# A Global Perspective on the Incidence of Monopoly Distortions*


Siying Ding — UIBE

Ahmad Lashkaripour — Indiana University

Volodymyr Lugovskyy — Indiana University


June 2024


**Abstract**

We develop a method to measure the incidence of monopolistic markup distortions in the global economy. Using semi-parametric formulas, we measure how trade modifies the deadweight loss of markups through two channels: (1) trade-induced change in markup dispersion, and (2) international *rent-shifting*. The latter, which has received less attention in the literature, constitutes zero-sum welfare effects similar to implicit tariffs that tilt the terms of trade in favor of countries exporting high-markup goods. To measure these effects, we estimate firm-level markups globally using demand-based and cost-based methods and compile new data on global profit ownership. Our findings reveal that trade has greatly reshaped the global incidence of monopoly distortions, reducing the deadweight loss of markups for high-income countries by 15% while increasing it by 44% among low-income nations. These asymmetric effects are primarily due to international rent-shifting and represent a 8% implicit tariff imposed by high-income countries on low-income partners. These findings are robust to accounting for global input-output linkages and fixed investment costs. Our results challenge the prevailing view that high-income countries have made disproportionately greater concessions under existing trade agreements. We show that rent-shifting externalities can be effectively mitigated through additional preferential tariff concessions under the WTO or a globally coordinated destination tax on profits.


## 1 Introduction

The past few decades were marked by two notable trends in the global economy: a substantial growth in international economic integration and a simultaneous rise in monopolistic markups, indicating heightened levels of market power. These developments pose two fundamental questions for economists:

1. to what extent has trade integration mitigated the deadweight loss of monopoly distortions on a global scale through pro-competitive pressures?
2. has trade integration resulted in an international shift in the burden of markup distortions?


*This paper was previously circulated as *"Who Pays for Markups in a Global Economy?"* in August 2021. We are grateful for comments and suggestions from Carsten Eckel, Alex Fakos, Sharat Ganapati, Colin Hottman, Isabel Mejean, Esteban Rossi-Hansberg, Kim Ruhl (discussant), Daniel Xu, Feodora Teti, and seminar participants at Georgetown University, Pennsylvania State University, NYU Abu Dhabi, CESifo Global Economy Conference, EIIT, Midwest Trade, and SED Meetings. All errors are our own.

While the existing literature has provided valuable insights into the first question, it offers limited guidance on the second. Theoretical studies that investigate the impact of trade in markup-distorted economies typically emphasize the pro-competitive effects of trade, which are internationally symmetric. Empirical studies, meanwhile, often focus on evaluating the effects within individual countries and cannot address the question of asymmetric trade effects across countries. Consequently, the question of how trade has shifted the incidence of monopoly distortions across countries remains largely unexplored, presenting a gap in the literature.

To bridge this gap, we develop a sufficient statistics methodology to measure the *unequal* incidence of monopolistic markup distortions in the global economy. We first derive semi-parametric formulas for the deadweight loss of markup distortions, isolating the extent to which trade integration has redistributed the burden of these distortions across countries. We then use these formulas to conduct measurement, which involves estimating firm-level markups at a global scale and assembling new data on global profit ownership.

Our analysis reveals that international trade has greatly shifted the burden of monopolistic markup distortions, reducing the deadweight loss for high-income countries while exacerbating it among low-income nations. These asymmetric effects are a result of international rent-shifting: trade creates a disconnect between the locations where monopolistic markups negatively impact consumer surplus and where the monopoly rents (excess profits) are distributed. Consequently, households in high-income countries, the main recipients of monopoly rents, benefit, while households in low-income nations lose out. We demonstrate that these rent-shifting externalities are analogous to implicit tariffs that shift the terms of trade in favor of high-income countries, challenging the view that high-income countries have made disproportionately greater tariff concessions under the WTO. We propose two global policy solutions to address rent-shifting externalities: one that is compatible with WTO rules and another that can be integrated into the evolving global minimum tax agreement.

Section 2 lays the groundwork by presenting the basic calculus of monopoly distortions in open economies. This is followed up by documenting a series of suggestive evidence in Section 3. First, we document a stark divergence in aggregate accounting profit margins between low- and high-income countries, despite the rate of R&D investment and firm entry remaining synchronized across these nations. Second, we show that this development is concomitant with export-led specialization of high-income countries in industries characterized by high profit margins. As a result, high-income countries now predominantly export goods with high profit margins to low-income countries, while importing goods with lower profit margins in return. Lastly, although profits are primarily accrued at the firms' physical location, we show that a non-trivial portion of profits is repatriated to foreign shareholders, predominantly to high-income countries. Taken together, these data regularities point to systematic North-South asymmetries in exposure to monopolistic distortions.

To measure these effects formally, we propose a non-parametric model of the global economy with markup distortions in Section 4 and derive closed-form formulas for the deadweight loss of markup distortions in Section 5. We show that the deadweight loss (DWL) of markups can be decomposed into

($i$) the welfare loss due to markup dispersion and ($ii$) international rent-shifting externalities. Namely:

$$\text{DWL} \approx \underbrace{\text{MLD}\left(\frac{1}{\mu}\right)}_{\text{markup dispersion}} + \underbrace{\ln\left(\frac{\text{avg expenditure-side markup}}{\text{avg ouput-side markup}}\right)}_{\text{international rent-shifting}}.$$

The first component, MLD $(1/\mu)$, represents *markup dispersion*, measured by the mean log deviation (MLD) of inverse markups $1/\mu$ in the entire economy. Markup dispersion is the only source of welfare loss in a closed economy or in open economies with internationally symmetric firm-level markups. And the *pro-competitive* effects emphasized in the previous literature (e.g., Melitz and Ottaviano (2008); Edmond et al. (2015); Arkolakis et al. (2019)) concern the change in MLD $(1/\mu)$ in response to trade.

The *international rent-shifting* component of the DWL represents a frequently overlooked externality arising from trade between nations with unequal market power. It is calculated as the ratio between the average expenditure-side markup (the excess markup a country pays on goods sourced domestically and internationally) and the average output-side markup (the excess markup a country collects on its output, sold domestically and internationally). This component indicates that the welfare loss from markup distortions is exacerbated for countries that are net importers of high-markup goods, as they incur higher excess markup payments than they collect.

The rent-shifting externalities constitute zero-sum welfare transfers between countries, stemming from a decoupling between where markup distortions are borne and where their financial benefits are realized. In a closed economy, markup rents are rebated to the same consumers whose surplus is negatively impacted by markups, precluding rent-shifting effects. However, in a global economy, the loss to consumer surplus may occur in one location while excess profits are rebated elsewhere. Countries that export high-markup goods capture net monopoly rents from global markets, experiencing reduced deadweight losses from markups, while countries specializing in low-markup industries pay net rents and endure disproportionately higher welfare losses.

Section 7 shows that rent-shifting externalities act as implicit tariffs, distorting the terms of trade in favor of countries that are net exporters of high-markup goods. This phenomenon is basically a *decentralized* form of terms-of-trade manipulation, wherein monopolistic firms distort prices in foreign markets to extract rents, which are ultimately rebated to their home country. Governments may be reluctant to regulate such anti-competitive practices if the resulting terms-of-trade gains outweigh domestic efficiency losses. Though overlooked, these considerations are essential when interpreting and applying the reciprocity principle within the WTO framework.

Measuring the deadweight loss of markups and the associated rent-shifting effects *non-parametrically* demands comprehensive data, which is unattainable—specifically, data on the full distribution of output and expenditure across admissible markup values for each country. To address this data constraint, we derive semi-parametric formulas that exactly characterize the deadweight loss of markups with less stringent data requirements. These formulas are valid under certain assumptions regarding market conduct, semi-parametric restrictions on demand, and parametric restrictions on the distribution of firm productivity. By utilizing these formulas, the data requirement is reduced to product-level expenditure and output data for each country, along with sales-weighted average markups within product categories.

Section 8 applies our semi-parmateric formulas to conduct welfare measurements. To this end, we compile international data on observable shares, such as output shares, expenditure shares, and multinational profits payment shares, as well as estimate firm-level markups on a global scale. The data on trade and production shares is sourced from the OECD Inter-Country Input-Output (ICIO) Tables, which span 36 industries representing the entire economy and cover 64 major countries along with an aggregate of the rest of the world. The ICIO data covers the period from 2005 to 2015. We supplement this data with original data on global profit ownership, constructed from the financial statements of multinational enterprises from the Orbis database.

We estimate firm-level markups globally using two methods: the cost-based method, as in De Loecker and Warzynski (2012), and the demand-based estimation method, as in Berry et al. (1995, BLP). For the cost-based estimation, we utilize a global sample of publicly-traded firms from the Worldscope database, similar to De Loecker and Eeckhout (2018). Large-scale implementations of BLP are more challenging and have less precedent in the literature. To address this challenge, we employ the computationally efficient linear approximation of BLP, proposed by Salanié and Wolak (2019), and leverage high-frequency transaction-level data to guide identification with limited information on observed product characteristics.

Our analysis reveals that the deadweight loss of markup distortions has increased over time, with markups eroding more than 7% of real global consumption in 2015. This rise in the deadweight loss is primarily driven by the increase in markup levels rather than changes in output compositions. Interestingly, some high-income countries, such as the Netherlands, benefit from markups, indicating that the positive effects of international rent-shifting outweigh the domestic efficiency loss caused by markup dispersion in these countries.

Our main finding is that trade integration has significantly reshaped the global incidence of monopoly distortions in a way that favors high-income countries. It has increased the burden of markup distortions for low-income countries by 44% while decreasing the burden for high-income nations by 15%. These asymmetric effects are driven by rent-shifting from low- to high-income countries, as low-income countries tend to specialize in less sophisticated, low-markup industries. This pattern of rent-shifting remains robust regardless of the method used to estimate firm-level markups and persists even after accounting for multi-national ownership, global input-output linkages, and fixed cost payments.

We show that international rent-shifting externalities are akin to an 8% implicit tariff unilaterally imposed by high-income countries on their low-income trading partners. This finding sheds fresh light on the current state of concessions within global trade agreements, challenging the growing narrative that high-income countries, such as the United States, have made disproportionately greater concessions under the status quo (Chow et al. (2018)). On a superficial level, high-income countries may appear to be making additional concessions and offering preferential treatment to their low-income counterparts under the WTO's Generalized System of Preferences (GSP). But in reality, rent-shifting externalities more than counteract the GSP concessions. After factoring in the implicit tariff due to rent-shifting externalities, high-income countries are effectively applying a 6% excess tariff on low-income partners.

