kucheryavyy, K., G. Lyn, and A. Rodríguez-Clare (2016). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. Technical report, National Bureau of Economic Research.

Lane, P. R. and A. Tornell (1996). Power, growth, and the voracity effect. *Journal of Economic Growth 1*(2), 213–241.

Lashkaripour, A. (2021). The cost of a global tariff war: A sufficient statistics approach. *Journal of International Economics 131*, 103419.

Lashkaripour, A. and V. Lugovskyy (2021). Profits, scale economies, and the gains from trade and industrial policy.

Levchenko, A. A. (2007). Institutional quality and international trade. *The Review of Economic Studies 73*(3), 791–819.

Liu, E. (2019). Industrial policies in production networks. *The Quarterly Journal of Economics 134*(4), 1883–1948.

Manova, K. (2013). Credit constraints, heterogeneous firms, and international trade. *Review of Economic Studies 80*(2), 711–744.

Matsuyama, K. (2005). Credit market imperfections and patterns of international trade and capital flows. *Journal of the European Economic Association 3*(2-3), 714–723.

Mehlum, H., K. Moene, and R. Torvik (2006). Institutions and the resource curse. *The Economic Journal 116*(508), 1–20.

Melitz, M. J. (2003, November). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Mrázová, M. (2011). Trade agreements when profits matter. *V Manuscript, London School of Economics*.

Mrázová, M., J. P. Neary, and M. Parenti (2021). Sales and markup dispersion: theory and empirics. *Econometrica 89*(4), 1753–1788.

Nocco, A., G. I. Ottaviano, and M. Salto (2014). Monopolistic competition and optimum product selection. *American Economic Review 104*(5), 304–09.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics 122*(2), 569–600.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Ossa, R. (2012). Profits in the" new trade" approach to trade negotiations. *American Economic Review 102*(3), 466–69.

Rajan, R. G. and L. Zingales (1998). Financial dependence and growth. *American Economic Review 88*(3), 559–586.

Salanié, B. and F. A. Wolak (2019). Fast,"robust", and approximately correct: estimating mixed demand systems. Technical report, National Bureau of Economic Research.

Tang, H. (2012). Labor market institutions, firm-specific skills, and trade patterns. *Journal of International Economics 87*(2), 337–351.

Van der Ploeg, F. (2011). Natural resources: curse or blessing? *Journal of Economic Literature 49*(2), 366–420.

Appendix (for online publication)

A The Welfare Cost Markups (Lemma 1)

Characterizing the Efficient Allocation: Lemma 1 builds on the observation that efficiency requires that prices feature a uniform (or no) markup over marginal cost. This observation follows trivially from the First Welfare Theorem. To fix ideas, we also present a formal proof with derivations below. As before, we use the superscript A to denote our focus on autarky in this appendix. Following Section 4, we can fully represent the economy (that consists of monopolistically competitive firms) with industry-level aggregates for price and quantity: $P_{i,k}$ and $Q_{i,k}$. The planner's problem under this representation can be expressed as

$$\max \; U(Q_{i,1}, ..., Q_{i,K}) \quad s.t \; \sum_k \left[Q_{i,k} / \overline{\varphi}_{i,k} \right] = L_i.$$

The first-order conditions associated with the above problem yield the following expression for relative prices that ensure the implementation of the efficient allocation:

$$\frac{P^{A\star}_{i,g}}{P^{A\star}_{i,k}} = \frac{\partial U(\mathbf{Q}) / \partial Q_{i,g}}{\partial U(\mathbf{Q}) / \partial Q_{i,k}} = \frac{1/\overline{\varphi}_{i,g}}{1/\overline{\varphi}_{i,k}} = \frac{w^{A\star}_i / \overline{\varphi}_{i,g}}{w^{\star A}_i / \overline{\varphi}_{i,k}}.$$

The above equation asserts that price indexes under the efficient allocation should be set equal to the marginal cost indexes, which leads to zero aggregate profits:

$$P^{A\star}_{i,k} = w^{A\star}_i / \overline{\varphi}_{i,k} \quad \longrightarrow \quad \Pi^{A\star}_i = 0.$$

Change in Welfare after Implementing the Efficient Allocation: Given the Cobb-Douglas-CES utility function, social welfare in country i (under autarky, A) can be formulated as

$$W^A_i = \frac{Y^A_i}{P^A_i} = \frac{Y^A_i}{\prod_k \left(P^A_{i,k} \right)^{e_{i,k}}} = \frac{w^A_i L_i + \Pi^A_i}{\prod_k \left(\mu_k w^A_i / \overline{\varphi}_{i,k} \right)^{e_{i,k}}}.$$

Next, we simplify the numerator in the above equation. Profits and total wage payments in economy i (under autarky) are given by

$$\Pi^A_i = \sum_k \left[\left(1 - \frac{1}{\mu_k} \right) P^A_{ii,k} Q^A_{ii,k} \right]; \qquad w^A_i L_i = \sum_k \left[\frac{1}{\mu_k} P^A_{ii,k} Q^A_{ii,k} \right]$$

Combining theses two expressions we can write total income in country i in terms of the average markup and wage payments. In particular,

$$
\begin{aligned}
w_i^A L_i + \Pi_i^A &= \sum_k \left[P_{ii,k}^A Q_{ii,k}^A\right] \times \frac{w_i^A L_i}{w_i^A L_i} = \frac{\sum_k \left[P_{ii,k}^A Q_{ii,k}^A\right]}{\sum_k \left[\frac{1}{\mu_k} P_{ii,k}^A Q_{ii,k}^A\right]} \times w_i^A L_i, \\
&= \frac{1}{\sum_k \left[\frac{1}{\mu_k} \frac{P_{ii,k}^A Q_{ii,k}^A}{\sum_{k'} \left[P_{ii,k'}^A Q_{ii,k'}^A\right]}\right]} \times w_i^A L_i = \left(\sum_k \frac{1}{\mu_k} e_{i,k}\right)^{-1} w_i^A L_i = \left(\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]\right)^{-1} w_i^A L_i
\end{aligned}
$$

where the second line uses the fact that under autarky, $\frac{P_{ii,k} Q_{ii,k}}{\sum_{k'} P_{ii,k'} Q_{ii,k'}} = e_{i,k}$. The third line follows from our choice of notation, whereby $\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] = \sum_k \left[\frac{1}{\mu_k} e_{i,k}\right]$. Appealing to the above expression for total income, we can calculate the change in welfare when moving from the market equilibrium to the efficient equilibrium—all under autarky:

$$
\frac{W_i^{A\star}}{W_i^A} = \frac{Y_i^{A\star}}{Y_i^A} \times \frac{P_i^A}{P_i^{A\star}} = \frac{w_i^{A\star} L_i}{\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{-1} w_i^A L_i} \times \frac{\prod_k \left(\mu_k w_i^A / \overline{\varphi}_{i,k}\right)^{e_{i,k}}}{\prod_k \left(w_i^{A\star} / \overline{\varphi}_{i,k}\right)^{e_{i,k}}}.
$$

Treating wage as the numeraire, i.e, $w_i^A = w_i^{A\star} = 1$, the above expression yields the expression for $\mathcal{D}_i^A$ presented under Lemma 1:

$$
\frac{W_i^{\star A}}{W_i^A} = \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] \times \prod_k \left(\mu_k^{e_{i,k}}\right) \quad \Longrightarrow \quad \mathcal{D}_i^A \equiv \log\left(\frac{W_i^{A\star}}{W_i^A}\right) = \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log\frac{1}{\mu}\right].
$$

Deriving the Approximation for $\mathcal{D}_i^A$. We use can use Taylor's theorem to link $\mathcal{D}_i^A$ to cross-industry markup dispersion. For a generic industry-level variable x_k, The Taylor expansion of function $\mathbb{E}_{e_i}[\log x] = \sum_i [e_i \log x_i]$ around $\bar{x}_i \equiv \mathbb{E}_{e_i}[x]$ can be expressed as

$$
\sum_k \left[e_{i,k} \log x_k\right] \approx \log \bar{x}_i + \sum_k \left[\frac{e_{i,k}}{\bar{x}_i}\left(x_k - \bar{x}_i\right)\right] - \frac{1}{2}\sum_k \left[\frac{e_{i,k}}{\bar{x}_i^2}\left(x_k - \bar{x}_i\right)^2\right]
$$

Note that, given $\bar{x}_i$'s definition, $\sum_k \left[\frac{e_{i,k}}{\bar{x}_i}(x_k - \bar{x}_i)\right] = \frac{1}{\bar{x}_i}\left(\bar{x}_i - \bar{x}_i\right) = 0$, so the second term on the right-hand side collapses to zero. Since, $\log \bar{x}_i = \log \mathbb{E}_{e_i}[x]$, rearranging the above equation yields

$$
\log \mathbb{E}_{e_i}[x] - \mathbb{E}_{e_i}[\log x] \approx \frac{1}{2\bar{x}_i{}^2}\sum_k \left[e_{i,k}\left(x_k - \bar{x}_i\right)^2\right] = \frac{1}{2\bar{x}_i^2}\mathrm{Var}_{e_i}(x).
$$

Plugging $x_k = \frac{1}{\mu_k}$ into the above equation delivers $\mathcal{D}_i \approx \frac{\bar{\mu}_i^2}{2}\mathrm{Var}_{e_i}\left(\frac{1}{\mu}\right)$, where $\frac{1}{\bar{\mu}_i} \equiv \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]$.

B The Case with Free Entry

In this appendix we characterize distance to the efficient frontier under free entry. Under free entry the price index of goods associated with closed economy i are given by

$$P_{ii,k} = \mu_k \tau_{ii,k} w_i M_{i,k}^{1-\mu_k},$$

where $M_{i,k}$ denotes the mass of firms serving industry k. Let $f^e_{i,k}$ denote the unit labor cost of entry into industry k in country i, and $L_{i,k}$ denote the number of workers employed in industry k for entry and production purposes. The free entry condition, $w_i L_{i,k} = M_{i,k} w_i \gamma_k f_e$, implies

$$M_{i,k} = \bar{\mathcal{F}}_{i,k} \times r_{i,k},$$

where $\bar{\mathcal{F}}_{i,k}$ is a constant and $r_{i,k} = \frac{P_{i,k} Q_{i,k}}{\sum_g P_{i,g} Q_{i,g}} = \frac{w_i L_{i,k}}{w_i L_i}$ is the revenue share associated with industry k. Following Theorem 1 in Lashkaripour and Lugovskyy (2021), the efficient allocation under free entry is implementable if the social planner offers a industry-level subsidy that equals the inverses markup in industry k. Our goal here is to characterize the welfare gains from implementing efficient subsidies—namely, $\mathcal{D}_i^A = \log \hat{Y}_i^A - \log \hat{P}_i^A$. With the understanding that we are analyzing an closed economy throughout this appendix, we henceforth drop the autarky superscript A to simplify the notation. Also, we assign $w_i = 1$ as the numeraire hereafter. Keeping in mind these choices, we begin with the characterization of $\hat{P}_i$, which in the case of a closed economy is $\hat{P}_i = \prod \hat{P}_{ii,k}^{e_{i,k}}$. Appealing to our earlier expression for $P_{ii,k}$, the change in industry-level price indexes after implementing efficient subsides ($\hat{P}_{ii,k} = P^*_{ii,k}/P_{ii,k}$) is

$$\hat{P}_{ii,k} = \frac{1}{\mu_k} \hat{r}_{i,k}^{1-\mu_k}.$$

To characterize $\hat{r}_{i,k}$, note that total payments to labor in industry k is the inverse subsidy, μ_k, times the consumer expenditure on industry k goods. Namely, $w_i' L_{i,k}' = \mu_k e_{i,k} Y_i$. Also, by definition, $w_i' L_{i,k}' = r_{i,k} w_i L_i$ where $w_i L_i$ is constant by choice of numeraire (i.e., $w_i = 1$). Considering the above, one can derive the following expression for $\hat{r}_{i,k} \equiv r^*_{i,k}/r_{i,k}$:

$$r^*_{i,k} = \frac{\mu_k e_{i,k}}{\sum_g \mu_g e_{i,g}} \quad \Longrightarrow \quad \hat{r}_{i,k} = \frac{\mu_k}{\sum_g \mu_g e_{i,g}} = \frac{\mu_k}{\mathbb{E}_{e_i}[\mu]}.$$

Note the consumer price index in closed economy i is $\hat{P}_i = \prod_k \hat{P}_{ii,k}^{e_{i,k}}$, which in combination with the above expressions for $\hat{P}_{ii,k}$ and $r_{i,k}$, implies

$$\log \hat{P}_i = \sum \left[e_{i,k} \log \left(\mu_k^{-\mu_k} \times \mathbb{E}_{e_i}[\mu]^{\mu_k - 1} \right) \right] = \mathbb{E}_{e_i}[\mu - 1] \log \mathbb{E}_{e_i}[\mu] - \mathbb{E}_e[\mu \log \mu].$$

Next we must characterize the change in total consumption expenditure (or nominal income) after implementing efficient subsidies. With subsidies in place, total consumption expenditure is the sum of wage payments and the negative tax revenue associated with subsidy payments. Specifically, let

T_i^*, denote the (negative) tax revenues associated with efficient subsidies such that $Y_i^* = w_i^* L_i + T_i^*$. It is straightforward to check that $T_i^* = \sum_k (1-\mu_k) e_{i,k} Y_i^*$, which yields the following expression for total consumption expenditure under efficient subsidies:

$$Y_i^* = \frac{1}{\sum_k \mu_k e_{i,k}} w_i^* L_i = \frac{1}{\mathbb{E}_{e_i}[\mu]} w_i^* L_i.$$

Since $w_i^* = w_i = 1$ by choice of numeraire, we immediately get $\log \hat{Y}_i = -\log \mathbb{E}_{e_i}[\mu]$. Plugging the values for $\log \hat{Y}_i$ and $\log \hat{P}_i$ into $\mathcal{D}_i^{\text{FE}}(\mathbf{X}) = \log \hat{Y}_i - \log \hat{P}_i$ yields

$$\mathcal{D}_i^{\text{FE}}(\mathbf{X}) = \mathbb{E}_{e_i}[\mu \log \mu] - \mathbb{E}_{e_i}[\mu] \log \mathbb{E}_{e_i}[\mu].$$

Appealing to the Taylor's expansion of functions $\mathbb{E}_{e_i}[x \log x]$, $\mathbb{E}_{e_i}[x]$, and $\log \mathbb{E}_{e_i}[x]$ around $\bar{x}_i \equiv \mathbb{E}_{e_i}[x]$, yields a simple first-order approximation for distance to the efficient frontier under free entry. In particular,

$$\mathcal{D}_i^{\text{FE}}(\mathbf{X}) \approx \frac{1}{2}\left[\text{CV}_{e_i}\left(\frac{1}{\mu}\right)\right]^2 + \text{Cov}_{e_i}(\mu, \log \mu).$$

C Impact of Trade on the Cost of Markups (Proposition 1)

C.1 Characterizing Real National Income

Given the CES demand function, the share of expenditure on domestic varieties in industry k is given by $\lambda_{ii,k} = \left(\frac{\tau_{ii} w_i / \overline{\varphi}_{i,k}}{P_{i,k}}\right)^{1-\sigma_k}$. Using this equation, we can write the real wage w.r.t. industry k goods as

$$\lambda_{ii,k} = \left(\frac{\tau_{ii} w_i / \overline{\varphi}_{i,k}}{P_{i,k}}\right)^{1-\sigma_k} \implies \frac{w_i \overline{L}_i}{P_{i,k}} = \overline{T}_{i,k} \lambda_{ii,k}^{\frac{1}{1-\sigma_k}},$$

where $\overline{T}_{i,k} \equiv \overline{\varphi}_{i,k} / \tau_{ii,k}$ is a term composed of constants. Noting that the consumer price index is the Cobb-Douglas average of industry-level price indexes, $P_i = \prod_k P_{i,k}^{e_{i,k}}$, the above equation yields

$$\frac{w_i \overline{L}_i}{P_i} = \underbrace{\prod_k \overline{T}_{i,k}^{e_{i,k}}}_{\overline{T}_i} \times \prod_k \lambda_{ii,k}^{\frac{e_{i,k}}{1-\sigma_k}}. \tag{17}$$

Notice that $\frac{w_i \overline{L}_i}{P_i}$ denotes real wage income, which is different from real income, $\frac{Y_i}{P_i} = \frac{w_i \overline{L}_i + \Pi_i}{P_i}$. To determine real income we characterize Π_i in terms of $w_i \overline{L}_i$. To this end, we can appeal to the definition of Π_i, whereby

$$\Pi_i = \sum_k \sum_j \left[\left(1 - \frac{1}{\mu_k}\right) P_{ij,k} Q_{ij,k}\right] = \sum_k \sum_j \left[(\mu_k - 1) w_i L_{i,k}\right] = \sum_k \sum_j \left[(\mu_k - 1) \rho_{i,k}\right] w_i \overline{L}_i. \tag{18}$$

The second line follows from the fact that sales in each industry equal the wage bill times markup—i.e., $\sum_j P_{ij,k}Q_{ij,k} = \mu_k w_i L_{i,k}$, where $L_{i,k}$ denotes the number of workers employed by industry k. Accordingly, $\rho_{i,k} \equiv \frac{L_{i,k}}{\bar{L}_i}$ denotes the industry-level labor share. We can express industry-level labor shares in terms of industry-level revenue shares, by appealing to the following properties:

$$\begin{cases} \frac{\mu_k \rho_{i,k}}{\mu_g \rho_{i,g}} = \frac{r_{i,k}}{r_{i,g}} \\ \sum \rho_{i,k} = \sum r_{i,k} = 1 \end{cases} \implies \rho_{i,k} = \frac{\frac{1}{\mu_k} r_{i,k}}{\sum \frac{1}{\mu_g} r_{i,g}}.$$

Note that $\frac{\mu_k \rho_{i,k}}{\mu_g \rho_{i,g}} = \frac{r_{i,k}}{r_{i,g}}$ follows directly from the fact that (1) $\frac{r_{i,k}}{r_{i,g}} = \frac{\sum_j \left(P_{ij,k}Q_{ij,k}\right)}{\sum_j \left(P_{ij,g}Q_{ij,g}\right)}$, (2) $\frac{\rho_{i,k}}{\rho_{i,g}} = \frac{w_i L_{i,k}}{w_i L_{i,g}}$, and (3) $\sum_j \left(P_{ij,k}Q_{ij,k}\right) = \mu_k w_i L_{i,k}$. Plugging the above expression for $\rho_{i,k}$ into Equation 18 yields the following expression for Π_i as a function of revenue share, markups, and wage income:

$$\Pi_i = \sum_g \left[\left(\mu_g - 1\right)\rho_{i,g}\right] w_i \bar{L}_i = \frac{\sum_g \left(1 - \frac{1}{\mu_g}\right) r_{i,g}}{\sum_g \frac{1}{\mu_g} r_{i,g}} w_i \bar{L}_i = \left[\left(\sum_g \frac{r_{i,g}}{\mu_g}\right)^{-1} - 1\right] w_i \bar{L}_i.$$