We propose two policy solutions to address international rent-shifting externalities, acknowledging that the first-best solution, *i.e.*, internationally coordinated markup correction, may be politically infeasible and challenging to implement through existing WTO mechanisms. The first solution entails additional preferential tariff concessions by high-income countries under the GSP. While this solution has the potential to completely eliminate rent-shifting externalities and restore *strict* reciprocity under the WTO, its implementation may be complicated by the increasing fragility of the WTO system. The second solution involves a destination tax on profits under Pillar One of the Global Minimum Tax Agreement. This latter solution is only partially effective in addressing rent-shifting externalities, but may be more politically viable in the current climate.

**Related Literature.** Our theoretical results relate to a broad literature that formalizes the impacts of trade on (markup) distortions. These studies typically focus on (1) inter-firm reallocation and (2) pro-competitive effects of trade liberalization. Given this focus, existing studies generally measure how trade modifies the *dispersion* in markup or non-markup wedges relative to autarky. Regarding inter-firm reallocation, Epifani and Gancia (2011), Bai, Jin, and Lu (2019), Berthou, Chung, Manova, and Bragard (2020), Dix-Carneiro, Goldberg, Meghir, and Ulyssea (2021), and Farrokhi, Lashkaripour, and Pellegrina (2024) examine how trade-induced reallocation across firms with different productivity and distortion levels impacts aggregate productivity and welfare.[1] Regarding pro-competitive effects, Melitz and Ottaviano (2008), Holmes, Hsu, and Lee (2014), De Blas and Russ (2015), Edmond, Midrigan, and Xu (2015), Feenstra and Weinstein (2017), Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019), and Edmond, Midrigan, and Xu (2023) consider settings where firms adjust their markup in response to import competition.[2] More recently, Atkin and Donaldson (2021), Baqaee and Farhi (2019), and Edmond, Midrigan, and Xu (2023) integrate these effects into unified frameworks. The former two papers accommodate non-markup wedges, while the latter paper also considers the impact of firm entry. Compared to these studies, we emphasize the international rent-shifting externalities which has received less attention. We adopt a non-parametric approach, contributing to an emerging literature examining the non-parametric impacts of trade (Adao, Costinot, and Donaldson (2017); Adão, Arkolakis, and Ganapati (2020); Errico and Lashkari (2022)). Moreover, we show that under some conditions, there is duality between tariffs and markups—a new result with salient implications for global policy design.

[1] Dix-Carneiro et al. (2021) and Farrokhi et al. (2024) allow for endogenous technology choice, which introduces an additional margin of allocative efficiency adjustment to trade. Our paper is also tangentially related to the literature examining the gains from trade relative to autarky in distorted economies (e.g., Costinot and Rodríguez-Clare (2014); Święcki (2017); Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019)). See Atkin and Donaldson (2021) for a synthesis of this growing literature.

[2] The results on the size of the pro-competitive effects of trade are somewhat mixed. 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. In a related vein, Nocco, Ottaviano, and Salto (2014); Dhingra and Morrow (2019); Behrens, Mion, Murata, and Suedekum (2020); Mrázová, Neary, and Parenti (2021) examine how the distribution of markups and the degree of misallocation in open economies depends on the underlying demand and supply structure.

The trade policy literature has explored rent-shifting externalities, but in the context of strategic interactions among governments (Brander and Spencer (1985); Eaton and Grossman (1986); Mrázová (2011); Ossa (2012); Bagwell and Staiger (2012); Head and Spencer (2017); Lashkaripour (2021); Firooz and Heins (2023); Lashkaripour and Lugovskyy (2023)). The central idea is that restoring allocative efficiency requires external interventions to reallocate resources towards producers with higher than average excess profit margins. When governments are unable to achieve this through domestic policies, they might resort to import tariffs or export subsidies. Such border policies may worsen the terms of trade in order to grant greater market access to high-profit domestic firms. This improves allocative efficiency at home but worsens it abroad, thereby creating a rent-shifting (or profit-shifting) externality. Our contribution to this literature is to highlight the reverse aspect of this arrangement: We demonstrate that monopolistic pricing practices by firms generate such large implicit terms of trade gains for some countries that they outweigh the associated allocative efficiency losses. As a result, governments may purposefully avoid regulating anti-competitive practices to preserve these implicit terms of trade benefits. This finding has important implications for reciprocity and tariff concessions under the WTO, contributing to quantitative assessments of reciprocity as in Bown, Parro, Staiger, and Sykes (2023).

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-based approach, paying particular attention to the challenges of production function estimation 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 markups 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. Keller and Yeaple (2020) examine the rise of market power from the lens of US multinational firms' sorting decisions. Coşar, Grieco, Li, and Tintelnot (2018) estimate markups across multiple international markets for the auto industry using the Berry et al. (1995) approach. However, large-scale demand-based estimation of markups across multiple industries remains limited due to computational constraints. Beyond basic markup estimation, Caliendo, Parro, and Tsyvinski (2022) develop a novel sufficient statistics approach to infers global wedges from the input-output structure.[3] Our contribution to this literature is twofold. First, we estimate both demand- and cost-based markups across a wide range of industries and countries, demonstrating that both methods yield remarkably similar macroeconomic implications. Second, while existing studies typically focus on the rise in average markups, we construct aggregate welfare indexes for the deadweight loss of markups and show that these indexes have risen globally over time.[4]

[3]While we focus on *output* market power, several papers have emphasized other aspects of market power. Hottman (2017) estimates retail markups for the United States and documents significant spatial variation. Alviarez, Fioretti, Kikkawa, and Morlacco (2023); Dhyne, Kikkawa, and Magerman (2022); Morlacco (2019) examine input market power and two-sided market power in international buyer-seller relationships. Hummels, Lugovskyy, and Skiba (2009); Ardelean and Lugovskyy (2023) examine market power in the international shipping industry.Asker, Collard-Wexler, and De Loecker (2019) estimate misallocation in global oil extraction by contrasting factual cost curves to undistorted counterfactual supply curves.

[4]Edmond et al. (2023) present a similar result for the US economy. Our result about the rising welfare cost of markups is

## 2 The Basic Calculus of Monopoly Distortions in Open Economies

Before presenting our general model of the global economy, it is helpful to discuss the basic calculus behind the welfare effects of monopoly distortions. Consider a simple two-sector, two-country economy consisting of North ($N$) and South ($S$). Assume one sector is perfectly competitive, while the other features monopolistic firms that charge a markup $\mu$ over their marginal cost $MC$, resulting in a price $P = \mu \times MC$. In the standard closed economy framework, the deadweight loss from monopoly pricing for each country $i = N, S$, denoted by $\mathscr{D}_i$, is calculated as the reduction in consumer surplus $\Delta CS_i$ minus the monopoly rent rebates.[5] This can be expressed as:

$$\mathscr{D}_i^{(closed)} = \Delta CS_i - \underbrace{\frac{\mu}{\mu - 1} PQ_i}_{\text{markup rents}} \equiv \Delta_i,$$

where $\Delta_i$ represent the Harbinger triangle. The intuition is that, from a macroeconomic standpoint, monopoly rents are ultimately redistributed back to consumers as supplementary income. Thus, the loss in consumer surplus due to monopoly markups is partially compensated by these rent transfers, resulting in a deadweight loss equal to the Harberger triangle. The top panel of Figure 1 illustrates these closed economy welfare effects for both countries.

In an open economy setting, the simple welfare calculus of monopoly pricing breaks down because monopoly rents are no longer redistributed within the same location where they create distortions and reduce consumer surplus. To illustrate this, let $Q_{ii'}$ represent the quantity of goods purchased by consumers in location $i'$ from producers in location $i$. Assume that there is two-way trade in the non-competitive (high-markup) sector, but the North has a revealed comparative advantage in this sector, manifested in positive net exports. Formally, this situation corresponds to:

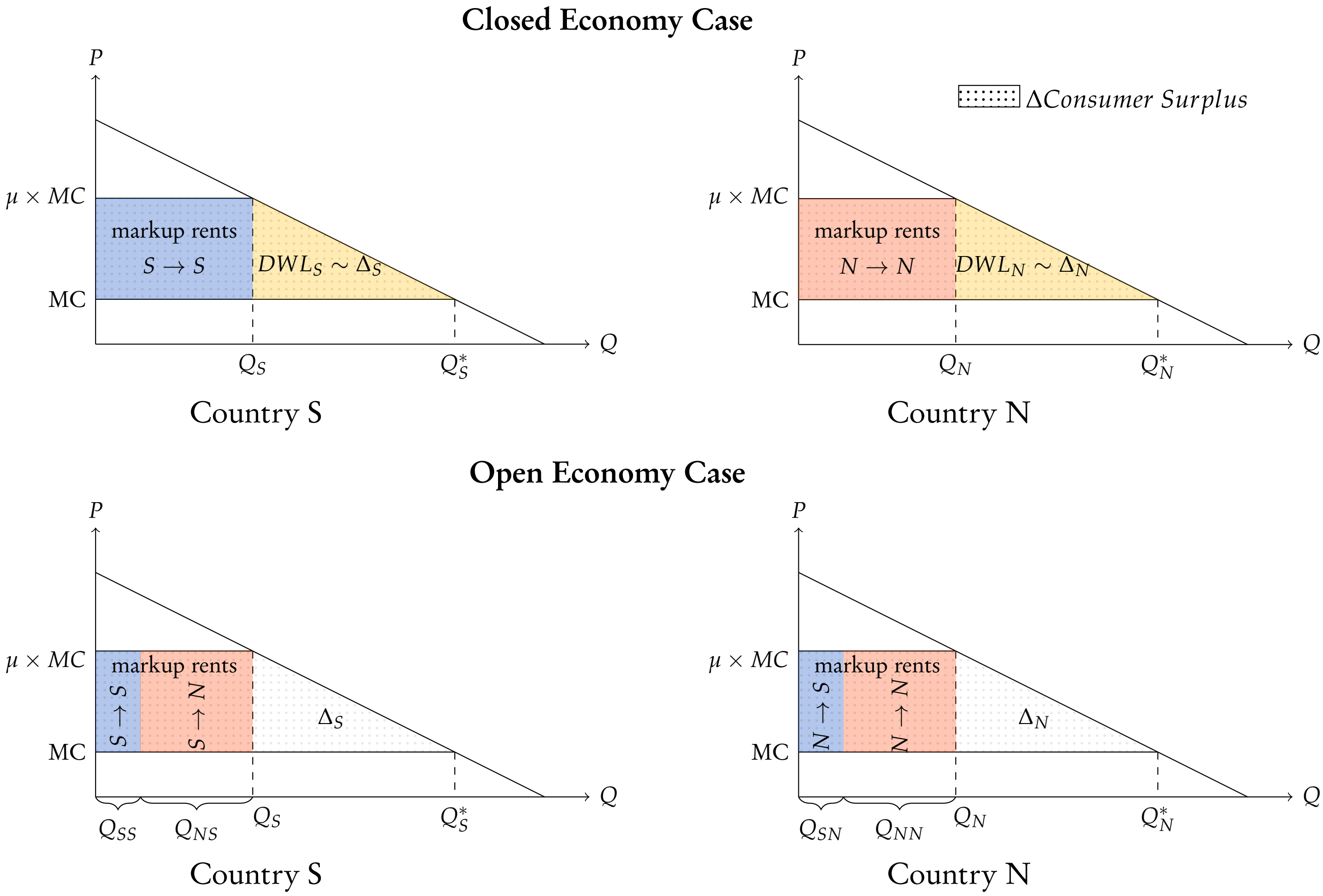
$$Q_{NS} > Q_{SN} \qquad (N \text{ is a net exporter of high-markup goods to } S),$$