Lastly, combining the above equation with Equation 17, we can arrive at the following expression for W_i:

$$W_i = \frac{\Pi_i + w_i \bar{L}_i}{P_i} = \frac{w_i \bar{L}_i}{P_i} \left(\sum_g \frac{r_{i,g}}{\mu_g}\right)^{-1} = \bar{T}_i \times \left(\prod_{k=1}^{K} \lambda_{ii,k}^{\frac{e_{i,k}}{1-\sigma_k}}\right) \left(\sum_{k=1}^{K} \frac{r_{i,k}}{\mu_k}\right)^{-1}.$$

C.2 Invariance of ToT to Restoring Marginal-Cost-Pricing

Recall from the main text that restoring efficiency in the global economy has two potential effects on welfare: (1) the pure gains from eliminating misallocation, and (2) terms-of-trade (ToT) effects that redistribute surplus from one country to another. State formally,

$$\log W_i^\star - \log W_i = \int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \boldsymbol{\mu}} \cdot \mathrm{d} \log \boldsymbol{\mu} + \underbrace{\int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \mathbf{w}} \cdot \mathrm{d} \log \mathbf{w}}_{\text{terms-of-trade effects}}.$$

Below we show that terms-of-trade, which are regulated by the wage vector $\mathbf{w} = \{w_i\}$, are invariant to restoring marginal cost pricing. That is, the second term on the right-hand side of the above equation is equal to zero. We establish this result using hat-algebra, by showing that $\hat{w}_i = w_i^\star / w_i = 1$ for all i. The vector $\{\hat{w}_i\}_i$ is regulated by the labor market clearing condition,

$$w_i \bar{L}_i = \sum_{k=1}^{K} \sum_{j=1}^{N} \left[\frac{1}{\mu_k} \lambda_{ij,k} e_{j,k} \pi_j w_j \bar{L}_j\right] = \sum_{k=1}^{K} \sum_{j=1}^{N} \left[\frac{\frac{1}{\mu_k}\left(\tau_{ij,k} w_i / \bar{\varphi}_{ij,k}\right)^{1-\sigma_k}}{\sum_{n=1}^{N}\left(\tau_{ni,k} w_n / \bar{\varphi}_{ni,k}\right)^{1-\sigma_k}} e_{j,k} \pi_j w_j \bar{L}_j\right],$$

where $\pi_j \equiv 1 + \frac{\Pi_j}{w_j L_j} \geq 1$ denotes the ratio of aggregate profits to wage payments in country j. After the restoring marginal cost pricing in each country, the labor market clearing condition can be solved in changes as

$$\hat{w}_i w_i \overline{L}_i = \sum_k \sum_j \left[\frac{1}{\mu_k} \frac{\lambda_{ij,k} \hat{w}_i^{1-\sigma_k}}{\sum_n \lambda_{nj,k} \hat{w}_n^{1-\sigma_k}} e_{j,k} \pi_j \hat{\pi}_j \hat{w}_j w_j \overline{L}_j \right], \tag{A}$$

where the change in the rent-to-wage ratios, $\hat{\pi}_j = \pi_j^\star / \pi_j$ solve the by the following system of equations for a given vector wage changes:

$$\pi_j \hat{\pi}_j = \frac{\sum_k \sum_j \left[\frac{\lambda_{ij,k} \hat{w}_i^{1-\sigma_k}}{\sum_n \lambda_{nj,k} \hat{w}_n^{1-\sigma_k}} e_{j,k} \pi_j \hat{\pi}_j \hat{w}_j w_j \overline{L}_j \right]}{\sum_k \sum_j \left[\frac{1}{\mu_k} \frac{\lambda_{ij,k} \hat{w}_i^{1-\sigma_k}}{\sum_n \lambda_{nj,k} \hat{w}_n^{1-\sigma_k}} e_{j,k} \pi_j \hat{\pi}_j \hat{w}_j w_j \overline{L}_j \right]}. \tag{B}$$

One can immediately check that $\{\hat{w}_i\} = \{\hat{\pi}_i\} = \{1\}$ is a solution to Equations (A) and (B). The intuition is that when transitioning to the efficient equilibrium, markups change uniformly for all suppliers in industry k. So, eliminating markups does not alter relative expenditure shares across suppliers *within* industries. Moreover, expenditure shares across industries are pinned to Cobb-Douglas weights and are, thus, invariant to markup elimination. Together, these features ensure that wages and the terms-of-trade are invariant to restoring marginal cost pricing. To state it formally,

$$\hat{\mathbf{w}} = \mathbf{1} \qquad \Longrightarrow \qquad \int_{\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \log \mathcal{W}_i}{\partial \log \mathbf{w}} \cdot \mathrm{d} \log \mathbf{w} = 0.$$

C.3 Deriving the Approximate Formula for $\Delta \mathcal{D}_i$

Applying Taylor's Theorem to $f(\mathbf{r}) = \log \mathbb{E}_{r_i} \left[\frac{1}{\mu} \right] = \log \sum_k \left[\frac{1}{\mu_k} r_{i,k} \right]$, we can derive the following approximation around $r_{i,k} \approx e_{i,k}$, which corresponds to a small deviation from autarky,

$$\log \mathbb{E}_{r_i} \left[\frac{1}{\mu} \right] \approx \log \mathbb{E}_{e_i} \left[\frac{1}{\mu} \right] + \sum_k \left[\frac{\frac{1}{\mu_k} (r_{i,k} - e_{i,k})}{\mathbb{E}_{e_i} \left[\frac{1}{\mu} \right]} \right].$$

We can use our notation for covariance to rewrite the above equations as

$$\log \mathbb{E}_{r_i} \left[\frac{1}{\mu} \right] - \log \mathbb{E}_{e_i} \left[\frac{1}{\mu} \right] \approx \bar{\mu}_i \, \mathrm{Cov}_{e_i}(\delta_i, \frac{1}{\mu}),$$

where, as before, $\frac{1}{\bar{\mu}_i} = \mathbb{E}_{e_i} \left[\frac{1}{\mu} \right]$ is a short-hand for the expenditure-weighted average markup in economy i.

C.4 $\Delta \mathcal{D}_i$ in Setups with Variable and Heterogeneous Markups

As argued in Section 5.2, our formula for $\Delta \mathcal{D}_i$ applies to can economy with variable and heterogeneous markups provided that consumer preferences fall under the homothetic class of preferences in

Arkolakis et al. (2019). The idea is that in such an environment, trade openness relocates resources across low- and high-markup industries but retains (a) the distribution of markups within industries, and (b) the relative sales (or labor shares) across firms from different markup brackets. Under (a) and (b) trade openness only impacts *inter-sectoral* misallocation but not *within-sectoral* misallocation. The proof of claim (a) is based on Arkolakis et al. (2019) and builds on the fact that preferences exhibit a choke price and the firm-level productivity distribution is Pareto in each origin with a common shape parameter. As an intermediate step, we can extrapolate from Arkolakis et al. (2019) that the markup charged in destination j by a firm in origin i–industry k with productivity φ is given by an industry-level function $m_k\left(\varphi/\varphi^*_{ij,k}\right)$, where $\varphi^*_{ij,k}$ is the zero-profit productivity cut-off as determined by the choke price. Importantly, $\mu_k\left(.\right)$ is strictly increasing, independent of the origin country, and independent of the underlying vector of trade costs, $\boldsymbol{\tau} \equiv \left\{\tau_{ij,k}\right\}_{i,j,k}$. The distribution of markups for goods sold from origin i to destination j in industry k is, accordingly,

$$\mathcal{M}_{ij,k}\left(\mu;\boldsymbol{\tau}\right) = \Pr\left\{m_k(\varphi/\varphi^\star_{ij,k}) \leq \mu \ \mid \ \varphi^\star_{ij,k} \leq \varphi\right\} = \frac{\Pr\left\{m_k(\varphi/\varphi^\star_{ij,k}) \leq \mu \ , \ \varphi^\star_{ij,k} \leq \varphi\right\}}{\Pr\left\{\varphi^\star_{ij,k} \leq \varphi\right\}},$$

where $\Pr\left\{.\right\}$ denotes probability and the last line follows from Bayes' rule. To evaluate the above probability expressions, note that $m_k(.)$ admits an inverse because it is strictly increasing and firm productivity distribution in origin i-industry k is Pareto, $G_{i,k}\left(\varphi\right) = 1 - b_{i,k}\varphi^{-\theta_k}$, with a shape parameter θ_k that is common across origin countries. The markup distribution can, accordingly, be expressed as

$$\mathcal{M}_{ij,k}\left(\mu;\boldsymbol{\tau}\right) = \frac{\int_{\varphi^\star_{ij,k}}^{\varphi^\star_{ij,k}m_k^{-1}(\mu)} dG_{i,k}\left(\varphi\right)}{\int_{\varphi^\star_{ij,k}}^{\infty} dG_{i,k}\left(\varphi\right)} = 1 - \left(m_k^{-1}\left(\mu\right)\right)^{-\theta_k} = \mathcal{M}_k(\mu).$$

Since $\left(\mu_k^{-1}\left(m\right)\right)^{-\theta_k}$ is independent of the origin country and the underlying vector of trade costs, one immediately concludes that the distribution of markups charged to the consumer price index in destination $j$is invariant to trade openness and origin, i.e., $\mathcal{M}_{ij,k}\left(\mu;\boldsymbol{\tau}\right) = \mathcal{M}_k(\mu)$. Next, we prove Claim (b) which follows from the fact that preferences are homothetic. Following the notation introduced in Section 5.2, the expenditure share on varieties with markup μ in a typical market is

$$\lambda_k(\mu) = \frac{\frac{\mu}{m_k^{-1}(\mu)}Q_k\left(\frac{\mu}{m_k^{-1}(\mu)}\right)}{\int_{\mu'=1}^{\infty}\left[\frac{\mu'}{m_k^{-1}(\mu')}Q_k\left(\frac{\mu}{m_k^{-1}(\mu)}\right)d\mathcal{M}_k(\mu')\right]}.$$

Since preferences are homothetic, Q_k is independent of nominal income and, thus, the level of trade costs, $\boldsymbol{\tau}$. Likewise and as detailed earlier, $m_k\left(.\right)$ and $\mathcal{M}_k\left(.\right)$ are both independent of $\boldsymbol{\tau}$. As such, $\lambda\left(\mu\right)$ is independent of $\boldsymbol{\tau}$ in each market. Finally, the invariance of $\lambda_k\left(\mu\right)$ and $\mathcal{M}_k(\mu)$ to $\boldsymbol{\tau}$ ensures that sales-weighted average markup in each location is invariant to trade liberalization.

Characterizing Counterfactual Welfare Changes. With non-CES preferences, we cannot specify welfare as the ratio of nominal income to the ideal price index. Instead, we appeal to the indirect utility function and relevant envelope conditions to characterize the change in welfare in response to markup correction. Let $V_i(Y_i, \mathbf{p}_i)$ denote the indirect utility function of country i, where $\mathbf{p}_i = \{p_{ni,k}(\omega)\}_{n,k,\omega}$ denotes the vector of all firm-level prices available to the consumer in country i. Suppose we counterfactually deflate all firm-0level markups in industry k by $\mathrm{d}\ln\mu_k$ per-cent. The price change for a generic variety is, then, $\mathrm{d}\ln p_{ni,k}(\omega) = \mathrm{d}\ln m_k + \mathrm{d}\ln w_n(\mathbf{m})$, where $\mathbf{m}$ denotes the vector of markups, which are counterfactually lowered from their status quo level $\mathbf{m} = \boldsymbol{\mu}$ to $\mathbf{m} = \mathbf{1}$. The correponding change in welfare, $W_i = V_i(Y_i, \mathbf{p}_i)$, is then given by

$$\mathrm{d}\ln W_i = \int_{\mathbf{m}=\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \ln V_i(.)}{\partial \ln Y_i} \mathrm{d}\ln Y_i(\mathbf{m}) - \sum_{n,k}\int_{\omega}\left(\int_{\mathbf{m}=\boldsymbol{\mu}}^{\mathbf{1}} \frac{\partial \ln V_i(.)}{\partial \ln p_{ni,k}}\left[\mathrm{d}\ln m_k + \mathrm{d}\ln w_i(\mathbf{m})\right]\right) d\omega.$$

Since aggregate profit margins are uniform across national origins, markup reduction preserves the vector of wages in the same spirit illustrated in Appendix C.2. The assumption that preferences are homothetic simplifies the above problem, as it ensures that $\frac{\partial \ln V_i(.)}{\partial \ln Y_i} = 1$ and drops out of the first integral. We can further simplify the above expression by appealing to Roy's Identity. Namely,

$$\frac{\partial V_i/\partial p_{ni,k}(\omega)}{\partial V_i/\partial Y_i} = -q_{ni,k}(\omega).$$

To simplify the notation, let $\lambda_{ni,k}(\omega) \times e_{i,k} = \frac{p_{ni,k}(\omega) q_{ni,k}(\omega)}{Y_i}$ denote the variety-specific expenditure share, where $e_{i,k}$ is the industry-level expenditure, pinned down by the Cobb-Douglas weight assigned to industry k. Applying Roy's identity and capitalizing on our choice of notation, we can simplify the expression for the change in welfare as

$$\Delta\ln W_i = \int_{\mathbf{m}=\boldsymbol{\mu}}^{\mathbf{1}} \mathrm{d}\ln Y_i(\mathbf{m}) - \sum_{k,n}\left(\int_{\mathbf{m}=\boldsymbol{\mu}}^{\mathbf{1}} e_{i,k} \times \left[\int_{\omega}\lambda_{ni,k}(\omega)\, d\omega\right] \mathrm{d}\ln m_k\right).$$

The above representation builds in the fact that $\mathrm{d}\ln m_k$is a uniform per-cent markup reduction that applies to all firms in industry k. Also, notice that $\int_{\omega}\lambda_{ni,k}(\omega)\, d\omega = 1$, by definition. Considering these points the above expression reduces to

$$\Delta\ln W_i = \Delta\ln Y_i - \sum_k\left[e_{i,k}\log\frac{1}{\mu_k}\right].$$

Recall that $\Delta\ln W_i = \ln W_i^\star - \ln W_i$ represents distance to the efficient frontier, which we denoted by $\mathcal{D}_i$. Relatedly, $\Delta\ln Y_i$ denotes the change in income after markup correction which is $\Delta\ln Y_i = \mathbb{E}_{e_i}\left[\ln\frac{1}{\mu}\right]$, following our earlier derivation. Combing these points, we get $\mathcal{D}_i = \ln\mathbb{E}_{r_i}\left(\frac{1}{\mu}\right) - \mathbb{E}_{e_i}\left[\ln\frac{1}{\mu}\right]$, which yields our baseline formula for the impact of trade on welfare cost of markups: $\Delta\mathcal{D}_i = \ln\mathbb{E}_{r_i}\left(\frac{1}{\mu}\right) - \ln\mathbb{E}_{e_i}\left(\frac{1}{\mu}\right)$.

C.5 $\Delta\mathcal{D}_i$ under Non-Markup Distortions

The definition of $\mathcal{D}_i$ (distance to the efficient frontier) is more nuanced under non-markup distortions. So, it is helpful to start with a formal definition of $\mathcal{D}_i$ under non-markup distortions before proceeding to our characterization of $\Delta\mathcal{D}_i$. Following Section 5.3, define non-markup distortions as price wedges that (unlike iceberg costs) do not shift factor demand and whose corresponding rents (unlike markups) leave the economy as a deadweight consumption loss. More specifically, let $\tilde{\Pi}_i$ denote the nominal rents associated with non-markup wedges, which are denoted by ψ_k. Social welfare in country i can be specified as

$$W_i = \frac{w_i L_i + \Pi_i + \tilde{\Pi}_i}{P_i} - \delta_i \frac{\tilde{\Pi}_i}{P_i}.$$

The first term is the indirect utility function from consumption. The second term is the deadweight loss from the dissipation of non-markup rents. $\delta_i \in [0,1]$ regulates the dissipation of non-markup rents and adjusts to policy interventions (e.g., financial reforms). Under the status quo, $\delta_i = 1$, in which case non-markup rents are competed away and entirely leave the economy as a deadweight consumption loss. Under the first-best policy intervention, $\delta_i = 0$, indicating that non-markup distortions are corrected at their sources and so is the dissipation of rents associated with them. To define $\mathcal{D}_i$, denote welfare as an explicit function of markups, non-markup wedges, and δ_i. Namely,

$$W_i = \mathcal{W}_i\left(\mu, \psi, \delta_i\right).$$

As shown at the end of this appendix, fixing $\delta_i = 1$ and setting markups to zero, the equilibrium allocation implied by non-markup wedges is contained-efficient. That is, no set of tax-cum-subsidies can improve allocative efficiency. This need not be the case if non-markup wedges distort relative input prices à la Liu, 2019. Here, the allocative efficiency gains from correcting non-markup wedges stem purely from non-price intervention that correct distortion at their source, thereby preventing rent dissipation. More formally, distance to the efficient frontier is defined as the gains from jointly (a) eliminating markups, *i.e.*, setting μ to 1, and (b) eliminating non-markup wedges at their source, *i.e.*, setting ψ to 1 and δ to zero:

$$\mathcal{D}_i = \log \mathcal{W}_i\left(1, 1, 0\right) - \log \mathcal{W}_i\left(\mu, \psi, 1\right).$$

Next, we specify $\mathcal{D}_i = \log \hat{W}_i$, where $\hat{W}_i = W_i^\star / W_i$ with the superscript $\star$ denoting the unconstrained-efficient allocation (i,.e., zero markups, zero non-markup wedges, and $\delta_i = 0$). We derive our expression for $\log \hat{W}_i = \log \hat{Y}_i - \log \hat{P}_i$, by characterizing the change in nominal income and the consumer price index. With a slight abuse of notation, we henceforth treat $Y_i = w_i L_i + \Pi_i + \tilde{\Pi}_i - \delta_i \tilde{\Pi}_i$ as nominal income. Notice that by definition, $\Pi_i^\star = \tilde{\Pi}_i^\star$. Also, under the status quo where $\delta_i = 1$, non-markup wedges do not contribute to net income, i.e., $\tilde{\Pi} - \delta_i \tilde{\Pi}_i = 0$. The change in income after restoring efficiency is this characterized by

$$\hat{Y}_i = \frac{Y_i^\star}{Y_i} = \frac{w_i^\star L_i}{w_i L_i + \Pi_i}.$$

Treating w_i as the numeraire and following the logic outlined in Appendix A, we derive a simplified expression for the change in nominal income, Y_i. Namely,

$$\log \hat{Y}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right].$$

Here, the change in the consumer price index encompasses the joint impact of removing markups (μ) and non-markup wedges (ψ), which is

$$\log \hat{P}_i = -\mathbb{E}_{e_i}\left[\log\left(\mu\psi\right)\right] = \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \psi\right].$$

Combining the expression for $\log \hat{Y}_i$ and $\log \hat{P}_i$, yields the following expression for distance to the (first-best) efficient frontier:

$$\mathcal{D}_i = \left(\log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right]\right) + \mathbb{E}_{e_i}\left[\log \psi\right].$$

Under autarky, where $r_{i,k}^A = e_{i,k}$, we get $\mathcal{D}_i^A = \left(\log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right]\right) + \mathbb{E}_{e_i}\left[\log \psi\right]$, which immediately implies

$$\Delta\mathcal{D}_i = \mathcal{D}_i - \mathcal{D}_i^A = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right].$$

This is the same exact expression as that specified under Proposition 1 for the case of our baseline model that featured only markup distortions.

The Constricted-Efficiency of Non-Rent-Generating Wedges. Below, we prove that wedges whose corresponding rents leave the economy are constrained -efficient. That fixing $\delta_i = 1$, there is no vector of prices (or taxes) that can improve allocative efficiency. We show this in a more general environment with arbitrary preferences in which utility from consumption is specified by a non-parametric indirect utility function $V_i\left(Y_i, \tilde{\mathbf{P}}_i\right)$ where $\tilde{\mathbf{P}}_i \equiv \left\{\tilde{P}_{i,1}, ..., \tilde{P}_{i,K}\right\}$ denotes the vector of tax-inclusive prices which are chosen by the government. As before, $P_{i,k}$ denotes the pre-tax price level set by the producer. We intent to prove that –in a closed economy i– the prices that maximize welfare coincide with producer producer prices, i.e., $\tilde{\mathbf{P}}_i = \mathbf{P}_i$, which indicates that the market allocation is constrained-efficient. Importantly, this will not be true if rents were not competed away. To proof our claim we must write the first-order conditions associated with

$$\max_{\tilde{\mathbf{P}}_i}\; W_i\left(\tilde{\mathbf{P}}_i\right) = V_i\left(Y_i\left(\tilde{\mathbf{P}}_i\right), \tilde{\mathbf{P}}_i\right) - \delta_i \frac{\partial V_i\left(.\right)}{\partial Y_i} \tilde{\Pi}_i,$$

where $Y_i = w_i L_i + \Pi_i + \left(\tilde{\mathbf{P}}_i - \mathbf{P}_i\right) \cdot \mathbf{Q}_i$ with the last term representing the revenue associate with choice $\tilde{\mathbf{P}}_i$. The term $\frac{\partial V_i(.)}{\partial Y_i}$ which can be interpreted as the the inverse price index, converts the dissipation of nominal rents to a loss in real welfare—consistent with our main CES model. The first-order conditions associated with the above problem can be written as

$$\frac{\partial W_i}{\partial \tilde{\mathbf{P}}_i} = \frac{\partial V_i\left(.\right)}{\partial Y_i}\left(\frac{\partial w_i L_i}{\partial \tilde{\mathbf{P}}_i} + \frac{\partial \Pi_i}{\partial \tilde{\mathbf{P}}_i} + \frac{\partial}{\partial \tilde{\mathbf{P}}_i}\left\{\left(\tilde{\mathbf{P}}_i - \mathbf{P}_i\right) \cdot \mathbf{Q}_i\right\}\right) + \frac{\partial V_i\left(.\right)}{\partial \tilde{\mathbf{P}}_i} - \delta_i \frac{\partial \Pi_i}{\partial \tilde{\mathbf{P}}_i} = 0.$$

Appealing to Roy's identity and treating w_i as the numeraire, simplifies the first-order condition as follows:

$$\frac{\partial W_i}{\partial \tilde{\mathbf{P}}_i} = \frac{\partial V_i\left(.\right)}{\partial Y_i}\left[\left(\tilde{\mathbf{P}}_i - \mathbf{P}_i\right) \cdot \frac{\partial \mathbf{Q}_i}{\partial \tilde{\mathbf{P}}_i} + \left(1 - \delta_i\right)\frac{\partial \Pi_i}{\partial \tilde{\mathbf{P}}_i}\right] = 0.$$

Setting $\delta_i = 1$ implies that the optimal price is equal to the equilibrium price: $\tilde{\mathbf{P}}_i = \mathbf{P}_i$. In other words, the equilibrium allocation is constrained-efficient subject to the full dissipation of distortion rents (i.e.,$\delta_i = 1$).

D CES Preferences across Industries

D.1 The Cost of Markups in a Closed Economy (Lemma 2)

The idea of the proof closely resembles that of our baseline Lemma 1. First, it is straightforward to check that the efficient allocation is obtainable under marginal cost-pricing, irrespective of the cross-industry utility aggregator. Next, suppose country i was operating under autarky. Extrapolating from Appendix A and treating w_i as the numeraire, the change in income after restoring marginal cost-pricing is

$$\hat{Y}_i^A = \frac{\Pi_i^A + w_i^A L_i}{w_i^A L_i} = \sum_k \left[\frac{1}{\mu_k} e_{i,k}^A\right] = \mathbb{E}_{e_i^A}\left[\frac{1}{\mu}\right]$$

where e_i^A corresponds to the autarky expenditure share on industry k in country i. Notice, the autarky expenditure share is strictly different from the factual expenditure share under CES preferences, i.e., $e_{i,k}^A \neq e_{i,k}$. We can, however, infer autarky expenditure shares from the factual expenditure share values using exact hat-algebra. First, it is straightforward to check that the change in industry-level expenditure shares if we shut down trade is

$$\hat{e}_{i,k} \equiv \frac{e_{i,k}^A}{e_{i,k}} = \frac{e_{i,k}\hat{P}_{i,k}^{1-\eta}}{\sum_g e_{i,g}\hat{P}_{i,g}^{1-\eta}},$$

where $\hat{P}_{i,k} = P_{i,k}^A / P_{i,k}$ is the change in industry k's price index after shutting down trade. Following Arkolakis et al. (2012), we know that $\hat{w}_i / \hat{P}_{i,k} = \lambda_{ii,k}^{-\frac{1}{1-\sigma_k}}$, where $\lambda_{ii,k}$ is the domestic expenditure share in industry k under the status quo. Rearranging the aforementioned expression delivers $\hat{P}_{i,k} = \hat{w}_i \lambda_{ii,k}^{\frac{1}{1-\sigma_k}}$.

Plugging the expression for $\hat{P}_{i,k}$ into the equation describing $\hat{e}_{i,k}$, yields

$$e_{i,k}^{A}=\frac{e_{i,k}\hat{w}_{i}\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}}{\sum_{g}e_{i,g}\hat{w}_{i}\lambda_{ii,g}^{\frac{1-\eta}{1-\sigma_{g}}}}=\frac{e_{i,k}\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}}{\sum_{g}e_{i,g}\lambda_{ii,g}^{\frac{1-\eta}{1-\sigma_{g}}}}=\frac{e_{i,k}\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}}{\mathbb{E}_{e_{i}}\left[\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right]}.$$

Stated verbally, we can infer the autarky expenditure share on industry k from the vector of factual industry-level expenditure shares $e_{i,k}$, domestic expenditure shares $\lambda_{ii,k}$, and substitution elasticities, σ_k and η. Plugging the above expression for $e_{i,k}^{A}$ in the our original expression for $\hat{Y}_{i}^{A}$, delivers the following expression

$$\hat{Y}_{i}^{A}=\sum_{k}\left[\frac{1}{\mu_{k}}\frac{e_{i,k}\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}}{\mathbb{E}_{e_{i}}\left[\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right]}\right]=\mathbb{E}_{e_{i}}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\right]$$

where $\tilde{\lambda}_{ii,k}$ is the normalized domestic expenditure share for each k, which is defined as $\tilde{\lambda}_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}\equiv\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}/\mathbb{E}_{e_{i}}\left[\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right]$. Following the same logic, the change in the consumer price index after restoring marginal cost pricing is given by

$$\hat{P}_{i}^{A}=\left[\sum_{k}\left(\frac{1}{\mu_{k}}\right)^{1-\eta}e_{i,k}^{A}\right]^{\frac{1}{1-\eta}}=\left[\sum_{k}\left(\frac{1}{\mu_{k}}\right)^{1-\eta}\frac{e_{i,k}\lambda_{ii,k}^{\frac{1-\eta}{1-\sigma_{k}}}}{\mathbb{E}_{e_{i}}\left[\lambda_{ii}^{\frac{1-\eta}{1-\sigma}}\right]}\right]^{\frac{1}{1-\eta}}=\left(\mathbb{E}_{e_{i}}\left[\frac{1}{\mu}\tilde{\lambda}_{ii,k}^{\frac{1}{1-\sigma}}\right]^{1-\eta}\right)^{\frac{1}{1-\eta}}.$$

Plugging the expressions for $\hat{Y}_{i}^{A}$ and $\hat{P}_{i}^{A}$ into $\mathcal{D}_{i}^{A}=\log\hat{Y}_{i}-\log\hat{P}_{i}^{A}$, we obtain an update expression for the autarky degree of misallocation under CES preferences

$$\mathcal{D}_{i}^{A}=\log\mathbb{E}_{e_{i}}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\right]-\frac{1}{1-\eta}\log\mathbb{E}_{e_{i}}\left[\left(\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1}{1-\sigma}}\right)^{1-\eta}\right].$$

Deriving the Approximate Formula for $\mathcal{D}_{i}^{A}$. Define the function $f\left(x_{1}...,x_{K}\right)=\frac{1}{1-\eta}\log\mathbb{E}_{\omega}\left[x^{1-\eta}\right]=\frac{1}{1-\eta}\log\sum_{k}\left[\omega_{k}x_{k}^{1-\eta}\right]$. Our goal is to derive the Taylor expression for $f(x_1...,x_K)$ around $\bar{\mathbf{x}}=\left(\mathbb{E}_{\omega}\left[x\right],...,\mathbb{E}_{\omega}\left[x\right]\right)$. For this we appeal to the following first- and second-order derivative of function $f(.)$,

$$\frac{\partial f}{\partial x_{k}}=\frac{\omega_{k}x_{k}^{-\eta}}{\sum_{k'}\omega_{k'}x_{k'}^{1-\eta}};\qquad\frac{\partial^{2}f}{\partial x_{k}\partial x_{g}}=\frac{-\eta\omega_{k}x_{k}^{-\eta-1}}{\sum_{k'}\omega_{k'}x_{k'}^{1-\eta}}\times1_{g=k}-\frac{(1-\eta)\omega_{k}\omega_{g}x_{k}^{-\eta-1}x_{g}^{-\eta}}{\left(\sum_{k'}\omega_{k'}x_{k'}^{1-\eta}\right)^{2}}.$$

Evaluating the above derivatives at $\bar{\mathbf{x}}=\left(\mathbb{E}_{\omega}\left[x\right],...,\mathbb{E}_{\omega}\left[x\right]\right)$, we can obtain the following second-order approximation for $f=\frac{1}{1-\eta}\log\mathbb{E}_{\omega}\left[x^{1-\eta}\right]$:

$$f(x_{1}...,x_{K})\approx f(\bar{\mathbf{x}})+\sum_{k}\left[\frac{\partial f\left(\bar{\mathbf{x}}\right)}{\partial x_{k}}\left(x_{k}-\mathbb{E}_{\omega}\left[x\right]\right)\right]+\frac{1}{2}\sum_{k}\sum_{g}\left[\frac{\partial^{2}f\left(\bar{\mathbf{x}}\right)}{\partial x_{k}\partial x_{g}}\left(x_{k}-\mathbb{E}_{\omega}\left[x\right]\right)\left(x_{k}-\mathbb{E}_{\omega}\left[x\right]\right)\right]$$

It is straightforward to check that the second term on the right-hand side is equal to zero

$$\sum_k \left[\frac{\partial f(\bar{\mathbf{x}})}{\partial x_k}(x_k - \mathbb{E}_\omega[x])\right] = \frac{\mathbb{E}_\omega[x]^{-\eta}}{\sum_k \omega_k \mathbb{E}_\omega[x]^{1-\eta}} \sum_k [\omega_k (x_k - \mathbb{E}_\omega[x])] = \frac{1}{\mathbb{E}_\omega[x]}(\mathbb{E}_\omega[x] - \mathbb{E}_\omega[x]) = 0.$$

Likewise the last term on the right-hand side can be simplified as

$$\begin{aligned}\sum_k \sum_g \left[\frac{\partial^2 f(\bar{\mathbf{x}})}{\partial x_k \partial x_g}(x_k - \mathbb{E}_\omega[x])(x_k - \mathbb{E}_\omega[x])\right] &= \frac{1-\eta}{2\mathbb{E}_{e_i}[x]} \sum_k \omega_k (x_k - \mathbb{E}_\omega[x]) \sum_g [\omega_g (x_k - \mathbb{E}_\omega[x])] - \frac{\eta}{2\mathbb{E}_{e_i}[x]^2} \sum_k \left[\omega_k (x_k - \mathbb{E}_\omega[x])^2\right] \\ &= \frac{1-\eta}{2\mathbb{E}_\omega[x]}(\mathbb{E}_\omega[x] - \mathbb{E}_\omega[x])(\mathbb{E}_\omega[x] - \mathbb{E}_\omega[x]) - \frac{\eta}{2}\frac{\mathrm{Var}_\omega(x)}{\mathbb{E}_\omega[x]^2} = -\frac{\eta}{2}[\mathrm{CV}_\omega(x)]^2.\end{aligned}$$

Plugging the above expressions back into our Taylor approximation for $f = \frac{1}{1-\eta}\log \mathbb{E}_\omega\left[x^{1-\eta}\right]$ and setting $x = \frac{1}{\mu}$ and $\omega = e_i^A$, we obtain

$$\frac{1}{1-\eta}\log \mathbb{E}_{e_i^A}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right] \approx \log \mathbb{E}_{e_i^A}\left[\frac{1}{\mu}\right] - \frac{\eta}{2}\left[\mathrm{CV}_{e_i^A}\left(\frac{1}{\mu}\right)\right]^2.$$

Noting that $\mathcal{D}_i^A = \log \mathbb{E}_{e_i^A}\left[\frac{1}{\mu}\right] - \frac{1}{1-\eta}\log \mathbb{E}_{e_i^A}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]$, we immediately arrive at the following approximation for the autarky degree of misallocation in economy i

$$\mathcal{D}_i^A \approx \frac{\eta}{2} \times \left[\mathrm{CV}_{e_i^A}(\frac{1}{\mu})\right]^2 = \frac{\eta}{2} \times \left[\mathrm{CV}_{e_i}(\frac{1}{\mu}\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}})\right]^2,$$

where the last line follows from our previous observation that $e_{i,k}^A = e_{i,k}\tilde{\lambda}_{ii,k}^{\frac{1-\eta}{1-\sigma_k}}$ for all k.

D.2 The Impact of Trade on the Welfare Cost of Markups (Proposition 2)

To characterize impact of trade on misallocation, we follow the same logic underlying the proof of Proposition 1. We first determine the degree of misallocation in an open economy, which is welfare distance from the globally efficient equilibrium under which marginal cost-pricing is restored universally in all countries and industries. Specifically, letting $\star$ denote the globally efficient equilibrium, $\mathcal{D}_i = \log \hat{Y}_i - \log \hat{P}_i$, where $\hat{Y}_i = Y_i^\star / Y_i$ and $\hat{P}_i = P_i^\star / P_i$. The change in open economy i's consumer price index after restoring marginal cost pricing is given by

$$\hat{P}_i = \left[\sum_k e_{i,k}\left(\frac{1}{\mu_k}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}} = \mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}.$$

Note that above equation differs from $\hat{P}_i^A$ in that it depends on the factual industry-level expenditure shares, $e_{i,k}$, rather than the counterfactual autarky expenditure shares. Likewise the change in income is

$$\hat{Y}_i = \sum_k \left[\frac{1}{\mu_k} r_{i,k}\right] = \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right],$$

where $r_{i,k}$ denotes the industry-level revenue share under the status quo. Capitalizing on the expressions for $\hat{P}_i$ and $\hat{Y}_i$ and can calculate the impact of trade on misallocation as $\Delta\mathcal{D}_i = \log\left(\frac{\hat{Y}_i}{\hat{Y}_i^A}\right) - \log\left(\frac{\hat{P}_i}{\hat{P}_i^A}\right)$. Specifically, appealing to the previously-derived expressions for $\hat{P}_i^A$ and $\hat{Y}_i^A$, we get

$$\frac{\hat{P}_i}{\hat{P}_i^A} = \frac{\mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}}{\mathbb{E}_{e_i}\left[\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\left(\frac{1}{\mu}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}}; \qquad \frac{\hat{Y}_i}{\hat{Y}_i^A} = \frac{\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right]}{\mathbb{E}_{e_i}\left[\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\frac{1}{\mu}\right]}.$$

Plugging the above two equations into $\Delta\mathcal{D}_i = \log\left(\frac{\hat{Y}_i}{\hat{Y}_i^A}\right) - \log\left(\frac{\hat{P}_i}{\hat{P}_i^A}\right)$, yields the formula presented under Proposition 2:

$$\Delta\mathcal{D}_i = \log\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log\mathbb{E}_{e_i}\left[\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\frac{1}{\mu}\right] - \frac{1}{1-\eta}\log\left(\frac{\mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]}{\mathbb{E}_{e_i}\left[\tilde{\lambda}_{ii}^{\frac{1-\eta}{1-\sigma}}\left(\frac{1}{\mu}\right)^{1-\eta}\right]}\right).$$

E Cost of Markups under Input-Output Linkages

E.1 Welfare Cost of Markups in a Closed Economy (Lemma 3)