This pattern could simply arise from a higher number of monopolistic firms operating in the North compared to the South. The bottom panel in Figure 1 illustrates the welfare effects of monopoly pricing in the open economy scenario. Assuming that profits earned by firms located in country $i$ are rebated to consumers in that country, the loss in consumer surplus in country $i$ is now offset by rents collected from both domestic and foreign sales. This leads to a locational decoupling between the payment of markup rents and the loss in consumer surplus due to markup pricing. Consequently, the deadweight loss from monopolistic markups no longer equals the Harberger triangle, $\Delta_i$. Instead, the losses are either greater or smaller than $\Delta_i$ depending on whether country $i$ is a net payer or recipient of markup rents to/from the other country. More formally, we can express the deadweight loss for countries $N$

subject to important caveats. While we account for firm entry and the scale externalities associated with it, we do not account for cost-reducing investment and internal economies of scale, as emphasized by Ganapati (2021) and Hsieh and Rossi-Hansberg (2023).

[5] Throughout this paper, we use "rents" to denote *excess* profits, which correspond to total firm profits from selling at a markup over marginal cost minus total investments in overhead costs – similar to Aghion et al. (2022).

Figure 1: The welfare consequences of monopoly distortions in a two-country economy




*Note*: This figure describes the welfare effects of markup wedges in a two-country and two-sector economy. Firms charge a constant markup $\mu$ over marginal cost MC in the sector for which demand and supply curves are displayed. The other sector is efficient, with goods priced at marginal cost. $Q_{ii'}$ denotes the quantity of markup-distorted goods produced in location $i$ and sold to location $i'$. $Q_i^*$ denotes the efficient quantity of consumption in location $i$.

(North) and $S$ (South) as:

$$\mathscr{D}_S = \Delta_S + \underbrace{\frac{\mu - 1}{\mu} P \left( Q_{NS} - Q_{SN} \right)}_{\text{rent-shifting}} > \Delta_S \qquad\qquad \mathscr{D}_N = \Delta_N - \underbrace{\frac{\mu - 1}{\mu} P \left( Q_{NS} - Q_{SN} \right)}_{\text{rent-shifting}} < \Delta_N.$$

Clearly, openness to trade in markup-distorted economies creates zero-sum welfare effects, which we refer to as *international rent-shifting*. In our example, rent-shifting effects reduce the deadweight loss of markup distortions for country $N$, while increasing the deadweight loss for country $S$. Perhaps surprisingly, these zero-sum welfare effects have received less attention in the literature compared to say pro-competitive effects. However, they are a crucial consideration in global policy design, as we will elaborate on shortly. Measuring international rent-shifting effects in a credible way requires a more comprehensive model that accounts for micro-to-macro general equilibrium effects and international profit payments. The following two sections provide the theoretical foundation for such measurement.

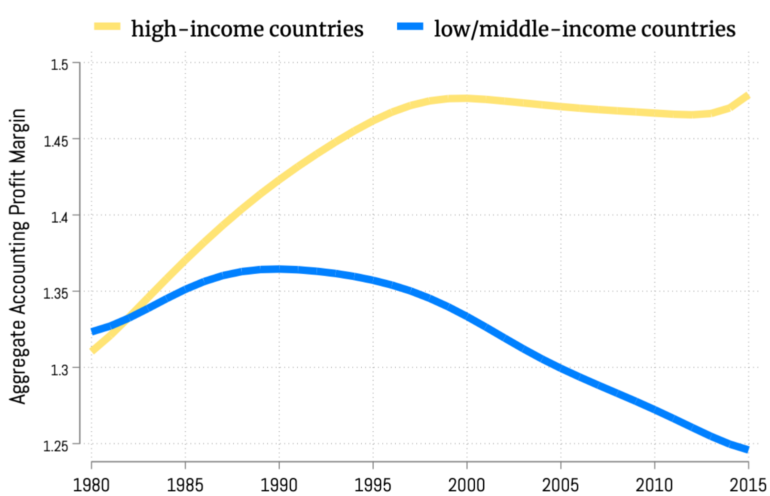
# 3 Suggestive Empirical Evidence

The international rent-shifting effects previously discussed arise when countries trade under asymmetric levels of market power, with some countries charging systematically higher markups than their trading partners. This section presents three stylized facts that hint at such asymmetries. The first pattern highlights diverging trends in aggregate accounting profit margins across low and high-income countries. The second pattern reveals that these trends are consistent with North-South specialization across low- and high-profit industries. The final fact demonstrates that the majority of profits are rebated within a firm's country of origin or to shareholders in high-income regions, suggesting that multinational profit payments exacerbate rent-shifting from low- to high-income countries.

**Fact 1.** *Aggregate* accounting profit margins have diverged between high- and low-income countries despite their rates of fixed capital formation and R&D growth remaining synchronized.

Figure 2 illustrates the trend in aggregate accounting profit margins for low and high-income countries between 1980 and 2015. These margins are computed as the ratio of sales to cost consolidated across all establishments within a country and industry. These measures are taken from the UNIDO-INDSTAT database, covering 196 countries and 23 ISIC rev.3 industries. The graph reveals that high-income economies, defined as countries in the top quartile of the GDP per capita distribution, experienced an upward trend in aggregate profit margins during this the 1980-2015 period. In contrast, low- and middle-income countries saw a decline in their aggregate profit margins.

Figure 2: North-South divergence in accounting profit margins




*Note*: the data is from UNIDO-INDSTAT. Aggregate accounting profit margins are calculate as the weighted average of sales-to-cost ratios across all ISIC industries. High-income countries are those in top quartile of the GDP per capita distribution. Low and middle income countries are classified as those in the bottom three quartiles of the distribution.

The divergence in aggregate profit margins across high and low/middle income countries does not coincide with a corresponding divergence in R&D expenditure or fixed capital formation, suggesting that higher accounting profits in high-income regions cannot be attributed solely to these investment factors. According to the United Nations' UIS database, the ratio of R&D expenditure to GDP has remained relatively stable between the two groups of countries during the same period. Investment trends can be analyzed at an even more granular level by examining the UNIDO-INDSTAT data. This dataset reports (*a*) fixed capital formation, which encompasses R&D by incumbent firms, at the industry level, and (*b*) firm entry dynamics, which captures the R&D associated with establishing new firms. Figure A4 in the appendix presents the longitudinal trends in fixed capital formation per worker and the number of establishments per industry. Neither of these indicators hint at a possible divergence in R&D expenditure that would be consistent with the observed divergence in accounting profit margins.

Considering the relative stability of R&D spending and fixed capital formation, the divergence in profit margins shown in Figure 2 likely reflects a divergence in *excess* markups. One potential driver is that firm-level markups have evolved asymmetrically across high-income and low/middle-income countries. For example, firm-level markups may have decreased in low-income countries due to heightened competition, while increasing in high-income regions as a result of cost reduction strategies. Another possible driver is increasing North-South specialization across industries with varying profit margins. Although determining the relative importance of each factor requires a model-based analysis, such as the one performed in Section 8 of this paper, an initial examination of the data suggests that inter-industry specialization plays a possible role.

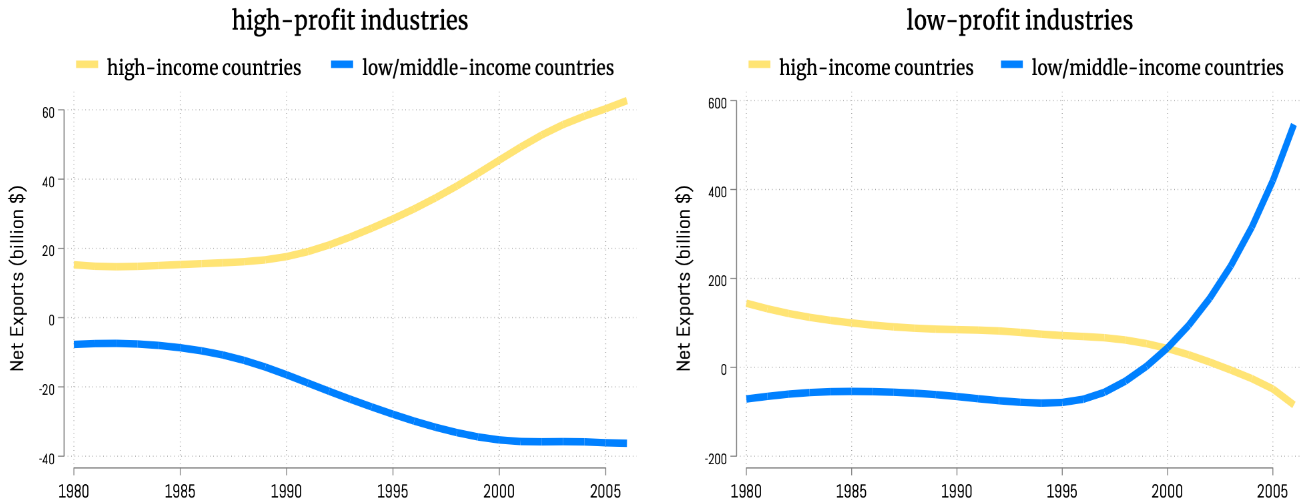
**Fact 2.** The North-South divergence in aggregate profit margins coincides with high-income economies, like the US, becoming increasingly specialized in high-profit industries.