As before, let superscript $\star$ denote the globally efficient equilibrium and A denote autarky. The national-level degree of misallocation for a closed economy can be calculated as $\mathcal{D}_i^A = \log W_i^{A\star} - \log W_i^A$, where W denotes welfare as measured by real consumption. Since consumption income in country i equals $w_i^A L_i + \Pi_i^A$, we can express welfare in closed economy i as

$$W_i^A = \frac{w_i^A L_i + \Pi_i^A}{P_i^A} = \pi_i^A \frac{w_i^A L_i}{P_i^A},$$

where $\pi_i^A \equiv 1 + \frac{\Pi_i^A}{w_i^A L_i}$ denotes the average profit margin in closed economy i. Use the hat notation, $\widehat{x} = x^*/x$ to describe the change in in a generic variable x, alter restoring efficiency. The degree of misallocation in a closed economy i can, accordingly, we specified as

$$\mathcal{D}_i^A = \log W_i^{A\star} - \log W_i^A = \log\left(\widehat{\pi}_i^A \widehat{w}_i^A\right) - \log\left(\widehat{P}_i^A\right). \tag{19}$$

Since we are dealing with a closed economy, we can normalize country i's wage rate by choice of numeraire, which ensures that $\widehat{w}_i = 1$. Recall from Section 6.2 that

$$\log\hat{P}_i = \underbrace{-\sum_k\left[\beta_{i,k}\log\mu_k\right]}_{\mathbb{E}_{\beta_i}\left[\log\frac{1}{\mu}\right]}, \quad \text{where} \quad \beta_{i,k} \equiv \sum_g\left[e_{i,g}a_{i,gk}\right]. \tag{20}$$

where $a_{i,gk}$ denotes the (g,k) entry of economy i's inverse Leontief matrix and $e_{i,g}$ denotes the net (i.e., final good) expenditure share. What remains to be characterized in Equation 19 is the change in the economy-wide profit margins, $\widehat{\pi}_i^A$.To this end, we first specify π_i^A as a function industry-level markups and shares. Recalling that $C_{i,k}$ denotes total payment to production inputs in origin i–industry k, we can write aggregate profits in a (closed or open) economy i as

$$\Pi_i = \sum_k [(\mu_k - 1)C_{i,k}]$$

Given our assumed production structure, total input costs are related to wage payments as $C_{i,k} = w_i L_{i,k}/\nu_{i,k}$, where $\nu_{i,k}$ denotes the value-added share in origin i–industry k. We can, thus, rewrite total profits as a function of labor shares, $\rho_{i,k} = L_{i,k}/\overline{L}_i$, and value-added shares as

$$\Pi_i = \sum_k \left[\frac{\mu_k - 1}{\nu_{i,k}} \frac{w_i L_{i,k}}{w_i \overline{L}_i}\right] w_i L_i = \sum_k \left[\frac{\mu_k - 1}{\nu_{i,k}} \rho_{i,k}\right] w_i \overline{L}_i. \tag{21}$$

Next, we need to write the labor shares as a function of revenue shares, $r_{i,k} = \frac{\sum_n P_{in,k} Q_{in,k}}{\sum_g \sum_n P_{in,g} Q_{in,g}}$. This step relies on the observations that revenue shares, by definition, are related to to labor shares as

$$\frac{\rho_{i,k}}{\rho_{i,g}} = \frac{\frac{\nu_{i,k}}{\mu_k}}{\frac{\nu_{i,k}}{\mu_g}} \frac{r_{i,k}}{r_{i,g}} \implies \rho_{i,k} = \frac{\frac{\nu_{i,k}}{\mu_k} r_{i,k}}{\sum_g \frac{\nu_{i,k}}{\mu_g} r_{i,g}}.$$

The second line invokes the accounting property that labor and revenue shares add up to one: $\sum_k \rho_{i,k} = \sum_k r_{i,k} = 1$. Plugging the above expression for $\rho_{i,k}$ back into Equation 21 yields

$$\Pi_i = \frac{\sum_k \left[(\mu_k - 1) \frac{r_{i,k}}{\mu_k}\right]}{\sum_k \left[\nu_{i,k} \frac{r_{i,k}}{\mu_k}\right]} w_i \overline{L}_i \qquad \implies \qquad \pi_i \equiv 1 + \frac{\Pi_i}{w_i \overline{L}_i} = \frac{\sum_k \left[(\mu_k + \nu_{i,k} - 1) \frac{r_{i,k}}{\mu_k}\right]}{\sum_k \left[\nu_{i,k} \frac{r_{i,k}}{\mu_k}\right]} = \frac{1 - \sum_k \left[\frac{1-\nu_{i,k}}{\mu_k} r_{i,k}\right]}{\sum_k \left[\frac{\nu_{i,k}}{\mu_k} r_{i,k}\right]}.$$

The above equation can be used to characterize $\widehat{\pi}_i$ by appealing to two observations: First, under autarky, each industry's gross revenue share should equal its gross expenditure share, i.e., $r_{i,k}^A = \tilde{e}_{i,k}^A$. Second, profits margins are zero under the efficient equilibrium, i.e., $\pi^{A\star} = 0$. Considering this, the above expression for π_i implies the following change in profit margins after restoring efficiency:

$$\widehat{\pi}_i^A = \frac{\pi_i^{A\star}}{\pi_i^A} = \frac{1 - \sum_k \left[\frac{1-\nu_{i,k}}{\mu_k} \tilde{e}_{i,k}^A\right]}{\sum_k \left[\frac{\nu_{i,k}}{\mu_k} \tilde{e}_{i,k}^A\right]} = \frac{1 - \mathbb{E}_{\tilde{e}_i^A}\left[\frac{1-\nu_i}{\mu}\right]}{\mathbb{E}_{\tilde{e}_i^A}\left[\frac{\nu_i}{\mu}\right]}. \tag{22}$$

Plugging Equations 20 and 22 back into our initial expression for $\mathcal{D}_i^A$ (Equation 19) yields

$$\mathcal{D}_i^A = \mathbb{E}_{\beta_i}\left[\log \frac{1}{\mu}\right] - \log\left(\frac{\mathbb{E}_{\tilde{e}_i^A}\left[\frac{\nu_i}{\mu}\right]}{1 - \mathbb{E}_{\tilde{e}_i^A}\left[\frac{1-\nu_i}{\mu}\right]}\right).$$

E.2 The Impact of Trade on the Cost of Markups (Proposition 3)

It is straightforward to check that the logic of Proposition 1 extends to an economy with input-output linkages, at least to a first-order approximation. That is, for a common input-output structure, restoring marginal cost pricing retains the age vector, **w**, and the terms-of-trade. Hence, letting superscript $\star$ denote the globally efficient equilibrium, the national-level degree of misallocation can be calculated as $\mathcal{D}_i = \log W_i^\star - \log W_i$. With this background, we wish to characterize $\Delta\mathcal{D}_i$, which is the change in misallocation as a result of trade engagement. Since consumption income in country i equals $w_i L_i + \Pi_i$, we can express welfare in country i as

$$W_i = \frac{w_i \overline{L}_i + \Pi_i}{P_i} = \pi_i \frac{w_i \overline{L}_i}{P_i},$$

where $\pi_i \equiv 1 + \frac{\Pi_i}{w_i \overline{L}_i}$ denotes the average profit margin in economy i. As before, the effect of trade on sectoral misallocation can be calculated as

$$\log \widehat{W}_i = \log\left(\widehat{w}_i / \widehat{P}_i\right) + \log\left(\widehat{\pi}_i\right). \tag{23}$$

Since industry-level markups are invariant to trade openness, it is straightforward to check that the expression for $\widehat{w}_i / \widehat{P}_i$ is the same with and without sectoral markup heterogeneity. Extrapolating the approach in CRC2014, we can derive the following relationship between real production cost, $C_{i,k}/P_{i,k}$, and the domestic expenditure share, $\lambda_{ii,k}$,

$$\lambda_{ii,k} = \left(\frac{P_{ii,k}}{P_{i,k}}\right)^{1-\sigma_k} = \left(\frac{\mu_k \tau_{ii,k} C_{i,k}}{P_{i,k}}\right)^{1-\sigma_k} \qquad \Longrightarrow \qquad \frac{\widehat{C}_{i,k}}{\widehat{P}_{i,k}} = \lambda_{ii,k}^{-\frac{1}{\sigma_k - 1}}.$$

Appealing to the expression for $C_{i,k}$ we can derive an equation relating the real real wage in each industry to real production cost. Namely,

$$\widehat{C}_{i,k} = \widehat{w}_i^{\nu_{i,k}} \prod_g \widehat{P}_{i,g}^{(1-\nu_{i,k})\alpha_{i,gk}} \qquad \Longrightarrow \qquad = \log \frac{\widehat{w}_i}{P_{i,k}} = \log \frac{\widehat{C}_{i,k}}{P_{i,k}} + \sum_g \left[(1-\nu_{i,k})\,\alpha_{i,gk} \log \frac{\widehat{w}_i}{P_{i,g}}\right]$$

The last line in the above equation specifies a system of equations, which can be inverted to characterize $\widehat{w}_i / \widehat{P}_{i,k}$ as a function of $\left\{\widehat{C}_{i,g}/\widehat{P}_{i,g}\right\}_g$. Doing so yields,

$$\frac{\widehat{w}_i}{\widehat{P}_{i,k}} = \prod_g \left(\frac{\widehat{C}_{i,g}}{P_{i,g}}\right)^{a_{i,kg}} = \prod_g \left(\lambda_{ii,g}^{\frac{1}{1-\sigma_k}}\right)^{a_{i,kg}} \qquad \Longrightarrow \qquad \frac{\widehat{w}_i}{\widehat{P}_i} = \prod_k \prod_g \left(\lambda_{ii,g}^{\frac{a_{i,kg} e_{i,k}}{1-\sigma_k}}\right).$$

where $a_{i,kg}$ is the (k,g) entry of economy i's inverse Leontief matrix and the last line follows from the fact that $\widehat{P}_i = \prod_k \widehat{P}_{i,k}^{e_{i,k}}$. The last line clearly indicates that $\widehat{w}_i / \widehat{P}_i$ is independent of the underlying vector of markups. Accordingly, $\log\left(\widehat{w}_i / \widehat{P}_i\right)$ on the right-hand side of Equation 23 corresponds to the gains from trade in an efficient economy. That is, $\log\left(\widehat{w}_i / \widehat{P}_i\right) = \log W_i^\star - \log W_i^{\star A}$, which immediately implies that $\Delta\mathcal{D}_i = \log(\widehat{\pi}_i)$. Considering this intermediate point, our goal herein is to derive a formula for $\log(\widehat{\pi}_i)$ under input-output linkages. Recall from earlier that the profits margin in economy i is

$$\pi_i \equiv 1 + \frac{\Pi_i}{w_i \bar{L}_i} = \frac{1 - \sum_k \left[\frac{1-\nu_{i,k}}{\mu_k} r_{i,k}\right]}{\sum_k \left[\frac{\nu_{i,k}}{\mu_k} r_{i,k}\right]}.$$

Note that under autarky the gross revenue and expenditure shares are exactly the same, i.e., $r_{i,k}^A = e_{i,k}^A$. We can, thus, produce the following expression for $\widehat{\pi}_i = \pi_i / \pi_i^A$,

$$\widehat{\pi}_i = \frac{\left(1 - \sum_k \frac{1-\nu_{i,k}}{\mu_k} r_{i,k}\right) \sum_k \left[\frac{\nu_{i,k}}{\mu_k} e_{i,k}^A\right]}{\left(1 - \sum_k \frac{1-\nu_{i,k}}{\mu_k} e_{i,k}^A\right) \sum_k \left[\frac{\nu_{i,k}}{\mu_k} r_{i,k}\right]}.$$

Taking logs from the above equations and using the expectation notation introducer earlier, we arrive that following expression for the effect of trade on misallocation as specified by Proposition 3

$$\Delta\mathcal{D}_i = \log(\widehat{\pi}_i) = \log\left(\frac{\mathbb{E}_{\tilde{e}_i^A}\left[\frac{\nu_i}{\mu}\right]}{1 - \mathbb{E}_{\tilde{e}_i^A}\left[\frac{1-\nu_i}{\mu}\right]}\right) - \log\left(\frac{\mathbb{E}_{r_i}\left[\frac{\nu_i}{\mu}\right]}{1 - \mathbb{E}_{r_i}\left[\frac{1-\nu_i}{\mu}\right]}\right).$$

E.3 Inferring Autarky Gross Expenditure Shares from Observable Shares

The formulas for $\mathcal{D}_i^A$ and $\Delta\mathcal{D}_i$ depend on gross expenditure shares under autarky, $\tilde{e}_{i,k}^A$. So, to complete the proofs of Proposition 2 and Lemma 2 , we must characterize $\tilde{e}_{i,k}$ as a function of observables. To this end, we use two relationships: First, that under autarky, $\mu_k C_{i,k}^A = E_{i,k}^A$—that is, total revenues in industry k, which are $\mu_k C_{i,k}^A$, equal total expenditure on that industry, $E_{i,k}^A$, under autarky. Second, the accounting identity,

$$E_{i,k} = e_{i,k} Y_i + \sum_g \left[(1-\nu_{i,g})\alpha_{i,kg} C_{i,g}\right],$$

which states that gross expenditure on industry k goods is the sum of consumption plus input expenditure. We can combine these two relationships to produce the following expression for gross expenditure under autarky

$$\tilde{e}_{i,k}^A \equiv \frac{E_{i,k}^A}{E_i^A} = \frac{e_{i,k} Y_i^A + \sum_g \left[(1-\nu_{i,g})\alpha_{i,kg} C_{i,g}^A\right]}{E_i^A} \implies \tilde{e}_{i,k}^A = e_{i,k} \frac{Y_i^A}{E_i^A} + \sum_g \left[\frac{(1-\nu_{i,g})\alpha_{i,kg}}{\mu_g} \tilde{e}_{i,g}^A\right]. \quad (24)$$

We can, furthermore, express Y_i^A / E_i^A (in the above equation) in terms the vector of gross expenditures shares $\left\{\tilde{e}_{i,k}^A\right\}_k$ and observables. For this, we use the accounting identity, $Y_i = E_i - \sum_k \left[(1-\nu_{i,k}) C_{i,k}\right]$, which states that aggregate consumption expenditure across all industries equals aggregate gross expenditure minus input expenditure. Plugging $C^A_{i,k} = E^A_{i,k}/\mu_k$ in the aforementioned identity yields

$$\frac{Y^A_i}{E^A_i} = 1 - \sum_k \left[\frac{1-\nu_{i,k}}{\mu_k}\tilde{e}^A_{i,k}\right].$$

Plugging the above expression back into Equation 24 deliveries the following equation which implicitly characterizes $\tilde{e}^A_{i,k}$ as a function of parameters and observable shares,

$$\tilde{e}^A_{i,k} = e_{i,k}\left[1 - \sum_g \left(\frac{1-\nu_{i,g}}{\mu_g}\tilde{e}^A_{i,g}\right)\right] + (1-\nu_{i,k})\sum_g \left[\frac{\alpha_{i,kg}}{\mu_g}\tilde{e}^A_{i,g}\right] = e_{i,k} + \sum_g \left[\frac{1-\nu_{i,g}}{\mu_g}\left(\alpha_{i,kg} - e_{i,k}\right)\tilde{e}^A_{i,g}\right].$$

We can write the above system of equations in matrix notation as

$$\mathbf{I}\tilde{\mathbf{e}}^A_i = \mathbf{e}_i + \mathbf{\Phi}_i\left(\mathbf{I} - \mathbf{e}_i \otimes \mathbf{1}\right)\tilde{\mathbf{e}}^A_i$$

where $\mathbf{\Phi}_i \equiv \left[\frac{(1-\nu_{i,g})\alpha_{i,kg}}{\mu_g}\right]_{k,g}$ is the $K \times K$ markup-adjusted input-output matrix in country i and $\tilde{\mathbf{e}}^A_i \equiv \left[\tilde{e}^A_{i,k}\right]_k$ and $\mathbf{e}_i \equiv [e_{i,k}]_k$ are $K \times 1$ column vectors. Inverting the above system yields the following closed-form expression for $\tilde{\mathbf{e}}^A_i \equiv [\tilde{e}_{i,k}]_k$ as a function of observables,

$$\tilde{\mathbf{e}}^A_i = \left(\mathbf{I} - \mathbf{\Phi}_i\left(\mathbf{I} - \mathbf{e}_i \otimes \mathbf{1}\right)\right)^{-1}\mathbf{e}_i.$$

F Firm Selection into Export Markets

F.1 Characterizing Net Profit Margins.

As an intermediate step, we characterize the aggregate profit margin (π_i) and consumer price index (P_i) in the presence of fixed marketing costs. These fixed costs, as explained in the main text, exhaust a fraction of the gross profits from markups, thereby reducing π_i. In what follows we characterize π_i in the presence of fixed costs for an economy that is either closed or open. Recall from Section 6.3 that fixed marketing costs in our model account for a constant fraction of origin i's sales to destination j. Namely,

$$M_{ij,k} w_j f_{j,k} = \left(1 - \frac{1+\theta_k}{\mu_k \theta_k}\right) P_{ij,k} Q_{ij,k}.$$

That is, a constant share, $\rho_k = 1 - \frac{1+\theta_k}{\mu_k\theta_k}$, of export sales, $P_{ij,k}Q_{ij,k}$, is paid as a fixed marketing cost to labor in destination j. So, the profits collected from sales of good ij,k net of fixed costs are $\Pi_{ij,k} = \left(1 - \frac{1}{\mu_k}\right) P_{ij,k}Q_{ij,k} - \rho_k P_{ij,k}Q_{ij,k}$. Let $\mathcal{R}_i \equiv \sum_k \sum_j P_{ij,k}Q_{ij,k}$ gross revenues in country i, then total profits in country i, $\Pi_i = \sum_j \sum_k \Pi_{ij,k}$, are given by

$$\Pi_i = \left[1 - \sum_k \left(\frac{1}{\mu_k} + \rho_k\right) r_{i,k}\right]\mathcal{R}_i, \tag{25}$$

where $r_{i,k}$, recall, denotes the industry-level revenue share. Total wage income in country i, meanwhile, equals factor compensation from domestic production plus fixed cost payments from foreign exporters. In particular,

$$w_i L_i = \left[\sum_k \frac{1}{\mu_k} r_{i,k}\right] \mathcal{R}_i + \sum_k \left(\rho_k e_{i,k}\right) E_i, \tag{26}$$

where $E_i = \sum_j \sum_k P_{ji,k} Q_{ji,k}$ denotes total expenditure. Observe that total expenditure in country i should equal wage plus profit income, i.e., $E_i = w_i L_i + \Pi_i$. Invoking this observation alongside Equations 25 and 26 yields the following relationship between national-level revenues and expenditure in country i

$$\mathcal{R}_i = \frac{1 - \sum_k \rho_k e_{i,k}}{1 - \sum_k \rho_k r_{i,k}} E_i.$$

Plugging the above expression back into Equations 25 and 26, yields the following formula for the aggregate profit margin in country i:

$$\pi_i \equiv \frac{w_i L_i + \Pi_i}{w_i L_i} = \frac{E_i}{\left[\sum_k \frac{1}{\mu_k} r_{i,k}\right] \mathcal{R}_i + \sum_k \left(\rho_k e_{i,k}\right) E_i} = \frac{1}{\frac{1-\sum_k \rho_k e_{i,k}}{1-\sum_k \rho_k r_{i,k}} \sum_k \left(\frac{1}{\mu_k} r_{i,k}\right) + \sum_k \left(\rho_k e_{i,k}\right)}. \tag{27}$$

F.2 The Cost of Markups in a Closed Economy (Lemma 4)

Following the logic presented earlier, The degree of misallocation in closed economy i is given by $\mathcal{D}_i^A = \log W_i^{A\star} - \log W_i^A$, where superscript "$\star$" corresponds to efficient equilibrium and superscript "A" denotes autarky. Noting that $W_i = \pi_i w_i L_i / P_i$ and assigning labor in country i as the numeraire (i.e., $w_i^A = w_i^{A\star} = 1$), we get

$$\mathcal{D}_i^A = \log \widehat{\pi}_i^A - \log \widehat{P}_i^A,$$

where $\widehat{\pi}_i^A = \pi_{i,k}^{\star A} / \pi_{i,k}^A$ and $\widehat{P}_i^A = P_i^{A\star} / P_i^A$. Next, we must define the efficient equilibrium. Unlike the baseline model, the optimal allocation is not obtained under marginal cost-pricing. Instead, all markups should be eliminated, but the excess markup that excludes the fraction covering the fixed marketing cost. Let $m_{i,k}(\varphi) < \mu_k$ denote the excess markup in market i, which depends on the firm productivity, φ. After eliminating the excess markup for the existing set of firms, the change in industry k's CES price index is given by

$$\hat{P}_i^{\star A} = \left[\int_{\varphi_{ii,k}^\star}^{\infty} m_{i,k}(\varphi)^{\gamma_k - 1} \lambda_{ii,k}(\varphi)\, dG_{i,k}(\varphi)\right]^{\frac{1}{1-\gamma_k}} = \left[\int_1^{\infty} m^{\gamma_k - 1} \lambda_{ii,k}(m)\, d\mathcal{G}_{i,k}(m)\right]^{\frac{1}{1-\gamma_k}},$$

where $\varphi_{ii,k}^*$ denotes the zero profit productivity cut-off and $\mathcal{G}_{i,k}(.)$ denotes the distribution of excess markup in market i–industry k. To economize on the notation, we use $\mathcal{M}_{i,k} \equiv \left[\int_1^\infty m^{\gamma_k - 1} \lambda_{ii,k}(m)\, d\mathcal{G}_{i,k}(m)\right]^{\frac{1}{\gamma_k - 1}}$ to denote the CES average *excess* markup in industry k. The change in the aggregate consumer price index is, correspondingly, $\log \widehat{P}_i^A = \log \left(\prod_k \hat{P}_{i,k}^{e_{i,k}}\right) = -\mathbb{E}_{e_i}\left[\log \mathcal{M}_i\right]$. To determine $\hat{\pi}_i^A$, we can appeal to Equation 27, noticing that expenditure and revenue shares coincide under autarky (i.e., $r_{i,k}^A = e_{i,k}^A$)

and excess profit margins are zero under the efficient equilibrium, $\pi_i^{A\star} = 1$. In particular,

$$\hat{\pi}_i^A = \frac{1}{\sum_k \left(\left[\frac{1}{\mu_k} + \rho_k \right] e_{i,k} \right)},$$

which implies that $\log \hat{\pi}_i = -\log \mathbb{E}_{e_i} \left[\frac{1}{\mu} + \rho \right]$. Plugging the expressions for $\log \widehat{P}_i^A$ and $\log \widehat{\pi}_i^A$ backs into our initial expression for $\mathcal{D}_i^A$, yields

$$\mathcal{D}_i^A = \mathbb{E}_{e_i} \left[\log \mathcal{M}_i \right] - \log \mathbb{E}_{e_i} \left[\frac{1}{\mu} + \rho \right].$$

F.3 The Impact of Trade on the Welfare Cost of Markups (Proposition 4)

It is straightforward to check that, under trade, the distribution of excess of markups is the same for firms from all origin countries selling to market i. This feature stems from two assumptions: the Pareto assumption on firm-level productivities and the assumption that fixed marketing cost are paid in terms of labor in destination j. Considering this the deflation in price index after eliminating excess markups is the same whether country i operates as a closed or open economy, i.e., $\log \widehat{P}_i^A = \log \widehat{P}_i = -\mathbb{E}_{e_i} \left[\log \mathcal{M}_i \right]$, where $\widehat{P}_i = P_i^\star / P_i$ and $\mathcal{M}_i$ is invariant to trade following the logic outlined in Section 5.2. As such, the impact of trade on misallocation is

$$\mathcal{D}_i - \mathcal{D}_i^A = \log \left(\widehat{\pi}_i / \widehat{\pi}_i^A \right) - \log \left(\widehat{P}_i / \widehat{P}_i^A \right) = \log \left(\widehat{\pi}_i / \widehat{\pi}_i^A \right),$$

where $\widehat{P}_i = P_i^\star / P_i$ and $\widehat{\pi}_i = \pi_i^\star / \pi_i$ are the change in the consumer price index and profit margins after restoring efficiency in an open economy. From the previous section, we know that $\widehat{\pi}_i^A = -\log \mathbb{E}_{e_i} \left[\frac{1}{\mu} + \rho \right]$. Applying our notation for expectations to Equation 27, implies

$$\log \left(\widehat{\pi}_i \right) = \log \left(\frac{\pi_i^\star}{\pi_i} \right) = \log \left(\frac{1 - \mathbb{E}_{r_i} \left[\rho \right]}{1 - \mathbb{E}_{e_i} \left[\rho \right]} \mathbb{E}_{r_i} \left[\frac{1}{\mu} \right] + \mathbb{E}_{e_i} \left[\rho \right] \right).$$

Notice that since $\mathbb{E}_{e_i} \left[\frac{r_i}{e_i} \right] = 1$, then $\mathbb{E}_{r_i} \left[\rho \right] = \mathrm{Cov}_{e_i} \left(\rho, \delta_i \right) + \mathbb{E}_{e_i} \left[\rho \right]$, where $\delta_i \equiv \frac{r_i}{e_i} - 1$. Replacing this expression into the above equation and plugging the resulting expressions for $\log \left(\widehat{\pi}_i \right)$ and $\log \left(\widehat{\pi}_i^A \right)$ back into the equation for $\Delta \mathcal{D}_i = \mathcal{D}_i - \mathcal{D}_i^A$ yields

$$\Delta \mathcal{D}_i = \log \mathbb{E}_{e_i} \left[\frac{1}{\mu} + \rho \right] - \log \left(\mathbb{E}_{r_i} \left[\frac{1}{\mu} + \rho \right] - \left(1 + \frac{\mathbb{E}_{r_i} \left[\frac{1}{\mu} \right]}{1 - \mathbb{E}_{e_i} \left[\rho \right]} \right) \mathrm{Cov}_{e_i} \left(\rho, \delta_i \right) \right)$$

Recall that ρ_k represents the share of sales that are paid to cover the fixed cost. Accordingly, setting $\rho = 0$ in the above expression delivers our baseline formula for the effect of trade on misallocation—which was presented under Proposition 1.

G Other Extensions

G.1 Capital as a Primary Production Input

Suppose production in industry, k, employs labor and capital inputs. Whereas labor is perfectly mobile across industries, capital inputs are industry-specific, with $\bar{\mathcal{K}}_k$ denoting the constant supply of capital for industry k's production. Suppose the production function has a Cobb-Douglas parametrization with $\kappa_{i,k}$ denoting the constant share of capital in production. The monopolistically competitive price index of goods sold by origin i to destination j in industry s is, thus,

$$P_{ij,k} = \mu_k \frac{\tau_{ij,k}}{\bar{\varphi}_{i,k}} w_i^{\varsigma_k} \nu_{i,k}^{1-\varsigma_k},$$

where $\nu_{i,k}$ denotes the rental rate of capital in origin i and industry k. Note that per cost minimization, $\nu_{i,k}\bar{\mathcal{K}}_{i,k} = \frac{1-\varsigma_k}{\varsigma_k} w_i L_{i,k}$, where $L_{i,k}$ denotes the labor employed by industry k in origin i, with $\sum_k L_{i,k} = 1$. Let $\ell_{i,k} = \frac{L_{i,k}}{L_i}$ denote the labor share of industry. Hence, the rental-rate-adjusted price indexes associated with economy i can be re-written as

$$P_{ij,k} = \mu_k \bar{a}_{ij,k} w_i \ell_{i,k}^{1-\varsigma_k}; \qquad\qquad P_i = \prod_k \left(\sum_i P_{ni,k}^{1-\sigma_k} \right)^{\frac{e_{i,k}}{1-\sigma_k}},$$

where $\bar{a}_{ij,k}$ encompasses all constant cost-shifters associated with ij,k. Total nominal income, meanwhile, is equal to wage and rental payments plus markup rent rebates. In particular,

$$Y_i = w_i L_i + \sum_k [\nu_{i,k}\bar{\mathcal{K}}_{i,k}] + \Pi_i.$$

Noting that $\nu_{i,k}\bar{\mathcal{K}}_{i,k} = \frac{1-\varsigma_k}{\varsigma_k} w_i \ell_{i,k} L_i$ and $\Pi_i = \sum_k [(\mu_k - 1)(w_i L_{i,k} + \nu_{i,k}\bar{\mathcal{K}}_{i,k})]$, we can re-write the expression for nominal income as

$$Y_i = \sum_k \left(\frac{\mu_k}{\varsigma_k} \ell_{i,k} \right) w_i L_i = \mathbb{E}_{\ell_i} \left[\frac{\mu}{\varsigma} \right] w_i L_i. \tag{28}$$

With the above background, we are ready to characterize the welfare cost of markup, which corresponds to the change in welfare after markups are eliminated. More specifically, we wish to characterize $\mathcal{D}_i \equiv \log \hat{W}_i = \log (W_i^\star / W_i)$, where $W_i = Y_i / P_i$ and the ($\star$) superscripts denotes the efficient equilibrium wherein marginal cost-pricing is restored. We first analyze a closed economy and then proceed to the open economy case.

The Closed Economy Case. To simplify the notation we avoid adding an additional superscript, A, to variables to denote autarky. But keep in mind that all variables are being evaluated as such. The change in the consumer price index for closed economy i can be expressed as

$$\hat{P}_i = \prod_k \left(\frac{1}{\mu_k} \hat{w}_i \hat{\ell}_{i,k}^{1-\varsigma_k} \right)^{e_{i,k}}.$$

We can set the change in the wage rate to one by choice of numeraire, i.e., $\hat{w}_i = 1$. The characterize $\hat{\ell}_{i,k}$, note that $w_i L_{i,k} = \frac{1-\kappa_k}{\mu_k} P_{i,k} Q_{i,k}$, which yields the following relationship between labor and revenue shares:

$$\ell_{i,k} = \frac{\frac{\varsigma_k}{\mu_k} r_{i,k}}{\sum_{k'} \frac{\varsigma_{k'}}{\mu_{k'}} r_{i,k'}}. \tag{29}$$

which given that under autarky, $r_{i,k} = e_{i,k}$ is constant, implies the following

$$\ell_{i,k} = \frac{\frac{\varsigma_k}{\mu_k} e_{i,k}}{\sum_{k'} \frac{\varsigma_{k'}}{\mu_{k'}} e_{i,k'}}; \qquad \ell^{\star}_{i,k} = \frac{\varsigma_k e_{i,k}}{\sum_{k'} \varsigma_{k'} e_{i,k'}}.$$

Combing the above expressions we can determine $\hat{\ell}_{i,k} = \ell^{\star}_{i,k} / \ell_{i,k} = \mu_k \frac{\sum_{k'} \frac{1}{\mu_{k'}} \varsigma_{k'} e_{i,k'}}{\sum_{k'} \varsigma_{k'} e_{i,k'}} \sim \mu_k \mathbb{E}_{\varsigma e_i} \left[\frac{1}{\mu} \right]$, which when plugged into our earlier expression for $\log \hat{P}_i$ delivers

$$\begin{aligned} \log \hat{P}_i &= \sum_k \left[\varsigma_k e_{i,k} \log \frac{1}{\mu_k} \right] + \log \mathbb{E}_{\varsigma e_i} \left[\frac{1}{\mu} \right] \sum_k (1 - \varsigma_k) e_{i,k} \\ &\sim \ \mathbb{E}_{e_i} \left[\varsigma \log \frac{1}{\mu} \right] \ + \ \log \mathbb{E}_{\varsigma e_i} \left[\frac{1}{\mu} \right] (1 - \mathbb{E}_{e_i} [\varsigma]) . \end{aligned}$$

We can characterize $\hat{Y}_i$ by appealing to Equations 28 and 29, which yields

$$\log \hat{Y}_i = \log \left(\frac{\sum_k \left(\frac{1}{\varsigma_k} \ell^{\star A}_{i,k} \right)}{\sum_k \left(\frac{\mu_k}{\varsigma_k} \ell^{A}_{i,k} \right)} \hat{w}_i L_i \right) = \log \left(\frac{\sum_{k'} \frac{1}{\mu_{k'}} \varsigma_{k'} e_{i,k'}}{\sum_{k'} \varsigma_{k'} e_{i,k'}} \right) \sim \log \mathbb{E}_{\varsigma e_i} \left[\frac{1}{\mu} \right],$$

where the last line uses $\hat{w}_i = 1$. Combing the expressions for $\hat{Y}_i$ and $\hat{P}_i$ yields the following formula for the welfare cost of markups in a closed economy, $\mathcal{D}_i = \log \hat{W}_i = \log \left(\hat{Y}_i / \hat{P}_i \right)$,

$$\mathcal{D}_i = \log \mathbb{E}_{\varsigma e_i} \left[\frac{1}{\mu} \right] \mathbb{E}_{e_i} [\varsigma] - \mathbb{E}_{e_i} \left[\varsigma \log \frac{1}{\mu} \right].$$

Based on the above formula we can infer the welfare cost of markups for country i under autarky with information markup levels as well as expenditure and labor input shares.

The Open Economy Case. To simplify the open economy case assume that $e_{i,k} = e_{n,k} = e_k$ for i and n. Following the logic outlined in Appendix C.2, a global restoration of marginal cost pricing relative international prices within industries, implying that $\hat{\ell}_{i,k} = \hat{\ell}_{n,k}$ and $\hat{w}_i = \hat{w}_n$ for all i and n. Moreover, given Equation 29 and internationally symmetric Cobb-Douglas preferences across industries, it is straightforward to check that

$$\hat{\ell}_{i,k} = \hat{\ell}_{n,k} = \mu_k \mathbb{E}_{\varsigma e}\left[\frac{1}{\mu}\right],$$

which given our choice of numeraire ($\hat{w}_i = 1$) and the steps presented in closed economy case, yields

$$\log \hat{P}_i = \mathbb{E}_e\left[\varsigma \log \frac{1}{\mu}\right] + \log \mathbb{E}_{\varsigma e}\left[\frac{1}{\mu}\right](1 - \mathbb{E}_e[\varsigma]).$$

Likewise, as before, we can combine Equations 28 and 29 to derive the following expression for the change in nominal income

$$\hat{Y}_i = \log \mathbb{E}_{\varsigma r_i}\left[\frac{1}{\mu}\right].$$

Combing the expressions for $\hat{Y}_i$ and $\hat{P}_i$ yields $\mathcal{D}_i = \log \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right](\mathbb{E}_{e_i}[\varsigma] - 1) - \mathbb{E}_{e_i}\left[\varsigma \log \frac{1}{\mu}\right] + \log \mathbb{E}_{\varsigma r_i}\left[\frac{1}{\mu}\right]$, implying the following formula for the welfare consequences of international rent-shifting for country i:

$$\Delta\mathcal{D}_i = \log \mathbb{E}_{\varsigma r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right].$$

Notice, the above formula differs from our baseline formula in that the averages are weight by labor input-adjusted revenues shares. In the limit where $\varsigma_k = 1$ for all k, the above equation reduces to our baseline formula for $\Delta\mathcal{D}_i$.

G.2 Origin and Destination-Specific Markups

Origin-Specific Markups. Suppose markups are origin-specific, with $\mu_{n,k}$ denting the constant markup charge by firms from country of origin n. In that case, the *conditional* change in the destination i's price index in response to a global restoration of marginal cost-pricing is[31]

$$\log \hat{P}_i = \sum_k \frac{e_{i,k}}{1-\sigma_k} \log \sum_n \left[\lambda_{ni,k}\left(\frac{1}{\mu_{n,k}}\right)^{1-\sigma_k}\right] = \mathbb{E}_{e_i}\left[\log \frac{1}{\mathcal{M}_i}\right], \quad \text{where } \frac{1}{\mathcal{M}_{i,k}} \equiv \left[\sum_n \lambda_{ni,k}\left(\frac{1}{\mu_{n,k}}\right)^{1-\sigma_k}\right]^{\frac{1}{1-\sigma_k}}.$$

Stated verbally, $\mathcal{M}_{i,k}$ corresponds to the sales-weighted average markup paid by country i on goods pertaining to industry k. The change in income is still given by $\log \hat{Y}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu_i}\right]$. Under autarky, where $\lambda_{ii,k} = 1$ and $r_{i,k} = e_{i,k}$, one can deduce that $\log \hat{P}_i = \mathbb{E}_{e_i}\left[\log \frac{1}{\mu_i}\right]$ and $\log \hat{Y}_i = \log \mathbb{E}_{e_i}\left[\frac{1}{\mu_i}\right]$, which yields the following formula for the welfare cost of markups, $\mathcal{D}_i^A = \log \hat{Y}_i - \log \hat{P}_i$:

$$\mathcal{D}_i^A = \log \mathbb{E}_{e_i}\left[\frac{1}{\mu_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu_i}\right].$$

[31] The conditional change corresponds to the change in price index holding wages constant. The corresponding change in welfare can be viewed as the distance to the efficient frontier if all countries restore marginal cost pricing while exchanging transfers in the form of wage subsidies/taxes, which maintain baseline wages to neutralize terms-of-trade externalities.