We provide evidence for this fact using two data sources: First, we use internationally representative but industry-level data from CEPII's TRADEPROD database. Second, we use US-specific firm-level data from COMPUSTAT NORTH AMERICA to establish this fact at a more granular level. The CEPII's TRADEPROD database supplements the manufacturing segment of the UNIDO-INDSTAT data with corresponding information on import and export values from 1980 to 2005, allowing examination of export activity across low and high-profit industries. For each country in the sample, we calculate *net* exports within an ISIC rev.2 manufacturing industry by subtracting imports from exports in that industry. High-profit industries are defined as those with an accounting profit margin in the top 25% of all manufacturing industries. Figure 3 illustrates the contrasting trends in exports between low and high-income countries from 1980 to 2005. High-income countries are net exporters in high-profit manufacturing industries, and over time, their manufacturing exports have become increasingly concentrated in high-profit industries with the opposite trend occurring in low/middle income nations. These observations indicate that the North-South divergence in aggregate profit margins can be partially attributed to diverging patterns of specialization across industries.

For the United States, we can demonstrate this trend using more granular firm-level data from COMPUSTAT NORTH AMERICA.[6] By sorting industries based on their accounting profit margins, which are derived from firm-level financial accounts data, we can examine the distribution of US production activity across industries in different profit percentiles. Our analysis shows that, concurrent with increasing

[6]See Appendix A for a detailed description of the COMPUSTAT data and other datasets used in our analysis.

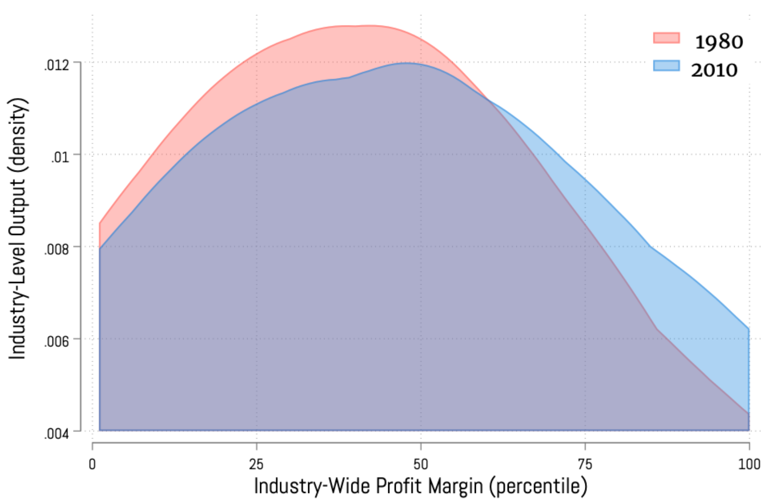
Figure 3: Net exports within high- versus low-profit manufacturing industries




*Note*: the data is from CEPII's TRADEPROD and covers manufacturing industries. High-profit industries are those with an accounting profit margin in the top quartile among all manufacturing industries. High-income countries are those in top quartile of the GDP per capita distribution. Low and middle income countries are classified as those in the bottom three quartiles of the distribution.

trade openness, the US economy has become progressively more specialized in high-profit margin industries. Figure 4 depicts this trend, illustrating that from 1980 to 2010, production activity among US firms has become increasingly concentrated in industries with high profit margins. All in all, the patterns suggest that the North-South divergence in profit margins is presumably due to inter-industry specialization.

Figure 4: The US economy has become increasingly specialized in high-profit industries

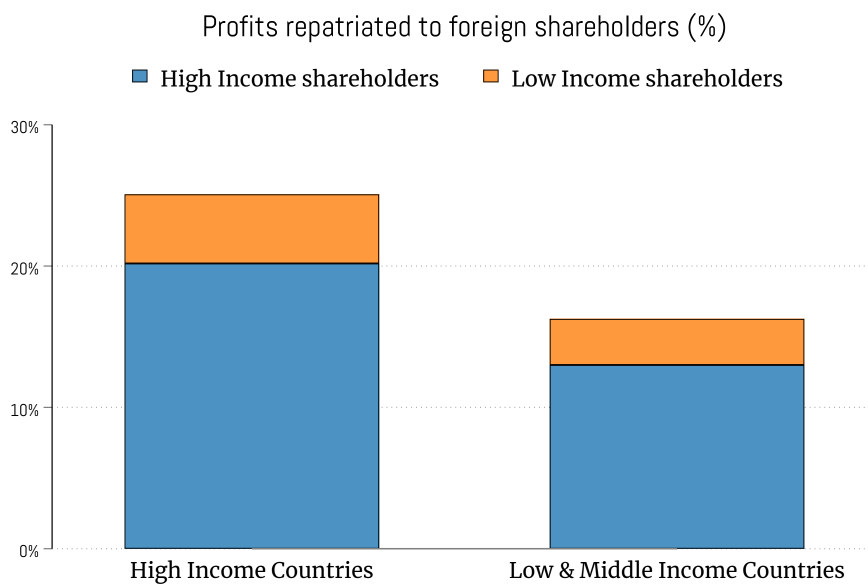



*Note*: The data is from COMPUSTAT. Accounting profit margins are measured as the weighted average of firm-level sales to cost ratios within SIC industries. The x-axis represents an industry's position (percentile 0 to 100) in the profit distribution based on it average accounting profit margin during the 1980-2005 period.

**Fact 3.** A minor fraction of profits are repatriated to foreign shareholders, but most repatriated profits payments accrue to high-income countries.

As noted earlier, the extent of international rent-shifting depends on the location in which profits are rebated. In theory, profits earned in one country could be repatriated to foreign shareholders, which may complicate the national-level relationship between profits and real income. Our last stylized fact, however, reveals that the majority of profits are distributed to domestic shareholders, with only a small portion being repatriated to foreign shareholders, primarily in high-income countries. Therefore, to the extent that we are concerned about rent-shifting from low to high-income countries, repatriated profit payments actually exacerbate the effect rather than mitigate it. We document this fact using firm-level ownership data from Orbis with results plotted in Figure 5. Evidently, over 85% of the profits earned by firms are distributed within the country of origin, and this percentage is even higher among high-income countries. The remaining profits are primarily repatriated to foreign shareholders located in high-income regions. These patterns suggest that repatriated profits contribute to transfer of rents from low and middle-income countries to high-income nations, amplifying the rent-shifting effects due to trade-led specialization.

Figure 5: The percent of profits repatriated to foreign shareholders




Note: The data is from Orbis. The share of profits repatriated to foreign shareholders are inferred from equity shares in multinational enterprises, using the algorithm described in Appendix A. The classification of countries into high-income and low/middle income is based on the United Nations Country Classification.

## 4 Non-Parametric Global Economy with Markup Distortions

We derive our main theoretical result using a non-parametric model of the global economy consisting of multiple countries, indexed by $n, i = 1, .., N$. Country $i$ hosts a fixed number of firms indexed by

$\omega \in \Omega_i$, each supplying a tradable and differentiated product variety. These firm-level varieties may span multiple sectors of economy. Labor is the only primary factor of production, and each country $i$ is endowed with an inelastic supply of labor, $L_i$, that is paid an equilibrium wage, $w_i$. Labor is internationally immobile but mobile across different production activities within a country.

**Demand.** The representative consumer in country $i$ maximizes a non-parametric utility function that aggregates over firm-level varieties sourced from various origin countries. Welfare in country $i$ is accordingly measured by the representative consumer's indirect utility,

$$W_i = V_i \left(E_i, \{\mathbf{p}_{ni}\}_n\right),$$

which depends on total expendable income, $E_i$, and the prices of all firm-level product varieties available to the consumer. Namely, $\{\mathbf{p}_{ni}\}_n$, where $\mathbf{p}_{ni} \equiv \{p_{ni}(\omega)\}$ contains the price of all goods sold by firms from origin country $n$ to country $i$.

**Supply.** Country $n$ is populated by a fixed set of firms that use labor as the sole primary factor of production and charge a possibly variable markup over marginal cost, reflecting profit maximization. The price of firm-level variety $\omega$ (sold from origin $n$ to destination $i$) can be specified as

$$p_{ni}(\omega) = \underbrace{\mu_{ni}(\omega)}_{\text{markup}} \times \tau_{ni} w_n / \varphi_n(\omega),$$

where $\mu_{ni}(\omega)$ is the variety-specific markup, $\varphi_n(\omega)$ denotes firm $\omega$'s productivity, $\tau_{ni}$ is the iceberg trade cost, and $w_n$ denotes the wage rate paid to workers in the country of origin $n$.

**General Equilibrium.** For a given set of parameters, equilibrium is vector of demand quantities, $\mathbf{q}$, prices, $\mathbf{p}$, wages, $\mathbf{w}$, and income, $\mathbf{Y}$, such that the representative consumer's utility is maximized in each country; firm-level profits are maximized; labor markets clear, so wage payments in country $i$ equal sales net of markups,

$$w_i L_i = \int_{\omega\in\Omega_i} \sum_{n=1}^{N} \left[\frac{1}{\mu_{in}(\omega)} p_{in}(\omega)\, q_{in}(\omega)\right] d\omega;$$

and total expenditure equals total income, $Y_i$, which is wage income plus lump-sum rebates of markup rents,

$$E_i = Y_i = w_i L_i + \underbrace{\int_{\omega\in\Omega_i} \left[\sum_{n=1}^{N} \left(1 - \frac{1}{\mu_{in}(\omega)}\right) p_{in}(\omega)\, q_{in}(\omega)\right] d\omega}_{\text{markup rents}}.$$

Note, that markups in this setting are implicitly determined as the optimal price divided by marginal cost, i.e., $\mu_{in}(\omega) = p_{in}(\omega) / (\tau_{in} w_i / \varphi_{in}(\omega))$. We should note that markup rents in our baseline setup are rebated to consumers in the firms' country of origin. Later, we relax this assumption and use data on firm ownership to infer the fraction of rents rebated internationally.