Note that the above formula is akin to that provided under Lemma 1, with the difference that mean operator is applied to the unique markup level charged by firms in country i. In an open economy, meanwhile, the welfare cost of markups can be specified as

$$\mathcal{D}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{\mu_i}{\mathcal{M}_i}\right].$$

Subtracting $\mathcal{D}_i^A$ from $\mathcal{D}_i$ yields the following formula describing how trade impacts the welfare cost in markups for country i:

$$\Delta\mathcal{D}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\mu_i}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu_i}\right] + \mathbb{E}_{e_i}\left[\log \frac{\mu_i}{\mathcal{M}_i}\right].$$

On the right-hand side, the term, $\log \mathbb{E}_{r_i}\left[\frac{1}{\mu_i}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu_i}\right] \approx \text{Cov}_{e_i}\left(\frac{1}{\mu_i}, \text{RCA}_i\right) \mathbb{E}_{e_i}\left[\frac{1}{\mu_i}\right]^{-1}$, represents international rent-shifting *if* foreign suppliers applied the same markup as domestic suppliers in country i. The last term on the right-hand side represents the spread between the markup paid to domestic and foreign suppliers, which can further contribute to international rent-shifting.

To evaluate $\Delta\mathcal{D}_i$, one needs data on industry-level expenditure and revenue shares and bilateral expenditure shares, $\lambda_{ni,k}$, and origin-specific markup levels, $\mu_{n,k}$, to determine $\mathcal{M}_{i,k}$. Data on bilateral expenditure shares are widely available and do not obstruct quantification. Origin-specific markups are, however, difficult to estimate with currently available data sets. Take the World Scope database as an example. First, many countries are represented by a small sample of firms, complicating extrapolation. Second, sales and cost data are consolidated for each firm across all international affiliates. The consolidation of data this way is a non-issue when dealing with a standard trade model in which the sales-weighted average markup is common across origins. We can infer the common sales-weighted average by pooling firms from different origins. This solution no longer works if we intend to recover the sales-weighted average markup on an origin-by-origin basis.

Destination-Specific Markups. Suppose firms in selling to destination i in industry k charge a destination-specific markup, $\tilde{\mu}_{i,k}$, which reflects pricing-to-market among other things. It is straightforward to verify that the change in the consumer price index in country i after restoring marginal cost-pricing is $\hat{P}_i = \sum_k \left[e_{i,k}\frac{1}{\tilde{\mu}_{i,k}}\right] = \mathbb{E}_{e_i}\left[\log \frac{1}{\tilde{\mu}_i}\right]$. The corresponding change in country i's total income depends on the average markup charged by country i's firms across different destination markets. Namely,

$$\log \hat{Y}_i = \mathbb{E}_{r_i}\left[\log \frac{1}{\tilde{\mathcal{M}}_i}\right], \quad \text{where} \quad \frac{1}{\mathcal{M}_{i,k}} \equiv \left[\sum_n r_{in,k}\frac{1}{\tilde{\mu}_{n,k}}\right].$$

In the above expression, $r_{in,k} \equiv \frac{P_{in,k}Q_{in,k}}{\sum_{n'} P_{in',k}Q_{in',k}}$ denotes the share of sales to destination n in country i's total sales in industry k. Under autarky, where $r_{ii,k} = 1$ and $r_{i,k} = e_{i,k}$, one can deduce that $\log \hat{P}_i = \mathbb{E}_{e_i}\left[\log \frac{1}{\tilde{\mu}_i}\right]$ and $\log \hat{Y}_i = \log \mathbb{E}_{e_i}\left[\frac{1}{\tilde{\mu}_i}\right]$, which yields the following formula for the welfare cost of markups, $\mathcal{D}_i^A = \log \hat{Y}_i - \log \hat{P}_i$:

$$\mathcal{D}_i^A = \log \mathbb{E}_{e_i}\left[\frac{1}{\tilde{\mu}_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\tilde{\mu}_i}\right].$$

Note that the above formula is akin to that provided under Lemma 1, with the difference that mean operator is applied to the unique markup level charged by firms in market *i*. In an open economy, meanwhile, the welfare cost of markups can be specified as

$$\mathcal{D}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\tilde{\mu}_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\tilde{\mu}_i}\right] - \mathbb{E}_{e_i}\left[\log \frac{\tilde{\mathcal{M}}_i}{\tilde{\mu}_i}\right].$$

Subtracting $\mathcal{D}_i^A$ from $\mathcal{D}_i$ yields the following formula describing how trade impacts the welfare cost in markups for country *i*:

$$\Delta\mathcal{D}_i = \log \mathbb{E}_{r_i}\left[\frac{1}{\tilde{\mu}_i}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\tilde{\mu}_i}\right] + \mathbb{E}_{e_i}\left[\log \frac{\tilde{\mathcal{M}}_i}{\tilde{\mu}_i}\right].$$

On the right-hand side, the term, $\log \mathbb{E}_{r_i}\left[\frac{1}{\tilde{\mu}_i}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\tilde{\mu}_i}\right] \approx \text{Cov}_{e_i}\left(\frac{1}{\tilde{\mu}_i}, \text{RCA}_i\right) \mathbb{E}_{e_i}\left[\frac{1}{\tilde{\mu}_i}\right]^{-1}$, represents international rent-shifting *if* all destination markets paid the same markup as market *i*. The last term on the right-hand side represents the spread between the markup collected from and paid to foreign entities, which can further contribute to international rent-shifting. As before, empirically evaluating $\Delta\mathcal{D}_i$ requires data on destination-specific markups, which are even more difficult to estimate. The cost-based estimation technique practically infeasible to implement when dealing with multi-product firms that charge differential markups across different destination markets. The demand-based estimation technique is equally difficult to implement due to excessive data requirements. One, in particular, needs transaction-level import data across many countries which is far from available.

G.3 Elastic Labor Supply

Suppose the representative consumer's welfare is equal to the utility from consumption minus the the disutility from labor provision. Namely,

$$W_i = \frac{w_i \ell_i + \Pi_i}{P_i} - \frac{1}{1+\frac{1}{\epsilon}} \ell_i^{1+\frac{1}{\epsilon}},$$

where $\ell_i \geq 0$ denotes the total labor supplied by the representative consumer in country *i* and P_i is the Cobb-Douglas-CES price index. Welfare maximization (taking rents as given) yields a labor supply function, $\ell_i = \left(\frac{w}{P_i}\right)^{\epsilon}$, with a constant elasticity ϵ. Following our previous choice of notation, let $\pi_i \equiv 1 + \frac{\Pi_i}{w_i \ell_i}$ to denote the ratio of markup rents to wage payments, with $\pi_i w_i \ell_i$ denoting total expenditure income in country i. Plugging the expression for ℓ_i into the welfare function yields

$$W_i = \left(\pi_i - \frac{\epsilon}{1+\epsilon}\right)\left(\frac{w_i}{P_i}\right)^{1+\epsilon}.$$

We can appeal to the hat-notation to express the change in welfare in response to restoring marginal cost-pricing as

$$\hat{W}_i = \frac{\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]}{1+\epsilon-\epsilon\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]} \times \left(\frac{1}{\hat{P}_i}\right)^{1+\epsilon},$$

where $\hat{P}_i = \prod_k \left(\frac{1}{\mu_k}\right)^{e_{i,k}}$, which implies that $\log \hat{P}_i = \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right]$. Plugging the expression for $\hat{P}_i$ cack into the above equation and rearranging the terms, yields

$$\mathcal{D}_i = (1+\epsilon)\left[\log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\log \frac{1}{\mu}\right]\right] - \log\left((1+\epsilon)\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{\epsilon} - \epsilon\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{1+\epsilon}\right).$$

Notice that even if markups are uniform, $\mathcal{D}_i$ is still positive, due to the additional term,

$$\log\left((1+\epsilon)\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{\epsilon} - \epsilon\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{1+\epsilon}\right).$$

This term reflects that a uniform markup on consumer goods distorts the choice between leisure and consumption, leading to a above-optimal supply of labor. The effect of trade on welfare cost of markups can be, accordingly, measured as

$$\Delta\mathcal{D}_i = (1+\epsilon)\left[\log \mathbb{E}_{r_i}\left[\frac{1}{\mu}\right] - \log \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]\right] + \log\left(\frac{(1+\epsilon)\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{\epsilon} - \epsilon\,\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]^{1+\epsilon}}{(1+\epsilon)\,\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right]^{\epsilon} - \epsilon\,\mathbb{E}_{r_i}\left[\frac{1}{\mu}\right]^{1+\epsilon}}\right).$$

The first term represents international rent-shifting, the impact of which is amplified by the labor supply elasticity, ϵ. The second term represents how trade aggravates or mitigates the inefficient labor supply problem. The data requirements to evaluate the above expression are as before with the addition of the labor supply elasticity, ϵ.

H Demand-Based Markup Estimation: *Details*

This appendix provides a more detailed description of our demand-based markup estimation procedure. Our identification strategy is borrowed from Lashkaripour and Lugovskyy, 2021, and amended to account for heterogeneity in individual-level demand elasticities. Since we are analyzing data from only one market, Colombia, we hereafter drop the destination market index, i, to simplify notation.

Our estimating equation can, thus, be stated as follows in first-differences:

$$\Delta \ln q_{j,kt}(\omega) = -\gamma_{kt}\Delta \ln p_{j,kt}(\omega) + \sigma^2_{\gamma t}\Delta K_{j,kt}(\omega) + \Delta D_{jkt} + \Delta\tilde{\varepsilon}_{\omega jkt}, \tag{30}$$

where $\Delta\tilde{\varepsilon}_{\omega jkt}$ represents a variety-specific demand shock and ΔD_{jkt} is an origin–product–year fixed effect. Of the remaining variables, $\Delta \ln p_{j,kt}(\omega)$ and $\Delta \ln q_{j,kt}(\omega)$ are directly observable for each import variety, while can be calculated using firm variety-level unit price data. In our estimation equation, k indexes an HS10 product category. But to calibrate our sufficient statistics formulas, we must estimate demand elasticities and markups for 19 broadly-defined and traded ICIO sectors. We, therefore, pool together HS10 products belonging to the same ICIO sector, $\mathcal{S}$, and estimate Equation 30 on this pooled sample assuming that σ_{kt} and γ_{kt} are uniform across products within the same industry. That is,

$$\gamma_{kt} = \gamma_{\mathcal{S}t}, \qquad \sigma_{kt} = \sigma_{\mathcal{S}t}; \qquad\qquad \forall k \in \mathbb{K}_{\mathcal{S}}$$

where $\mathbb{K}_{\mathcal{S}}$ denotes the set of HS10 products pertaining to ICIO sector $\mathcal{S}$. We handle outliers by trimming our sample to exclude observations that report a price and quantity changes above the 97.5th percentile of the relevant product-year cell.

To handle the identification challenge underlying demand estimation, we borrow the shift-share research design developed by Lashkaripour and Lugovskyy, 2021. We, in particular, construct a shift-share instrument for price changes that interact the monthly change in annual exchange rates with the *lagged* monthly export shares to Colombia (at the firm variety-level). To present our instrument, let $\Delta \ln \mathcal{E}_{jt}(m)$ denote the year-over-year change in origin j's exchange rate with the Colombian Peso in month m of year t. Also, let $\pi_{j,kt}(\omega; m)$ denote the share of month m sales in variety ωjkt's annual export sales to Colombia. Our shift-share instrument can be specified as,

$$z_{j,kt}(\omega) = \sum_{m\in\mathcal{M}} \pi_{j,kt-1}(\omega, m)\, \Delta \ln \mathcal{E}_{jt}(m).$$

where note that $\pi_{j,kt-1}(\omega, m)$ is the lagged (rather than concurrent) monthly export share. The above instrument, by design, measures exposure to exchange rate shocks at the *firm*×*origin*×*product*×*year* level based on prior export activity. It satisfies relevancy as implied by the first-stage F-statistics reported in Table 2.. A formal discussion on exclusion restrictions can be found in Lashkaripour and Lugovskyy, 2021. We also use the number of (alternative) product codes served by firm ω in year t to instrument for $\Delta K_{j,kt}(\omega)$. Following standard practices in the literature, such count variables satisfy exclusion restrictions if entry decisions are made prior to the realization of idiosyncratic demand shocks, $\Delta\tilde{\varepsilon}_{\omega jkt}$.

We estimate γ_{kt} as a moving average demand elasticity by fitting Equation 30 to pooled data for years t, $t-1$, and $t+1$. This way, we are able to track the change in market power over time. We also perform a pooled estimation over all years in our sample, with results reported in Table 2. This table reports the ICIO sectors for which we want to obtain markup estimates, the estimate demand elasticity (for the pooled sample), as well the first stage F-statistics from the Kleibergen-Paap Wald test for weak identification. The reported first-stage F-statistics average a little over 60, validating the relevance of our shift-share instrument.

Table 2: Pooled Demand estimation results by ICIO sector

			Estimated Parameter			
ICIO	Description	ISIC	γ_k	std error	F-stat	Observations
1-3	Agriculture & Mining	1-9	5.26	3.45	16,331	3.72
4	Food	10-12	3.01	0.43	35,266	21.96
5	Textiles, Leather & Footwear	13-15	9.56	0.75	75.05	186,489
6	Wood	16	3.29	7.74	0.08	7,178
7	Paper	17-18	5.32	2.35	3.76	24,467
8	Petroleum	19	2.90	0.76	5.19	4,842
9	Chemicals	20-21	2.61	2.03	17.90	192,020
10	Rubber & Plastic	22	2.51	0.37	131.87	140,798
11	Minerals	23	3.66	0.32	70.58	38,848
12	Basic Metals	24	5.54	1.23	13.62	38,831
13	Fabricated Metals	25	3.75	1.68	120.73	153,793
14	Electronics	26	8.07	16.90	2.24	191,012
15	Electrical Equipment	27	4.39	1.13	76.58	166,646
16	Machinery	28	3.57	0.25	236.84	330,676
17	Motor Vehicles	29	3.53	0.50	209.52	145,053
18	Other Transport Equipment	30	3.97	0.98	9.09	10,534
19	N.E.C. & Recycling	31-33	4.99	4.55	22.38	123,613

Notes. Estimation results of Equation (30). Standard errors in parentheses. The estimation is conducted with HS10 product-year-origin fixed effects. The weak identification test statistics is the F statistics from the Kleibergen-Paap Wald test for weak identification of all instrumented variables. The test for over-identification is not reported due to the pitfalls of the standard over-identification Sargan-Hansen J test in the multi-dimensional large datasets pointed by Angrist et al. (1996).

I The Deep Origins of International Rent-Shifting

International rent-shifting benefits countries with a comparative advantage in high-markup industries at the expense of others. This type of comparative advantage often has deep roots in national institutions and resource endowments. This appendix explores how these deep characteristics may explain the cross-national heterogeneity in $\Delta\mathcal{D}_i$, which denotes the increase in the cost of markups due to international rent-shifting.

Figure 8 displays the relation between $\Delta\mathcal{D}_i$ and some well-known determinants of comparative advantage. The upper panel of Figure 8 examines the role of three institutional factors: First, the "rule of law," which is the ability to enforce contracts in a given country. Following Nunn (2007), Costinot (2009), and Chor (2010), we use the national indicator for contracting institutions from the World Bank's WORLDWIDE GOVERNANCE INDICATOR. This indicator captures individuals' perceptions of the quality of contract enforcement, property rights, the police, the courts, and the likelihood of crime and violence, with a larger number corresponding to a higher quality of contracting institutions. Based on Figure 8, countries that score better on the "rule of law" indicator are net beneficiaries from rent-shifting (i.e., exhibit a negative $\Delta\mathcal{D}_i$). Intuitively, high-markup industries, which are more differentiated, require more relationship-specific and non-contractable inputs, necessitating a well-developed contracting institution (Acemoglu et al., 2007; Levchenko, 2007; Nunn, 2007). Our results on this front echo those in (Kaufmann et al., 2010) that countries with better contracting institutions tend to experience a reduction in misallocation after opening to trade.

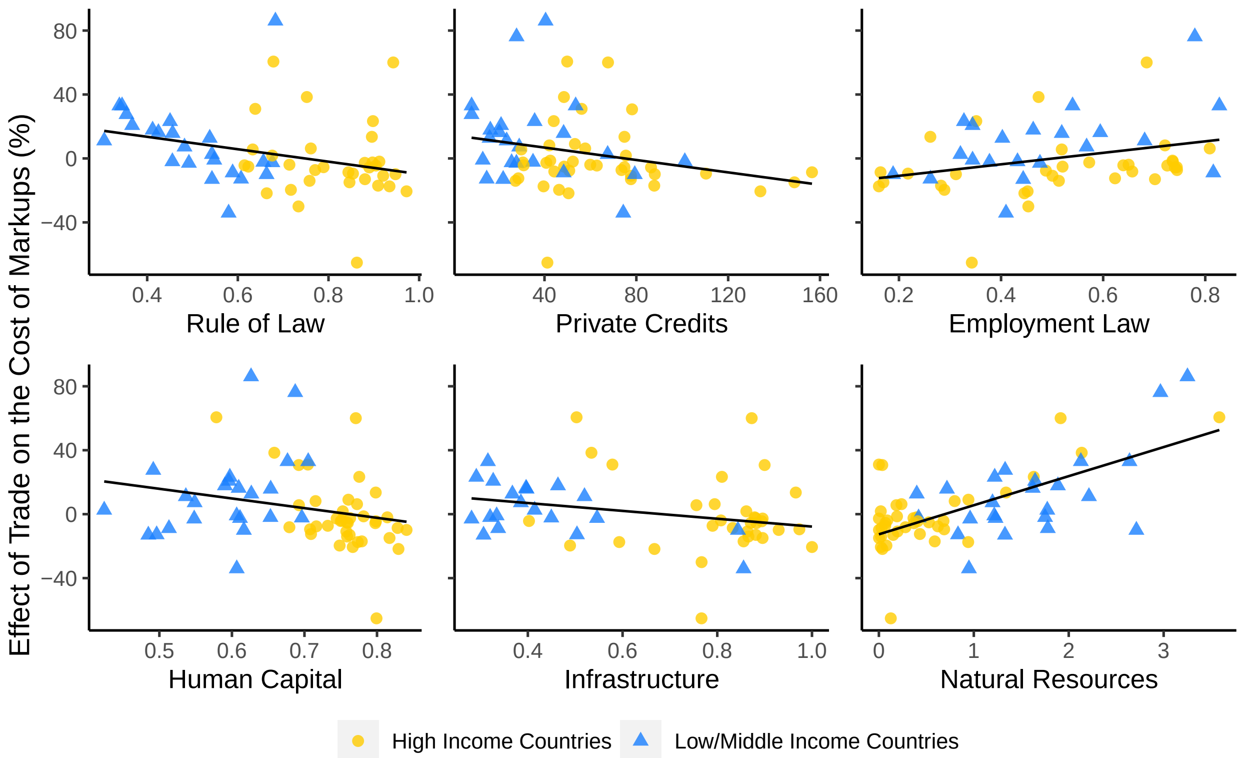
Second, we examine the role of financial development, which is associated with specialization in industries that rely more heavily on external finance (Kletzer and Bardhan, 1987; Beck, 2002; Matsuyama, 2005; and Manova, 2013). We observe that financial development (as proxied by private credits) is associated with specialization in high-markup industries and being a net beneficiary of international rent-shifting. Private credit, here, is defined as the share of private credit by deposit money banks and other financial institutions to GDP. Our results are robust to alternative measures of financial development, such as the stock market capitalization, the ratio of liquid liabilities to GDP, the importance of banks relative to the central bank, and the ratio of claims on the non-financial private sector to total domestic credit used by Rajan and Zingales (1998) and King and Levine (1993).