**Alternative Representation of Equilibrium Outcomes.** Our theory can be streamlined by expressing firm-level variables as a function of the underlying markup, assuming the demand function exhibits sufficient symmetry to ensures a one-to-one mapping from firm-level productivity to markup within origin-destination dyads. Let $\mathcal{M} \subset [1, \infty)$ denote the compact set of markups charged by firms worldwide. For any $\mu \in \mathcal{M}$, and let $p_{in}(\mu) = \mu \times \tau_{in} w_i / \varphi_i(\mu)$ denote the price of firms charging markup $\mu$ (within the origin-destination dyad $in$), with $\varphi_i(\mu)$ denoting their productivity. With a slight abuse of notation, let $\tilde{q}_{in}(\mu)$ denote the quantity of firm-level varieties with markup $\mu$, which is the demand per firm times the corresponding measure of firms. The share of country $i$'s expenditure on varieties with markup $\mu$ can be specified as

$$e_n(\mu) = \sum_i \left[p_{in}(\mu)\, \tilde{q}_{in}(\mu)\right] / E_n,$$

where $E_n = \int_{\mathcal{M}} \sum_n p_{in}(\mu)\, \tilde{q}_{in}(\mu)\, d\mu = Y_n$, given the representative consumer's budget constraint. Denote by $\lambda_{in}(\mu)$ country $n$'s share of expenditure on goods originating from country $i$, conditional on the markup level, $\mu$. In particular,

$$\lambda_{in}(\mu) = \frac{p_{in}(\mu)\, \tilde{q}_{in}(\mu)}{\sum_{i'} \left[p_{i'n}(\mu)\, \tilde{q}_{i'n}(\mu)\right]}$$

To track output activity, let $y_i(\mu)$ denote the share of country $i$'s gross revenues attributed to (global) sales of goods with markup $\mu$. Namely,

$$y_i(\mu) = \frac{\sum_n \lambda_{in}(\mu)\, e_n(\mu)\, Y_n}{\int_{\mathcal{M}} \sum_n \lambda_{in}(\mu)\, e_n(\mu)\, Y_n d\mu}.$$

Likewise, we can specify $\ell_i(\mu) \equiv L_i(\mu) / L_i$, which is the share of labor used for producing goods with markup $\mu$. In particular, noting that $L_i(\mu) = \frac{1}{\mu} Y_i(\mu)$, yields

$$\ell_i(\mu) = \frac{\sum_n \frac{1}{\mu} \lambda_{in}(\mu)\, e_n(\mu)\, E_n}{\int_{\mathcal{M}} \sum_n \frac{1}{\mu} \lambda_{in}(\mu)\, e_n(\mu)\, E_n d\mu}.$$

Appealing to the above relationship, we can express total income in country $i$ as a function of wage payments and average (output-side) markup as follows:

$$Y_i = \left(\int_{\mathcal{M}} \mu \ell_i(\mu)\, d\mu\right) w_i L_i = \left(\int_{\mathcal{M}} \frac{1}{\mu} y_i(\mu)\, d\mu\right)^{-1} w_i L_i. \tag{1}$$

**Notation:** To condense the notation, we hereafter use $\mathbb{E}_\omega[.]$ and $\widetilde{\mathbb{E}}_\omega[.]$ to denote the *arithmetic* and *harmonic* mean operators. In particular, for a generic function, $F(\mu)$, define the operators

$$\mathbb{E}_\omega[F(\mu)] \equiv \int_{\mathcal{M}} F(\mu)\,\omega(\mu)\,d\mu \qquad \text{(Arithmetic mean)}$$

$$\widetilde{\mathbb{E}}_\omega[F(\mu)] \equiv \left(\int_{\mathcal{M}} F(\mu)^{-1}\,\omega(\mu)\,d\mu\right)^{-1} \qquad \text{(Harmonic mean)}$$

where $\omega(\mu) \geq 0$ is a well-behaved weight function that satisfies $\int_{\mathcal{M}} \omega(\mu)\,d\mu = 1$. To showcase how the above definition simplifies notation, we can specify the relationship between total income and wage payments (Equation 1) more compactly as

$$Y_i = \widetilde{\mathbb{E}}_{y_i}[\mu]\,w_i L_i; \qquad Y_i = \mathbb{E}_{\ell_i}[\mu]\,w_i L_i,$$

where $\widetilde{\mathbb{E}}_{y_i}[\mu]$ is country $i$'s output-weighted harmonic mean markup and $\mathbb{E}_{\ell_i}[\mu]$ is the employment-weighted arithmetic mean markup.

## 5 The Deadweight Loss of Markups in a Global Setting

In this section, we derive sufficient statistics formulas for the deadweight loss of markups in open economies, contrasting it with the better known closed economy case.

**Characterizing the Pareto efficient frontier.** The global economy's Pareto efficient frontier can be obtained by solving a planning problem where the central planner selects the global vector of after-tax-cum-subsidy prices, $\widetilde{\mathbf{p}} = \{\widetilde{\mathbf{p}}_i\}_i$, and uses lump-sum transfers to redistribute surplus inter-nationally based on Pareto weights. Letting $\alpha_i$, denote national-level Pareto weights, the noted planning problem amounts to the following maximization problem

$$\max_{\widetilde{\mathbf{p}},\mathbf{T}} \sum \alpha_i \ln V_i\left(E_i, \widetilde{\mathbf{p}}_i\right)$$

subject to the availability of lump-sum income transfers ($T_i$) and feasibility constraints,

$$E_i = \widetilde{\mathbb{E}}_{y_i}[\mu]\,w_i L_i + \sum_n \left[\int_{\mathcal{M}} \left(\widetilde{p}_{in}(\mu) - p_{in}(\mu)\right)\tilde{q}_{in}(\mu)\,d\mu\right] + T_i, \qquad \sum_i T_i = 0.$$

As shown in Appendix B, the solution to the above problem restores marginal cost pricing globally,

$$\widetilde{p}^*_{ni}(\mu) = \frac{1}{\mu} p_{ni}(\mu),$$

and assign transfers $T_i^*$ based on the national-level exposure to policy and the underlying Pareto weights. The *unconstrained* Pareto frontier can then be traced by varying $\alpha_i$ and adjusting the optimal transfers.

**Distance to the efficient frontier.** We are interested in the welfare consequences of moving from the decentralized equilibrium to the efficient (marginal cost pricing) frontier, holding technology and the set of operating firms fixed. As before, we denote equilibrium outcomes pertaining to the efficient frontier with superscript $*$. Suppose preferences are homothetic and the initial markup schedule represents a sufficiently small departure from efficient pricing.[7] The welfare impacts from moving to a point on the efficient frontier, $\Delta \ln W_i = \ln \left( W_i^* / W_i \right)$, are approximated by

$$\Delta \ln W_i \approx \Delta \ln E_i - \sum_n \int_{\mathcal{M}} \Delta \ln p_{ni} (\mu) \, \lambda_{ni} (\mu) \, e_i (\mu) \, d\mu, \tag{2}$$

where $\Delta \ln E_i$ denotes the corresponding change in country $i$'s nominal expenditure after the restoration of efficient pricing and the assignment of appropriate transfers, and $\Delta \ln p_{ni} (\mu)$ denotes the price change due to markup correction. We begin by specifying the price changes in Equation 2. The price of firm-level varieties sold by origin $n$ to destination $i$ in the decentralized equilibrium are $p_{ni} (\mu) = \mu \times \tau_{ni} w_n / \varphi_n (\mu)$. After markup correction, the prices are revised to $p_{ni}^* (\mu) = \tau_{ni} w_n / \varphi_n (\mu)$. The corresponding price change, $\Delta \ln p_{ni} (\mu) = \ln p_{ni}^* (\mu) - \ln p_{ni} (\mu)$, can be, thus, specified as a function of the initial markup and general equilibrium wage adjustments. In particular,

$$\Delta \ln p_{ni} (\mu) = - \ln \mu + \Delta \ln w_n, \tag{3}$$

where $\Delta \ln w_{in} = \ln ( w_n^* / w_n )$. Aggregating over the price changes (as they appear in Equation 2) and noting that $\sum_n \lambda_{ni} (\mu) = 1$ for all $\mu \in \mathcal{M}$, yields

$$\sum_n \int_{\mathcal{M}} \Delta \ln p_{ni} (\mu) \, \lambda_{ni} (\mu) \, e_i (\mu) \, d\mu = - \underbrace{\int_{\mathcal{M}} \ln \left( \frac{1}{\mu} \right) \sum_n \left[ \lambda_{ni} (\mu) \right] e_i (\mu) \, d\mu}_{\mathbb{E}_{e_i} \left[ \ln \frac{1}{\mu} \right]} - \sum_n \lambda_{ni} \Delta \ln w_n, \tag{4}$$

where $\lambda_{ni} \equiv \int_{\mathcal{M}} \lambda_{ni} (\mu) \, d\mu$ denotes country $i$'s total expenditure share on varieties sourced from origin $n$. To specify the income effects in Equation 2, note that $E_i = Y_i = \widetilde{\mathbb{E}}_{y_i} [\mu] \, w_i L_i$ and $E_i^* = \mathscr{T}_i^* w_i^* L_i$, where $\mathscr{T}_i^* = 1 + T_i^* / Y_i^*$ is the transfer to income ratio. By construction, $\mathscr{T}_i = 1$ in the decentralized transfer-free equilibrium. Hence, we can specify the change in country $i$'s expendable income as

$$\Delta \ln Y_i = \ln \left( \mathscr{T}_i^* w_i^* L_i \right) - \ln \left( \widetilde{\mathbb{E}}_{y_i} [\mu] \, w_i L_i \right) = \Delta \ln w_i + \Delta \ln \mathscr{T}_i + \ln \mathbb{E}_{y_i} \left[ \frac{1}{\mu} \right]. \tag{5}$$

As one can see, the welfare effects specified by Equations 5 and 4 are contaminated with *factoral* terms of trade effects. More specifically, markup correction modifies international wages as represented by $\{ \Delta \ln w_n \}_n$, thereby disrupting relative factor prices among countries. However, these effects are typically negligible, as demonstrated in Figure A5 of the appendix. Another issue is that there are numerous

[7] The above approximation tends to equality if high-market-share variaties exhibit a sufficiently low markup ($\ln \mu \approx 0$) and high-markup variaties absorb a sufficiently low market share ($\lambda_i (\mu) \approx 0$).

points on the Pareto efficient frontier, each characterized by a different transfer schedule, $\{\Delta \ln \mathscr{T}_n\}_n$. To handle these technical issues, we define the deadweight loss of monopolistic markup distortions for country $i$ as the (foregone) welfare gains from restoring marginal cost pricing globally net of transfers and factoral terms of trade effects—namely, $\mathscr{D}_i \equiv (\Delta \ln W_i)_{\mathbf{w},\mathscr{T}}$. In other words, $\mathscr{D}_i$ signifies the welfare change associated with moving to a point on the efficient frontier where factoral terms of trade effects are exactly offset by appropriate lump-sum transfers. By substituting Equations 5 and 4 into 2 and eliminating the mirror terms related to factoral terms of trade effects and transfers, we obtain the expression $\mathscr{D}_i = \ln \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right]$. Reorganizing this expression allows us to decompose the deadweight loss of monopoly distortions into two components: the loss associated with markup dispersion (mean log deviation of inverse markups) and the exposure to international rent-shifting externalities.[8]

**Proposition 1.** *The deadweight loss from monopolistic markup distortions for open economy i is given by*

$$\mathscr{D}_i \approx \underbrace{\left(\ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right]\right)}_{\textit{markup dispersion}} + \underbrace{\ln\left(\widetilde{\mathbb{E}}_{e_i}\left[\mu\right] / \widetilde{\mathbb{E}}_{y_i}\left[\mu\right]\right)}_{\textit{international rent-shifting}}, \tag{6}$$

*where the first element represents the (expenditure-weighted) markup dispersion in country i and the second element represents international rent-shifting externalities.*

The *markup dispersion* component of the deadweight loss represents to the *mean log deviation* (MLD) of inverse markups. We henceforth condense the notation by re-writing the dispersions terms simply as

$$\ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right] \sim \text{MLD}_{e_i}\left(\frac{1}{\mu}\right).$$

While the *markup dispersion* term is well-understood, the *international rent-shifting* component of $\mathscr{D}_i$ has less precedents in the literature and merits further elaboration. In a closed economy, markup rents are rebated to the same consumers whose surplus is negatively impacted by markups, precluding rent-shifting effects. However, in an open economy, the loss to consumer surplus may occur in one location while excess profits (rents) are rebated elsewhere. Consequently, the deadweight loss from markup distortions may be elevated or mitigated for households in country $i$, depending on whether country $i$ is a net receiver or a net payer of markup rents to the rest of the world. Accordingly, rent-shifting effects can be specified in terms of a country's revealed comparative advantage across low and high-markup goods, as follows:

$$\ln\left(\widetilde{\mathbb{E}}_{e_i}\left[\mu\right] / \widetilde{\mathbb{E}}_{y_i}\left[\mu\right]\right) \approx \widetilde{\mathbb{E}}_{e_i}\left[\mu\right] \times \text{Cov}\left(\frac{y_i(\mu)}{e_i(\mu)}, \frac{1}{\mu}\right),$$

where Cov (.) is the covariance operator. The above formulation relates a country's exposure to rent-shifting externalities to its pattern of specialization, as measured by net exports. If $y_i(\mu)/e_i(\mu) > 1$,

[8] We later present formulas for $\mathscr{D}_i$ when markups distort entry decisions. However, our framework does not account for distortions to dynamic invest decisions, as emphasized by Voronina (2022) and Adhami et al. (2024).

country $i$ is a net exporter of goods with markup $\mu$. And, conversely, if $y_i(\mu)/e_i(\mu) < 1$, country $i$ is a net importer of these goods. Accordingly, this formulation posits that, in a global setting, the burden of monopoly distortions is relatively lower for countries that are net exporters of high-markup goods and relatively higher for others. The above approximation hints that international rent-shifting are being zero-sum in nature—a point we formally show later when dissecting the impacts of trade.

**Closed Economy Benchmark.** A special case of Proposition 1 is the closed economy case, wherein the industry-level expenditure and revenue shares coincide—i.e., $y_i(\mu) = e_i(\mu)$ for all $\mu \in \mathcal{M}$. Plugging this identity into the expression for $\mathscr{D}_i$ from Proposition 1, yields

$$\mathscr{D}_i^{closed} \approx \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right] \sim \text{MLD}_{e_i}\left(\frac{1}{\mu}\right). \tag{7}$$

That is, the deadweight loss of monopoly distortions in a closed economy is determined entirely my the extent of markup dispersion, as measured by the mean log deviation (MLD) of inverse markups. This result echoes our previous assertion that in a closed economy, markup rents are rebated to the same set of households who initially pay for the markups, precluding any excess gains or losses from international rent-shifting. With this background in mind, we now isolate the impacts of trade on the welfare cost (or incidence) of monopoly distortions.

**Dissecting the Impacts of Trade on Markup Distortions.** Our results, thus far, signal that trade openness modifies the deadweight loss from monopoly distortions across countries. We can formalize this implication by contrasting the counterfactual deadweight loss of monopoly distortions under autarky (Equation 7) to its factual loss (Equation 6)—calculating the trade-led change in deadweight loss or incidence of monopoly distortions as $\Delta\mathscr{D}_i \equiv \mathscr{D}_i - \mathscr{D}_i^{closed}$. Doing so delivers

$$\Delta\mathscr{D}_i = \underbrace{\Delta\text{MLD}_{e_i}\left(\frac{1}{\mu}\right)}_{\Delta \text{ markup dispersion}} + \underbrace{\ln\left(\widetilde{\mathbb{E}}_{e_i}[\mu]/\widetilde{\mathbb{E}}_{y_i}[\mu]\right)}_{\text{international rent-shifting}}.$$

The above equation divides the impacts of trade into two effects. First, trade can change the extent of markup dispersion as measured by $\Delta\text{MLD}_{e_i}\left(\frac{1}{\mu}\right)$. These changes can reflect either trade-led consumption reallocation or *pro-competitive* pressures. Second, trade leads to a decoupling between markup payments and markup dividends as measured by the international rent-shifting term, $\ln\left(\widetilde{\mathbb{E}}_{e_i}[\mu]/\widetilde{\mathbb{E}}_{y_i}[\mu]\right)$. While the former effect could impact all countries in the same direction, the latter effect is *zero-sum* by nature. More specifically, international rent-shifting effects are all but a pure redistribution of surplus from one set of countries to another, which satisfy $\ln \sum_i \left[\frac{w_i L_i}{w \cdot L} \times \frac{\widetilde{\mathbb{E}}_{y_i}[\mu]}{\widetilde{\mathbb{E}}_{e_i}[\mu]}\right] = 0$. That is, the (log) weighted average of rent-shifting effects across countries is zero, implying that trade increases the incidence of monopoly distortions among one set of countries while decreasing it for others. The beneficiaries of these trade-led adjustments are countries specializing in high-markup product categories and net receivers of markup rents from the rest of the world.

### 5.1 The Dissipation of Quasi-Rents under Free Entry

In our baseline model, we simplified the analysis by disregarding firm entry and the fact that monopoly rents encompass quasi-rents used to cover sunk entry costs. We now demonstrate that even with firm entry, international imbalances in monopoly rents lead to zero-sum welfare effects, similar to rent-shifting externality in our baseline model. Though, the nature of these zero-sum welfare effects are different under firm entry. In this case, the market failure arises not from excessive markups but rather from the excessive entry of firms into certain industries. We illustrate that in open economies, exposure to entry distortions mirrors the rent-shifting externalities highlighted earlier.

To demonstrate this point, suppose firms pay a sunk entry cost, $w_i f_{i,k}$, to develop a blueprint. The number of entrants that pay the noted cost is determined by the free-entry condition, which equates variable profits to the entry cost in each industry and country. For simplicity, suppose demand exhibits a CES parametrization. Following Lashkaripour and Lugovskyy (2023), the efficient allocation can be implemented via markup-correcting output taxes (i.e., $\tau_i^*(\mu) = 1/\mu$).[9] Technical details regarding the free entry equilibrium and efficient tax implementation are provided in Appendix D. The same appendix shows that a closed economy's distance to the efficient frontier *under free entry* is

$$\mathscr{D}_i^{closed} = \mathbb{E}_{e_i}[\mu \ln \mu] - \mathbb{E}_{e_i}[\mu] \ln \mathbb{E}_{e_i}[\mu] .$$

The above formula represents the deadweight loss from firms' entry decisions not internalizing the social benefits of adding new product varieties. Crucially, the marginal value of additional variety is associated with the markup in each industry, implying inefficiently low entry among high-markup firms and excess entry among low-markup firms. The extent of inefficiency is, thus, determined by the degree of markup dispersion, i.e., $\mathscr{D}_i^{closed} \approx \text{Var}_{e_i}[\mu]$. As before, openness to trade modifies the national-level incidence of entry distortions, as denoted by $\Delta\mathscr{D}_i$. As demonstrated in Appendix D, these trade-induced welfare effects are given by the following equation under free entry:

$$\Delta\mathscr{D}_i = \mathbb{E}_{e_i}\left[(\mu - 1)\ln \frac{y_i^*(\mu)}{y_i(\mu)}\right],$$

where $y_i^*(\mu)$ is the counterfactual output share associated with markup $\mu$ under the efficient allocation. The expression above echoes the result presented under Proposition 2. Specifically, a global markup-correcting policy will increase the relative wage of countries that are net exporters of high-markup goods. The higher wage would suppress demand relatively more for these countries' output of low-markup goods, as these goods face a more price-elastic demand function. As a result, $y_i^*(\mu)/y_i(\mu)$ will be positively correlated with $\mu$ among net exporters of high-markup goods, implying that trade lowers

[9] The tax revenue is rebated to consumers in a lump-sum fashion. Though, the tax policy could be designed to be revenue-neutral, since all tax policies of the form $\tau_i^*(\mu) = \overline{\mu}/\mu$ are efficient, where $\overline{\mu}$ is a constant tax shifter which can be set to yield zero revenues.

the incidence of entry distortions for these countries, *i.e.*, $\Delta\mathscr{D}_i > 0$. The opposite effect will occur in countries that are net importers of high-markup goods, as the policy will lead to a reduction in their relative wage and a corresponding shift in their production towards low-markup goods.[10]

Figure A6 in the appendix illustrates this point by simulating a generic model featuring two countries. The countries are symmetric except for their revealed comparative advantage in low-markup versus high-markup goods. The simulation demonstrates that under both free and restricted entry, trade openness amplifies the deadweight loss of monopoly distortions for the country that is a net importer of high-markup goods. Conversely, it reduces these costs for the other country. This affirms that monopoly distortions have internationally zero-sum effects, even in scenarios where firm entry dissipates quasi-rents. The underlying logic is that the degree of market power is correlated with entry distortions, and exposure to these distortions mirrors exposure to rent-shifting under restricted entry.