Third, we look at the role of labor market institutions, which are considered an important determinant of comparative advantage (Costinot, 2009; Cuñat and Melitz, 2012; Tang, 2012). Countries with more flexible labor market institutions tend to have a comparative advantage in more volatile industries—they are better poised to respond to shocks by hiring and firing workers as necessary (Cuñat and Melitz, 2012). We measure labor market flexibility using the "Employment Laws Index" developed by Botero et al. (2004). This index measures the protection of labor and employment laws, with the larger number corresponding to a higher level of protection. It is an average of (1) alternative employment contracts; (2) cost of increasing hours worked; (3) cost of firing workers; and (4) dismissal procedures. Interestingly, Figure 8 indicates that stricter labor protection laws are associated with specialization in low-markup industries and increased exposure to adverse rent-shifting effects.[32]

The lower panel of Figure 8 displays the effect of infrastructure and factor endowment on $\Delta\mathcal{D}_i$. Extrapolating from the Heckscher-Ohlin theorem, these factors can influence comparative advantage across low- and high-markup industries. Country-level measures of human capital and infrastructure are taken from Hall and Jones (1999). Figure 9 suggests that better infrastructure and human capital contribute to specialization in high-markup industries and becoming net beneficiaries of rent-

[32] Botero et al. (2004) also provide the "Collective Relations Laws Index" measuring the protection of collective relations laws as the average of labor union power and collective disputes. Our results are robust when using the "Collective Relations Laws Index".

Figure 8: Determinants of $\Delta\mathcal{D}$: national institutions and resource endowment

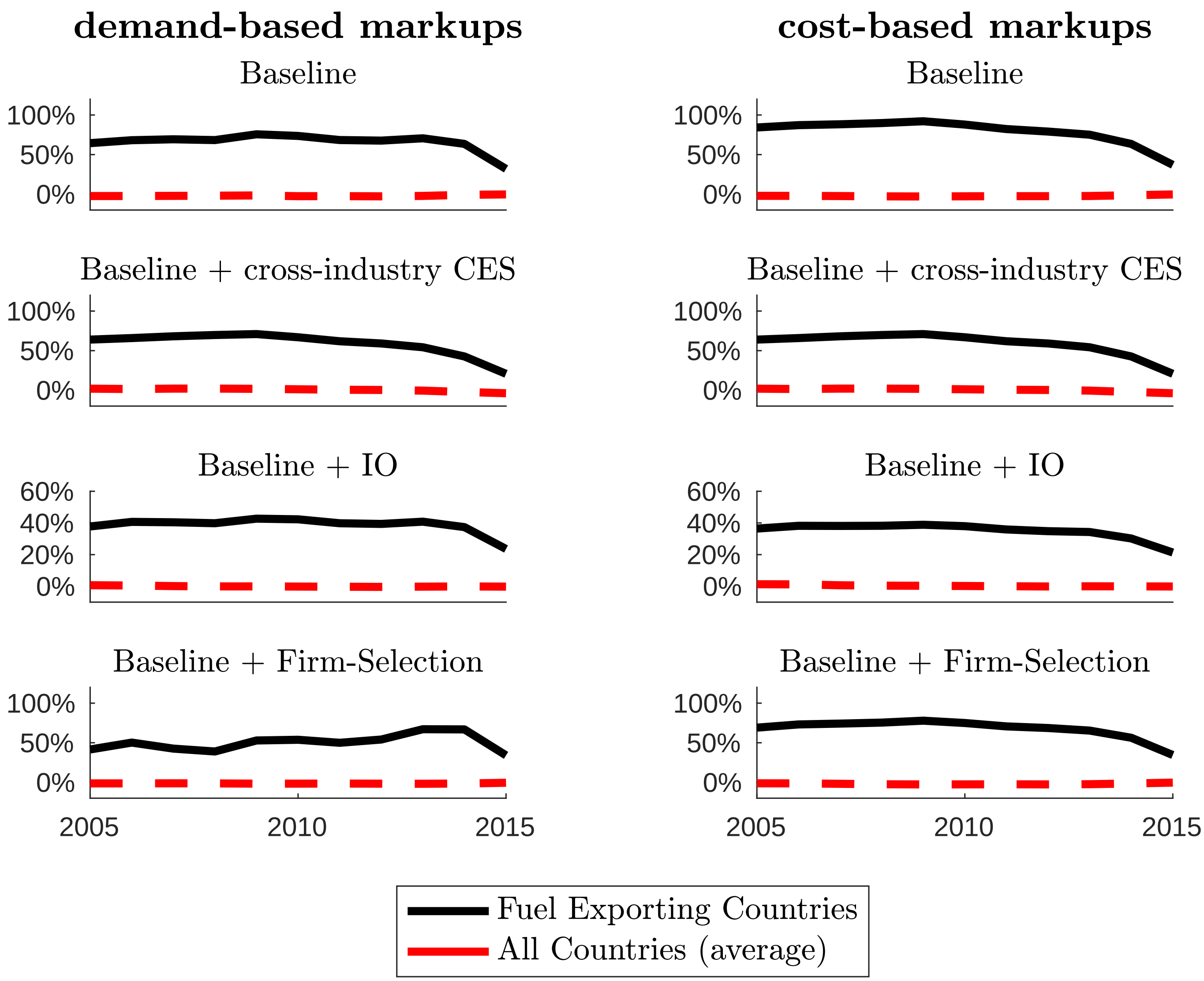


Note: The variable in the y-axis is the effect of trade on the cost of markups ($\Delta\mathcal{D}_i$) in each countries in 2015 estimated by the baseline model with demand-based markups (see column 1 in Table 3). The "Rule of Law" index is taken from the World Bank's WORLDWIDE GOVERNANCE INDICATOR, averaging from 2000 through 2014. The "Private Credits" is the share of private credit by deposit money banks and other financial institutions to GDP in each country, taken from World Bank's GLOBAL FINANCIAL DEVELOPMENT and averaged from 2000 through 2014. The "Employment Law" index is taken from Botero et al. (2004), which captures different aspects of the regulation of labor markets in each country. The "Human Capital" and "Infrastructure" are the human capital per worker (log) and the index of social infrastructure taken from Hall and Jones (1999). The "Natural Resources" is the share of total natural resources rents to GDP (log) taken from the World Bank's WORLD DEVELOPMENT INDICATOR, averaging from 2000 through 2014.

shifting (i.e., exhibiting a negative $\Delta\mathcal{D}_i$). Our measure of natural resource endowment is from the World Bank.

Among the many factors examined in Figure 8, natural resource abundance plays a pivotal role. Figure 9, therefore, explores the dependence of rent-shifting on natural resource-abundance in more depth. It displays the increase in the cost of markups due to international rent-shifting among fuel exporting countries, where fuel is broadly defined to include most energy sources. International rent-shifting is visibly more detrimental for these countries. Sectors associated with fuel and energy production tend to have low markup margins. So natural resource-abundant countries specializing in these industries experience a shifting of markup rents from their economy to the rest of the world. These findings add a new perspective to the vibrant literature on the *resource curse* (e.g., Krugman, 1987; Lane and Tornell, 1996; Hodler, 2006; Mehlum et al., 2006; Van der Ploeg, 2011).

Figure 9: The effect of trade on misallocation: fuel-exporting countries



Note: This figure displays the inferred % change in sectoral misallocation, $\mathcal{D}_i$, in response to trade openness. For example, a change of 5% corresponds to a 5% increase in misallocation while a change of -5% corresponds to a 5% decrease. Fuel exporting countries are those whose fuel exports constitute more than 20% of national exports (United Nations Country Classification). These countries include Australia, Canada, Norway, Brunei, Kazakhstan Russia, and Saudi Arabia. The figures in the rows 1-4 are respectively computed using the formulas under Propositions 1-4. Data on industry-level expenditure, production and input-output shares are from the ICIO.

J The Evolution of International Rent-Shifting Patterns

The results presented in Section 7 revealed a dampening of rent-shifting from low- to high-income nations over time. Two primary factors can drive this pattern. First, low- and middle-income countries may have become more specialized in high-markup industries. Second, markup levels are evolving to favor the pattern of specialization in low- and middle-income nations.

Figure 10 examines these two possibilities by plotting the increase in the cost markups due to international rent-shifting, $\Delta\mathcal{D}_i$, and its evolution under different scenarios. The top panel corresponds to our benchmark result, and accounts for the longitudinal change in both markups and trade shares. The middle panel in Figure 10 isolates the contribution of markup changes to the evolution of $\Delta\mathcal{D}_i$. It plots $\Delta\mathcal{D}_i$ for each year holding trade shares constant at their 2005 level. The evolution of$\Delta\mathcal{D}_i$ in the middle panel, as a result, merely reflects the change in markups over time. The bottom panel in Figure 10 isolates the contribution of changes in trade shares. It plots $\Delta\mathcal{D}_i$ for each year, holding markups constant at their 2005 level. As such, the evolution of $\Delta\mathcal{D}_i$ in the bottom panel merely reflects the impact of changing production specialization over time.

Comparing the three cases in Figure 10 indicates that changes in the pattern of specialization and trade shares account for most of the dampening in low-to-high income rent-shifting. In other words, it appears that low- and middle income nations have become increasingly specialized in sophisticated, high-markup industries. These developments have, in turn, dampened the extent to which markup rents flow out of these economies to high-income trading partners.

K Duality Between International Rent-Shifting and Tariffs

The proof adopts the notation in Section 8, expressing welfare as an explicit function of tariffs and markups. Namely,

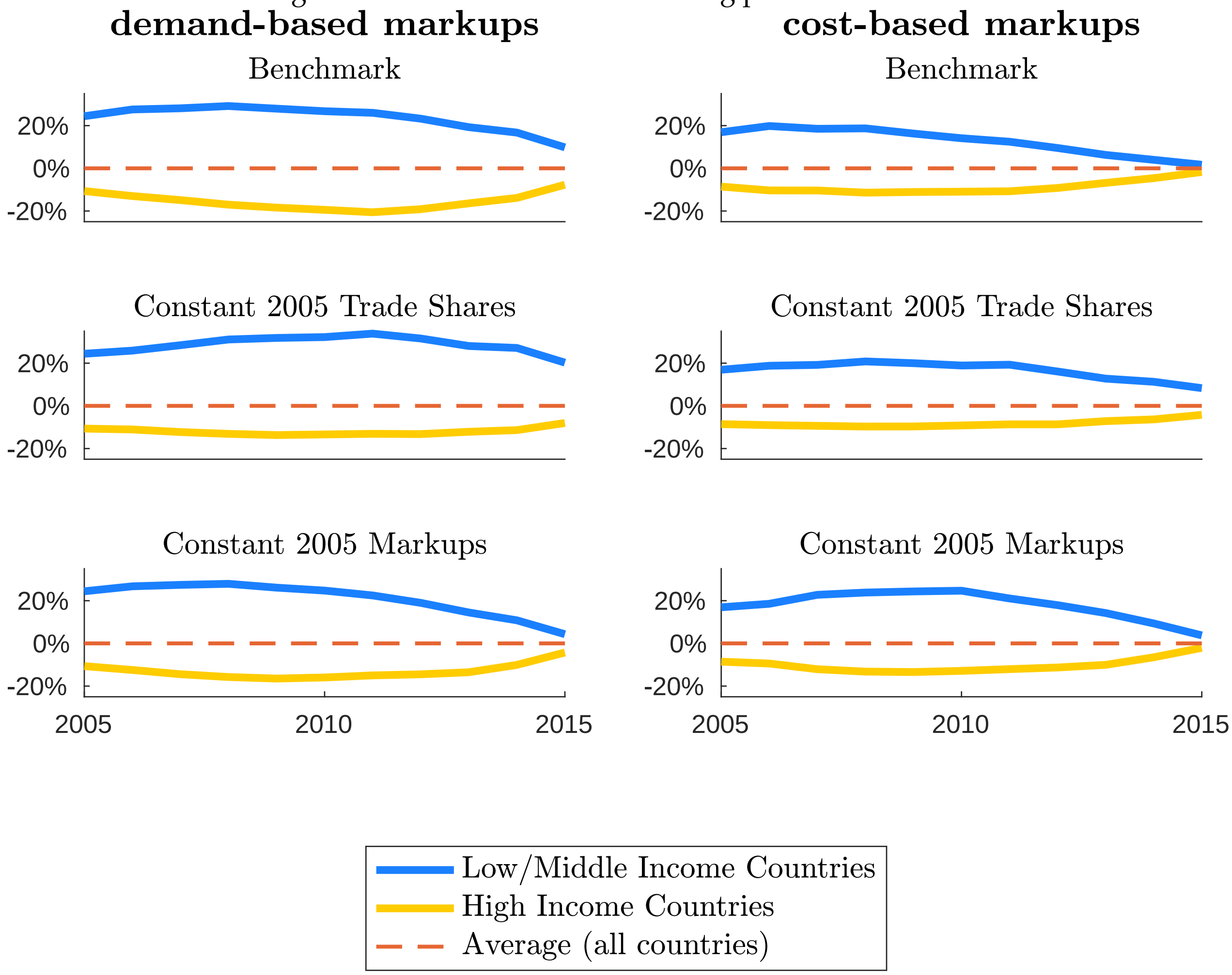
$$W_i = \mathcal{W}_i\left(\mathbf{t}, \boldsymbol{\mu}\right).$$

Our goal is to show that there exists a vector of tariffs that mimic the welfare impacts of markup distortion, including international rent-shifting. Stated formally,

$$\exists\ \mathbf{t}: \qquad \mathcal{W}_i\left(\mathbf{t}, \mathbf{0}\right) = \mathcal{W}_i\left(\mathbf{0}, \boldsymbol{\mu}\right).$$

Observe that markups are akin to industry-level production taxes that exhibit the same rate in all countries. Extrapolating from this observation, a uniform global tariff, $1 + \check{t}_{i,k} = \mu_k$, is more distortionary than equal rate markups from a global standpoint. That is, $\sum_i \mathcal{W}_i\left(\check{\mathbf{t}}, \mathbf{0}\right) \leq \sum_i \mathcal{W}_i\left(\mathbf{0}, \boldsymbol{\mu}\right)$. To elaborate, there may be countries for which $\mathcal{W}_i\left(\check{\mathbf{t}}, \mathbf{0}\right) > \mathcal{W}_i\left(\mathbf{0}, \boldsymbol{\mu}\right)$, since the welfare gains/losses from tariffs is the sum of tariff dispersion and wage-driven terms-of-trade effects—analogous to Equation 6 in the main text. The dispersion in tariffs, $\check{\mathbf{t}}$, is always greater and more costly than $\boldsymbol{\mu}$, since the tariff rate on the domestically produced goods is zero. But a country that experiences losses from profit-shifting under the markup schedule, $\boldsymbol{\mu}$, may experience terms-of-trade improvements under

Figure 10: The drivers of rent-shifting patterns over time



Note: The above graph reports the change in the cost of markups, $\Delta\mathcal{D}$, du to international rent-shifting. For example, a change of 5% corresponds to a 5% increase in the cost driven by rent-shifting. The top panel account for the change in both markups and trade shares over time. The middle panel computes $\Delta\mathcal{D}$ holding trade shares constant at their 2005 level. The bottom panel computes $\Delta\mathcal{D}$ holding markups constant at their 2005 level. In each case $\Delta\mathcal{D}$ is computed using the formula under Proposition 1. Data on industry-level expenditure, production and input-output shares are from the ICIO.

the tariff schedule, $\check{\mathbf{t}}$. In such cases, it is plausible that $\mathcal{W}_i(\check{\mathbf{t}}, \mathbf{0}) > \mathcal{W}_i(\mathbf{0}, \boldsymbol{\mu})$.

With the above background in mind, we establish three intermediate claims. First, following Lashkaripour and Lugovskyy (2021), the unilaterally optimal tariff in each country is uniform across industries if markups were eliminated to zero. Accordingly, for the vector of tariffs, $\check{\mathbf{t}} \equiv \{\check{t}_{i,k}\}_{i,k}$, there exists a uniform tariff equivalent, $\bar{\mathbf{t}} \equiv \{\bar{t}_i\}_i$, that preserves welfare in country i and is strictly lower than the optimal tariff rate. Second, relative to the efficient equilibrium benchmark, markups (i.e., production taxes) yield a strictly lower welfare for a country than the unilaterally optimal export tax, $x^*_{ij,k}$.[33] Also, following Lashkaripour and Lugovskyy (2021), the optimal export tax for an efficient

[33] Stated differently, given the rest of the world's tax schedule, replacing markups with the optimal export tax in country i is welfare-improving.

small open economy is

$$1 + x^*_{ij,k} = \frac{\sigma_k}{\sigma_k - 1} < \mu_k$$

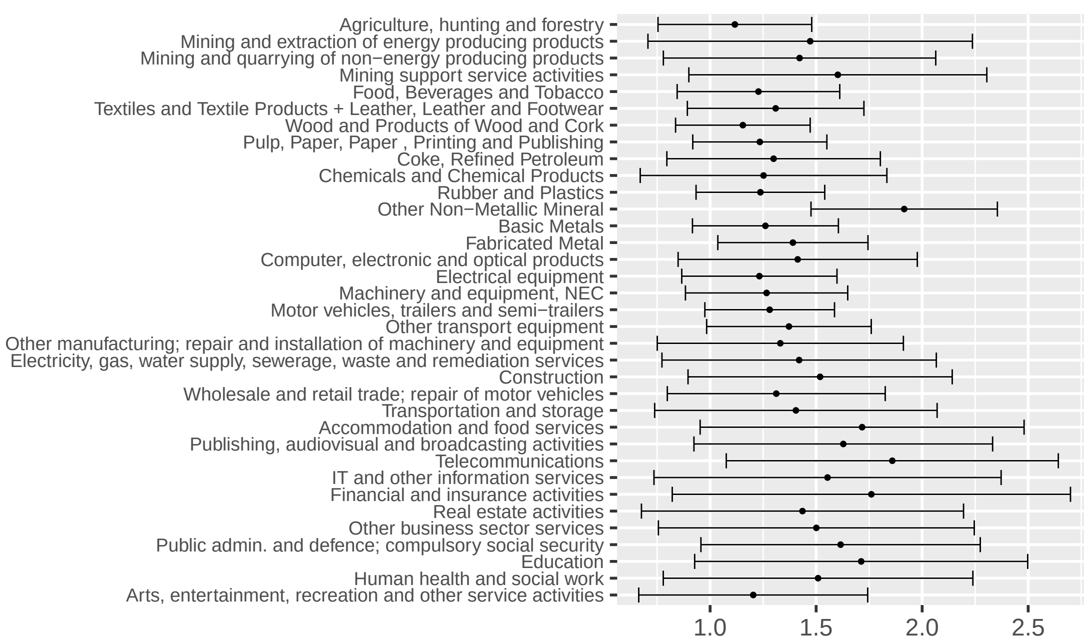
which is strictly smaller than the markup under the standard assumption that the degree of national-level market power exceeds firm-level market power, i.e., $\gamma_k > \sigma_k$. Third, if trade elasticities, σ_k, are uniform, then the optimal import tax is welfare-equivalent to the optimal export by the Lerner symmetry (Costinot and Werning (2019)). Combining these results, we can apply the Intermediate Value Theorem to the continuous function, $\mathcal{W}_i(.)$ to conclude the following. If trade elasticities are sufficiently homogeneous across industries, there exists a vector $\{a_i\}_i$, with $a_i \in (-\infty, 1)$, such that

$$\mathcal{W}_i\left(\{a_1\bar{t}_1, ..., a_N\bar{t}_N\}, \mathbf{0}\right) = \mathcal{W}_i\left(\mathbf{0}, \boldsymbol{\mu}\right),$$

where recall that $\bar{t}_i$ is the uniform tariff-equivalent of the tariff schedule, $\check{\mathbf{t}}_i$—which we defined earlier. This statement proves our initial claim, but under the assumption that applied tariffs are zero. It is straightforward to verify that the proof follows if global applied tariffs are sufficiently small (which is the case in the on real world) or sufficiently different from the unilaterally optimal schedule for individual countries.

L Additional Tables and Graph

Figure 11: Distributions of Firm-level Markups Across Industries



Note: The figure shows the within-industry distribution of firm-level markups in 2010. The mean markup of a given industry is graphed as a dot in the figure, while the error bars that extend below and above the mean markup represent one standard deviation below and above the mean, respectively.

Table 3: Country specific results: impact of international rent-shifting

Country	$\Delta\mathcal{D}$ — Demand-Based Markups				$\Delta\mathcal{D}$ — Cost-Based Markups			
	Baseline	Baseline + CES	Baseline + IO	Baseline + FS	Baseline	Baseline + CES	Baseline + IO	Baseline + FS
Australia	29.2%	17.8%	10.2%	36.1%	17.7%	16.6%	8.4%	23.1%
Austria	-9.7%	-8.2%	-5.2%	-20.2%	-1.6%	-7.4%	-3.7%	-7.7%
Belgium	-12.1%	-10.5%	-4.6%	-28.8%	10.2%	-9.4%	1.4%	9.9%
Canada	20.1%	6.4%	8.9%	18.7%	19.1%	5.8%	10.8%	20.2%
Chile	48.5%	31.2%	23.6%	50.4%	37.9%	29.0%	25.9%	43.4%
Czech Republic	-5.6%	-11.5%	-6.7%	3.3%	7.1%	-10.4%	-4.4%	4.2%
Denmark	4.7%	2.0%	1.3%	-18.0%	31.7%	1.9%	13.0%	30.5%
Estonia	14.9%	5.3%	0.9%	15.1%	-5.8%	5.0%	-5.2%	-13.3%
Finland	5.0%	1.1%	-3.7%	-10.3%	5.4%	1.0%	-3.4%	-4.7%
France	-6.5%	-7.3%	-3.7%	-4.3%	-1.5%	-6.7%	-2.6%	0.9%
Germany	-10.2%	-12.2%	-6.3%	-23.2%	20.6%	-10.9%	4.6%	16.5%
Greece	5.7%	-0.0%	7.2%	9.7%	-10.0%	-0.0%	-1.0%	-11.5%
Hungary	-3.6%	-12.8%	-5.3%	12.7%	22.4%	-11.4%	4.1%	28.6%
Iceland	3.3%	-0.7%	-1.7%	-11.8%	27.8%	-0.6%	7.5%	38.7%
Ireland	-34.6%	-25.6%	-34.5%	-166.0%	167.4%	-21.6%	33.7%	167.2%
Israel	-26.6%	-20.7%	-6.1%	-25.2%	-21.0%	-18.7%	-7.0%	-20.2%
Italy	-0.5%	-2.4%	-0.5%	-12.3%	1.1%	-2.2%	-0.0%	-6.8%
Japan	-10.8%	-10.4%	-5.6%	-7.8%	-2.2%	-9.4%	-3.9%	-2.7%
Korea	-24.3%	-22.4%	-15.9%	-24.5%	-0.0%	-20.3%	-12.0%	-1.7%
Latvia	8.8%	5.6%	0.4%	17.9%	-14.4%	5.3%	-7.8%	-17.8%
Lithuania	-8.3%	-6.8%	-7.5%	-27.2%	6.4%	-6.4%	-2.8%	0.2%
Luxembourg	43.4%	36.4%	1.9%	41.2%	37.5%	34.6%	1.7%	36.2%
Mexico	13.1%	3.1%	6.5%	42.5%	8.9%	3.0%	7.4%	21.0%
Netherlands	4.6%	-0.5%	-6.8%	-15.2%	28.9%	-0.5%	1.9%	26.7%
New Zealand	-18.0%	-12.6%	-6.3%	-32.9%	4.6%	-11.6%	-0.9%	8.1%
Norway	76.4%	45.0%	47.3%	71.5%	67.2%	42.4%	53.8%	72.0%
Poland	-4.7%	-6.4%	-0.5%	-5.5%	-3.3%	-6.0%	-0.5%	-4.8%
Portugal	7.6%	1.6%	3.5%	12.3%	-17.1%	1.5%	-5.8%	-23.8%
Slovak Republic	-9.7%	-15.7%	-10.6%	3.3%	7.4%	-14.5%	-8.0%	6.2%
Slovenia	2.5%	-2.3%	0.0%	-5.6%	16.5%	-2.1%	6.1%	11.8%
Spain	-9.7%	-9.1%	-1.9%	-8.5%	-4.4%	-8.5%	-1.2%	-4.5%
Sweden	-1.3%	-3.2%	-4.3%	-13.7%	10.3%	-2.9%	-0.1%	3.9%
Switzerland	-21.1%	-19.4%	-7.6%	-34.2%	3.8%	-17.3%	0.2%	4.2%
Turkey	11.4%	7.4%	5.8%	26.3%	-14.9%	7.2%	-3.2%	-12.7%
United Kingdom	-25.8%	-19.8%	-5.5%	-17.0%	-32.2%	-18.6%	-9.3%	-28.8%
United States	-12.1%	-13.7%	-2.8%	-9.3%	-14.3%	-12.4%	-4.9%	-13.4%
Argentina	-0.4%	-0.2%	0.1%	-3.1%	8.7%	-0.2%	4.1%	11.6%
Brazil	11.3%	2.2%	4.4%	11.2%	10.7%	2.0%	6.0%	11.2%
Brunei Darussalam	104.4%	68.0%	67.9%	103.9%	83.6%	63.3%	67.0%	91.8%
Bulgaria	20.1%	11.6%	-4.6%	36.0%	-10.7%	10.7%	-15.3%	-9.1%
Cambodia	20.7%	21.1%	19.3%	49.1%	42.2%	25.0%	33.5%	60.9%
China	0.3%	-2.0%	-4.5%	-1.5%	-7.6%	-1.8%	-7.8%	-11.9%
Colombia	23.6%	12.2%	15.5%	31.7%	7.6%	11.3%	12.2%	12.8%
Costa Rica	-9.9%	-8.5%	3.9%	4.4%	-12.2%	-7.9%	2.5%	-0.4%
Croatia	-15.2%	-13.8%	1.9%	-3.4%	-31.4%	-13.0%	-5.9%	-29.2%
Cyprus	-3.9%	-6.6%	3.0%	14.6%	-23.4%	-6.2%	-4.5%	-16.3%
India	-18.2%	-12.8%	-5.4%	-15.5%	-24.2%	-11.6%	-11.0%	-26.5%
Indonesia	12.9%	9.2%	7.0%	15.3%	9.2%	9.0%	7.6%	13.6%
Hong Kong, China	-30.5%	-22.9%	4.8%	-11.1%	-54.3%	-22.8%	-4.9%	-50.1%
Kazakhstan	94.6%	64.9%	45.2%	100.1%	77.1%	59.9%	44.5%	85.6%
Malaysia	-10.8%	-13.4%	-3.6%	-12.1%	-2.8%	-12.2%	-2.6%	-0.6%
Malta	43.8%	27.0%	2.5%	50.8%	19.4%	24.9%	-0.4%	22.7%
Morocco	-22.7%	-20.1%	-2.1%	-6.1%	-32.4%	-18.4%	-10.2%	-22.7%
Peru	21.2%	12.8%	13.4%	24.0%	9.7%	12.3%	11.1%	13.4%
Philippines	-8.8%	-9.8%	0.4%	6.6%	-25.2%	-9.8%	-7.9%	-20.7%
Romania	24.1%	14.6%	7.2%	46.9%	-4.0%	14.1%	-0.1%	-0.2%
Russian Federation	47.9%	27.7%	20.2%	36.2%	45.5%	25.6%	25.5%	42.0%
Saudi Arabia	71.3%	39.0%	74.2%	57.9%	67.9%	35.5%	79.8%	72.5%
Singapore	5.2%	0.6%	-22.4%	-22.8%	49.5%	0.5%	-11.5%	46.8%
South Africa	5.5%	-4.8%	3.2%	6.5%	4.2%	-4.4%	3.2%	5.1%
Chinese Taipei	-38.9%	-30.6%	-18.1%	-32.4%	-33.3%	-26.8%	-23.5%	-36.3%
Thailand	-44.3%	-33.0%	-14.7%	-44.3%	-27.2%	-33.2%	-15.5%	-26.2%
Tunisia	-12.0%	-14.6%	-4.7%	-4.2%	-23.6%	-13.8%	-12.0%	-19.4%
Viet Nam	42.4%	28.2%	6.3%	55.0%	4.8%	30.8%	0.3%	13.3%
Average	-0.5%	-4.0%	0.0%	-0.6%	0.8%	-3.6%	0.3%	0.8%

Figure 12: The cost of markups under different levels of cross-industry substitutability

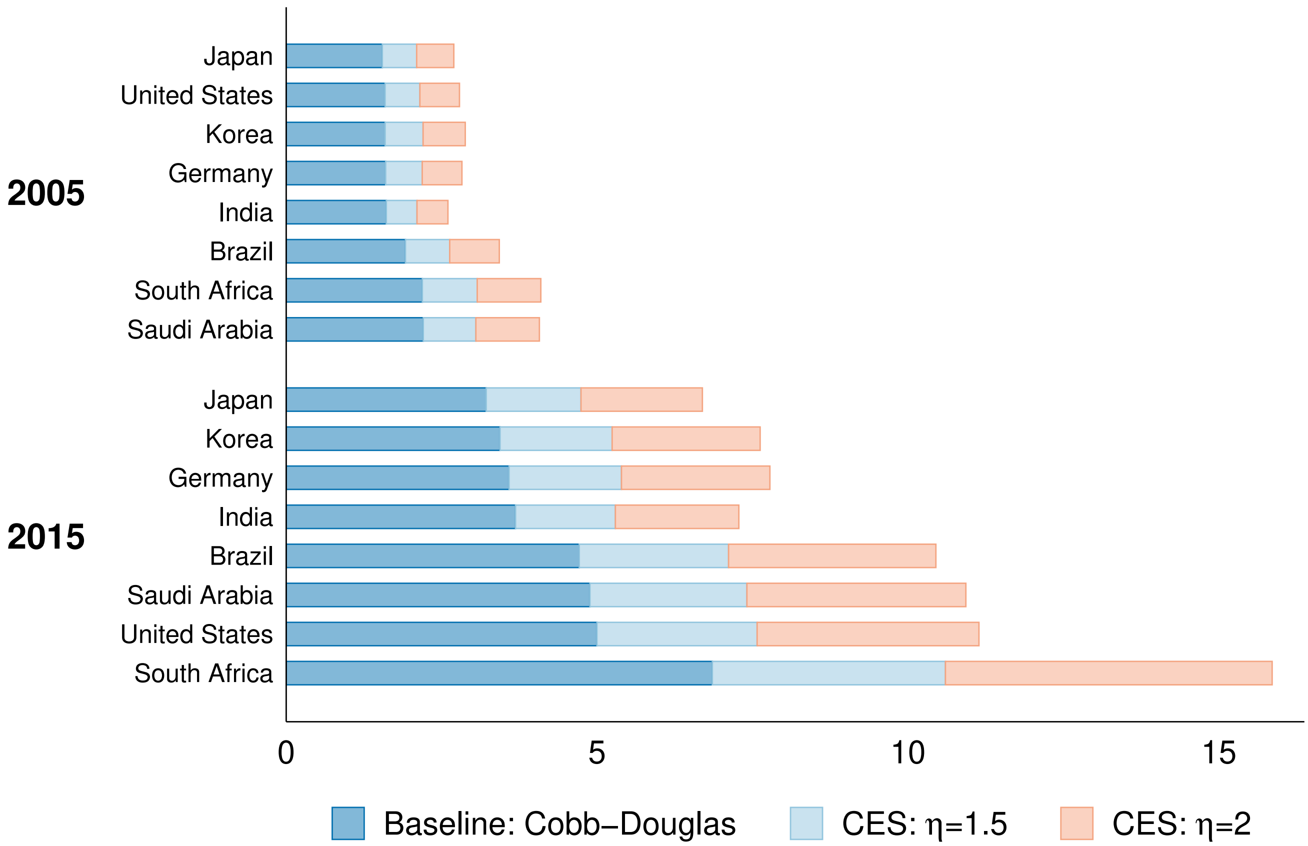
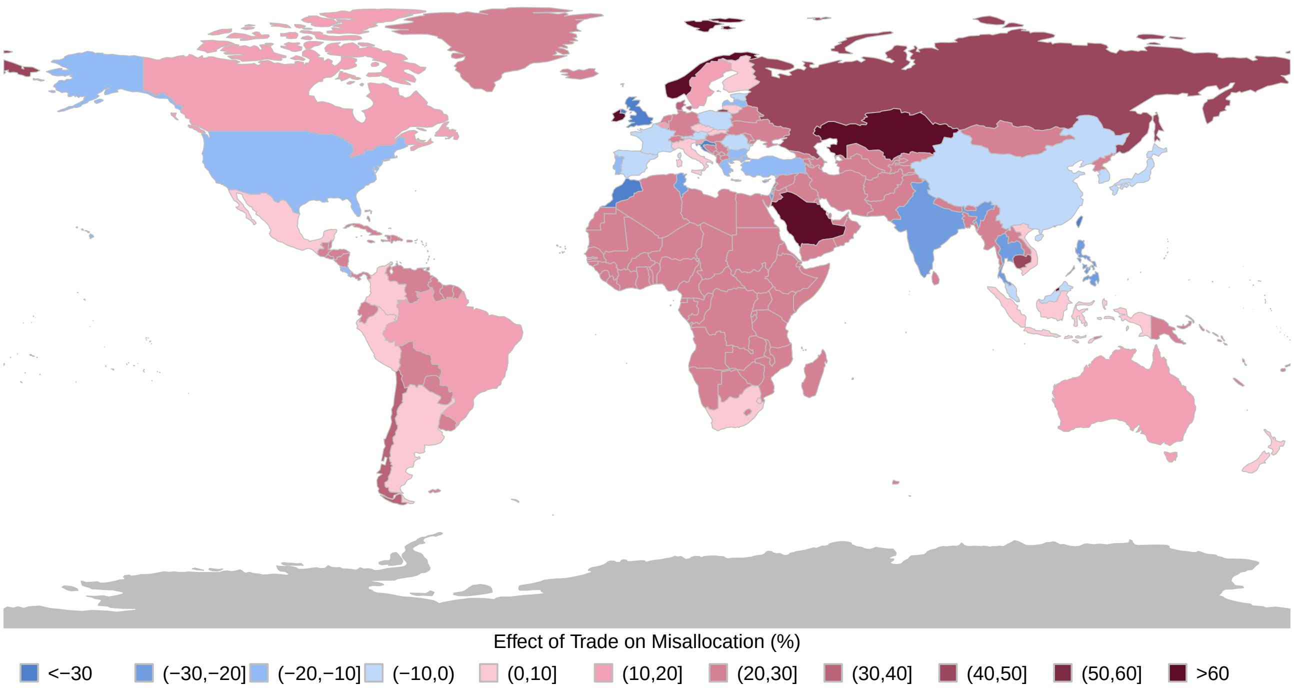


Figure 13: The welfare impacts of international rent-shifting



Note: The map displays the per-cent change in national welfare due to international rent-sifting ($\triangle D_i$) for different countries in 2015. The reported values are calculated using Proposition 1. The markups levels are taken from our cost-based estimation. Data on national and industry-level revenue and expenditure shares are from ICIO.