## 6 Exact Measurement using Semi-Parametric Model

According to Proposition 1, the data requirements to measure the deadweight loss of markups *non-parametrically* are virtually prohibitive in an global economy setting. Measurement requires granular international data on expenditure and output shares by markup level, namely, $e_i(\mu)$ and $y_i(\mu)$. The former variable can, in principle, be recovered using internationally representative firm-level balance sheet data, but recovering $e_i(\mu)$ from existing data is practically infeasible. Moreover, isolating the impact of trade on the deadweight of markups ($\Delta\mathscr{D}_i$) is even more challenging. It requires that one infers the counterfactual mean log deviation of markups under autarky from observables.

We can overcome these data challenges by turning to a parametric model of the global economy. But, Ideally, we would like to proceed with the *minimal* amount of parametric assumptions that allow measurement using existing data. On the preference side, we assume that there are $k = 1, ..., K$ narrowly-defined industries and the utility aggregator across industries is Cobb-Douglas, $W_i = \prod_k Q_{i,k}^{e_{i,k}}$. The country-specific weights, $e_{i,k}$, represent the share of earned income spent on industry $k$ goods.

Expenditure shares within industries are regulated by a *homothetic with aggregator demands system*, as defined by Errico and Lashkari (2022). To present the demand system concisely, we temporarily drop indices that specify origin, and destination and only maintain the index $\omega$ denoting the firm variety. The demand system characterizes consumer choices over a set $\Omega_k$ of firm-level product varieties within industry $k$, with price $\mathbf{p}_k = \{p(\omega)\}_{\omega\in\Omega_k}$. It is characterized by a collection of expenditure-share functions $\Lambda_k(\acute{p}(\omega); \boldsymbol{\varsigma})$, where $\acute{p}(\omega)$ is the normalized price of variety $\omega$ and $\boldsymbol{\varsigma} \in \mathbb{R}^D$ is a vector of parameters. The share functions satisfy an adding-up constraint, $\int_{\omega\in\Omega} \Lambda_k(\acute{p}(\omega); \boldsymbol{\varsigma})\, d\omega = 1$. The price normalization, $\acute{p}(\omega) \equiv p(\omega) / \mathscr{P}_k(\mathbf{p}_k; \boldsymbol{\varsigma})$, uses a linear homogeneous aggregator satisfying $\mathscr{P}_k(\alpha\mathbf{p}_k; \boldsymbol{\varsigma}) = \alpha\mathscr{P}_k(\mathbf{p}_k; \boldsymbol{\varsigma})$ for all $\alpha > 0$. This normalization ensures that the composition of demand

[10] In Appendix C.2, we explore the dissipation of quasi-rents by fixed operational costs that differ from sunk entry costs. We derive formulas for the deadweight loss (DWL) of markups and its response to trade in this context, highlighting that accounting for quasi-rents does not necessarily diminish the DWL of markup distortions. The intuition behind this is that the DWL of markup wedges is determined by the dispersion in excess markups. This dispersion could be either lower or higher than the gross measure of dispersion that does not take quasi-rents into account.

depends only on relative prices, not the price level. The expenditure share functions are assumed to satisfy the following functional form

$$\Lambda_k\left(\acute{p}\left(\omega\right);\varsigma\right)=\frac{\acute{p}\left(\omega\right)D_k\left(\acute{p}\left(\omega\right);\varsigma\right)}{\int_{\omega'}\acute{p}\left(\omega'\right)D_k\left(\acute{p}\left(\omega'\right);\varsigma\right)d\omega'},\tag{8}$$

where $D_k\left(.,\varsigma\right)$ are positive-valued single-argument functions that are decreasing over interval $\acute{p}\left(\omega\right)\in\left(0,\underline{\acute{p}}\right)$. There is also a constant *relative* choke price $\underline{\acute{p}}\in\mathbb{R}_+\cup\{\infty\}$, beyond which demand for a variety drops to zero. In particular, $\lim_{\acute{p}(\omega)\to\underline{\acute{p}}}D_k\left(\acute{p}\left(\omega\right);\varsigma\right)=0$ and $D_k\left(\acute{p}\left(\omega\right);\varsigma\right)=0$ for $\acute{p}\left(\omega\right)\geq\underline{\acute{p}}$. This demand system is referred to as homothetic with aggregator since it specifies expenditure share based on two aggregate indices $P_k\equiv\mathscr{P}_k\left(\mathbf{p}_k;\varsigma\right)$ and $Y_k\equiv\int_\omega\frac{p(\omega)}{P_k}D_k\left(\frac{p(\omega)}{P_k};\varsigma\right)d\omega$. The share of expenditure on variety $\omega$ is, specifically, given by $\lambda_k\left(\omega\right)\equiv\frac{1}{Y_k}\frac{p(\omega)}{P_k}D_k\left(\frac{p(\omega)}{P_k};\varsigma\right)$.

On the production side, we assume that the firm-level productivity distribution is Pareto with a good-specific shape parameter and country and good-specific scale parameter. In particular,

$$G_{i,k}\left(\varphi\right)=1-\left(\overline{\varphi}_{i,k}/\varphi\right)^{\theta_k}.\tag{9}$$

The semi-parametric model allows us to determine $e_i\left(\mu\right)$ and $y_i\left(\mu\right)$ by combining aggregate production and expenditure data with a global sample of firm-level balance sheets. We can then leverage these share variables to calculate the deadweight loss of markups on an international scale. The key insight underpinning this measurement is that the *within-industry* expenditure share on varieties with markup $\mu$ is common across countries and independent of trade costs.

To illustrate this result, we specify the variables associated with a generic firm variety with marginal cost $c$, as a function of $\nu\equiv P_{i,k}/c$, which represents the firm's competitiveness.[11] The marginal cost is described by $c_{ni,k}\left(\omega\right)=\tau_{ni,k}w_i/\varphi_{n,k}\left(\omega\right)$, where $\varphi_{n,k}\left(\omega\right)$ is the firm-level productivity that is distributed according to 9. With the relative choke price set to $\underline{\acute{p}}=1$, varieties sourced from any origin country fall within a competitiveness range of $\nu\in(1,\infty)$, irrespective of trade costs.

Building on the approach in Arkolakis et al. (2019), we show that a firm's markup is uniquely determined by its competitiveness $\nu$ through a function $m_k\left(\nu\right)$ that is strictly increasing, concave, and common across countries (see Appendix C). Consequently, the price charged by each firm and its share of total sales and expenditure are also uniquely determined by $\nu$. We denote by $\tilde{\lambda}_{i,k}\left(\nu\right)$, country $i$'s share of expenditure on varieties with competitiveness $\nu$ *within* industry $k$; and by $\tilde{\rho}_{i,k}\left(\nu\right)$ the share of these varieties from total output sales in industry $k$. Appendix C shows that these *within-industry* share variables are country-blind (i.e., $\tilde{\lambda}_{i,k}\left(\nu\right)\sim\tilde{\lambda}_k\left(\nu\right)$ and $\tilde{\rho}_{i,k}\left(\nu\right)\sim\tilde{\rho}_k\left(\nu\right)$) and given by

$$\tilde{\lambda}_k\left(\nu\right)=\tilde{\rho}_k\left(\nu\right)=\frac{\frac{m_k(\nu)}{\nu}D_k\left(\frac{m_k(\nu)}{\nu};\varsigma\right)\nu^{-\theta_k-1}}{\int_1^\infty\frac{m_k(\nu)}{\nu}D_k\left(\frac{m_k(\nu)}{\nu};\varsigma\right)\nu^{-\theta_k-1}d\nu}.$$

[11] Note that $P_{i,k}$ is the finite choke price for industry $k$ goods in market $i$ under the normalization $\underline{\acute{p}}=1$. More specifically, if the *relative* choke price is $\underline{\acute{p}}\equiv\frac{p}{P}=1$, then demand would be zero for any variety for which $p<P$.

The above equation also asserts that $\tilde{\lambda}_k(\nu)$ and $\tilde{\rho}_k(\nu)$ are independent of the underlying trade costs. Accordingly, the expenditure and output shares associated with variaties containing markup $\mu$ are also country-blind and invariant to trade costs, given by $\lambda_{i,k}(\mu) \sim \lambda_k(\mu) = \tilde{\lambda}_k\left(m_k^{-1}(\mu)\right)$ and $\rho_{i,k}(\mu) \sim \rho_k(\mu) = \tilde{\rho}_k\left(m_k^{-1}(\mu)\right)$. The *total* expenditure and output shares associated with markup $\mu$ (across all industries) can be then obtained from the within-industry shares as

$$e_i(\mu) = \sum_k \lambda_k(\mu)\, e_{i,k} = \sum_k \rho_k(\mu)\, e_{i,k} \qquad\qquad y_i(\mu) = \sum_k \rho_k(\mu)\, y_{i,k},$$

both of which are invariant to trade costs. The above equations allow us to calculate $e_i(\mu)$ and $y_i(\mu)$ using two pieces of information: the global distribution of sales across markup levels within industries ($\rho_k(\mu)$) and *aggregate* share variables across industries for each country ($y_{i,k}$ and $e_{i,k}$). Importantly, $\rho_k(\mu)$ can be derived from global data on firm balance sheets via standard markup estimation techniques, while $y_{i,k}$ and $e_{i,k}$ can be calculated using widely-accessible aggregate datasets. By plugging the derived $e_i(\mu)$ and $y_i(\mu)$ into Proposition 1, we can approximate the deadweight loss of markup distortions across a wide range of countries. Moreover, the semi-parametric model allows for an exact calculation of the effect of trade on the welfare cost or incidence of markups, as summarized below.

**Proposition 2.** *If consumer preferences and the firm productivity distribution satisfy 8 and 9, the impact of trade on deadweight loss of markups for country i is exactly specified by*

$$\Delta\mathscr{D}_i = \ln\left(\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right),$$

*in terms of the following sufficient statistics: (1) sales-weighted average markup by industry,* $\widetilde{\mathbb{E}}_{\rho_k}[\mu]$, (2) *industry-level output shares,* $y_{i,k}$, *and* (3) *industry-level expenditure share,* $e_{i,k}$.

The above proposition asserts that trade modifies the burden of markup distortions purely through rent-shifting effects. Trade, in other words, has no pro-competitive effects resonating with the argument in Arkolakis et al. (2019). Moreover, one can compute these welfare consequences with widely-available data on macro-level output and expenditure shares as well as a globally representative sample of firm-level markups. The gains from trade (relative to autarky) in our semi-parametric model are described by the following formula, $GT_i = 1 - \exp(\Delta\mathscr{D}_i) \times \prod_k \lambda_{ii,k}^{e_{i,k}/\theta_k}$, which departs from the canonical ACR formula in an international rent-shifting adjustment, $\Delta\mathscr{D}_i$, as specified by Proposition 2.[12] In other words, monopolistic markup distortions magnify the gains from trade for countries that (on net) collect monopoly rents from the rest of the world while lowering them for others.

[12]The gains from trade formula specified above generalizes the two-by-two formula derived concurrently by Firooz and Heins (2023) to a semi-parmateric multi-country, multi-industry setting with variable markups. It also bears resemblance to, but differs from, the gains from trade formula derived by Lashkaripour (2020) and Kucheryavyy et al. (2023) in the presence of scale distortions.

### 6.1 Measurement in Richer Environments

In this section we extend our analysis to richer environments. We, more specifically, re-derive the expressions for $\mathscr{D}_i$ and $\Delta\mathscr{D}_i$ under multi-national ownership, input-output linkages à la Caliendo and Parro (2015), and fixed overhead costs. In the interest of brevity we present a verbal description of each extension here, with detailed derivations provided in the appendix.

$(a)$ **Multinational Ownership and Cross-Border Profit Payments.** As documented earlier, only a minor fraction of profits are repatriated to foreign shareholder—hence, the abstraction from cross-border profit payments in our baseline model. However, we can easily extend our baseline formulas to account for such payments. Appendix E derives updated formulas for $\Delta\mathscr{D}_i$, under the condition where a constant share $\pi_{ni}$ of country $n$'s profits are repatriated to international shareholders in country $i$. The new formula for $\Delta\mathscr{D}_i$ features an additional term that accounts for the cross-border profit payments to foreign shareholders. More formally,

$$\Delta\mathscr{D}_i = \overbrace{\ln\left(\sum_k \left[y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}\right] / \sum_k \left[e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}\right]\right)}^{\text{international rent-shifting}}$$
$$\underbrace{+\ln\left(1+\sum_{n\neq i}\left[\pi_{ni}\frac{Y_n}{Y_i}\sum_k y_{n,k}\left(1-\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}\right)-\pi_{in}\sum_k y_{i,k}\left(1-\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}\right)\right]\right)}_{\text{cross-border profits payments}}, \tag{10}$$

The last term in the above equation represents the net inflow of repatriated profits for country $i$, calculated as the difference between the inflow and outflow of such profits. Importantly, the semi-parametric model allows us to evaluate this term using aggregate profit ownership shares, denoted as $\{\pi_{ni}\}_{n,i}$, which can be inferred from firm-level ownership data. The calculation also requires data on industry-level output and expenditure shares and the sales-weighted average markup for each industry, as in our baseline model.

$(b)$ **Global Input-Output Networks.** Appendix G examines a global economy in which production relies on labor and internationally traded intermediate inputs. In this extension, the magnitude of international rent-shifting depends on the degree to which the markup paid on imported inputs is re-exported and passed on to foreign consumers after production. As a result, the formulas that describe the impacts of trade on the national-level incidence of monopoly distortions depend on the elements of the global input-output matrix. We present these sufficient statistics formulas in the Appendix G and observe that the core logic from our baseline model continues to hold. Specifically, trade intensifies the incidence of monopoly distortions for countries that are net exporters of high-markup goods while reducing it for others, where net exports now take into account global input-output linkages.

$(c)$ **Accounting for Fixed Overhead Costs.** Earlier, we showed that considering *sunk* entry cost payments, does not eliminate the zero-sum welfare effects associated with market power. In Appendix H, we explore how accounting for fixed overhead costs affects the zero-sum international rent-shifting effects. Specifically, we analyze a global economy where serving individual markets requires paying a fixed cost that consumes a portion of the monopoly rents. We provide updated formulas for calculating the deadweight loss of monopoly distortions, isolating how trade alters these costs. Our updated formulas demonstrate that a country's exposure to international rent-shifting in the presence of fixed overhead costs is influenced by two factors: the shape of the firm productivity distribution and how this industry-specific shape parameter correlates with a country's net exports. These factors determine the net profits paid to the rest of the world via fixed cost payments.

# 7 Duality Between Tariffs and Rent-Shifting Externalities

In this section, we argue that zero-sum rent-shifting effects are similar to implicit tariffs that tilt the terms of trade in favor of the beneficiary countries. This equivalence is useful because it highlights important nuances underlying the World Trade Organization's (WTO) principle of reciprocity and identifies potential policy solutions to counteract the rent-shifting externalities.

While our baseline model abstracted away from policy wedges to improve exposition, we now introduce tariffs and additional notation to articulate our duality result. Let $\mathbf{t} = \left\{ t_{ji,k} \right\}_{j,i,k}$ denote the vector of applied tariffs, where element $t_{ji,k}$ represents the tariffs collected by destination country $i$ on goods imported from origin country $j$ in industry $k$. Recall that industries in our semi-parmateric model differ in their underlying trade elasticities denoted by $\theta_k$ and the sale-weighted average markup, $\widetilde{\mathbb{E}}_{\rho_k}[\mu]$. Tariffs generate revenue for the government by creating an additional wedge between the price and marginal cost of internationally traded goods. The consumer price of goods after tariffs is given by

$$\widetilde{p}_{ni,k}(\mu) = (1 + t_{ni,k})\, p_{ni,k}(\mu),$$

where $p_{ni,k}(\mu)$ is producer price inclusive of the markup. The total expendable income in each country $i$, thus, equals the sum of wage income, markup rents, and tariff revenues:

$$E_i = \widetilde{\mathbb{E}}_{y_i}[\mu]\, w_i L_i + \sum_k \sum_n \left[ \frac{t_{ni,k}}{1 + t_{ni,k}} \lambda_{ni,k} e_{n,k} E_n \right].$$

Notice, tariffs and markups are similar in that they both create distortionary wedges that raise income from foreign consumers. Tariffs are, however, more streamlined and targeted instruments for collecting such revenues as they explicitly discriminate between domestic and foreign suppliers.

Specify welfare as a function of monopoly markups and tariffs as $W_i \equiv \mathbb{W}_i(\mathbf{t}, \boldsymbol{\mu})$. Next, consider a hypothetical scenario where all markups are eliminated globally, reducing them from their current level $\boldsymbol{\mu}$ to $\mathbf{1}$. The welfare gains from this change determine the deadweight loss of monopolistic markups as

$\mathscr{D}_i(\mathbf{t}) = \mathbb{W}_i(\mathbf{t}, \mathbf{1}) - \mathbb{W}_i(\mathbf{t}, \boldsymbol{\mu})$ for a given set of applied tariffs. As discussed earlier, $\mathscr{D}_i$ comprises two effects: first, the welfare effect of removing markup dispersion, which is beneficial for all countries; second, international rent-shifting effects, which are zero-sum transfers between countries. From the lens of our semi-parmateric model, the rent-shifting effects the sole channel through which trade modifies the deadweight loss of monopolistic distortions, which is denoted by $\Delta\mathscr{D}_i$.[13]

We demonstrate that the zero-sum rent-shifting effects of monopolistic markups are equivalent to the terms of trade effects that arise from imbalanced import tariffs. Specifically, starting from a markup-free equilibrium, if countries that benefit from rent-shifting externalities in the markup-distorted equilibrium increase their tariffs while other countries reduce theirs, the resulting welfare effects are identical to $\Delta\mathscr{D}$. Consequently, international rent-shifting effects can be interpreted as implicit tariffs that favor net exporters of high-markup goods. The following proposition, proven in Appendix J , presents this finding.

**Proposition 3.** *Suppose applied tariffs,* $\mathbf{t}$ *and trade elasticities,* $\boldsymbol{\theta}$*, are sufficiently small. The international rent-shifting effects associated with monopolistic markups are observationally equivalent to implicit tariffs denoted by* $\tilde{\mathbf{t}}$*. In particular, there exists a* $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$ *such that*

$$\Delta\mathscr{D}_i = \mathbb{W}_i(\mathbf{t}, \mathbf{1}) - \mathbb{W}_i(\mathbf{t} + \tilde{\mathbf{t}}, \mathbf{1}) \qquad \forall i = 1, .., N$$

*where* $\Delta\mathscr{D}_i = \ln\left(\frac{\sum_k y_{i,k} \widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k} \widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right)$ *denotes country i's exposure to rent-shifting effects and* $\tilde{t}_i$ *is country i's implicit tariff that reproduces* $\Delta\mathscr{D}_i$ *and is increasing in the net rents extracted from foreign consumers.*

The above proposition establishes a *weak* duality between tariffs and monopolistic markups, stating that tariffs can replicate the rent-shifting externalities caused by markups. However, Appendix J goes further by demonstrating a *strong* duality between tariffs and monopolistic markups, provided that trade elasticities are sufficiently uniform across industries. When this additional condition is met, tariffs can replicate not only the rent-shifting effects but also the welfare losses that arise from markup dispersion. Later in this paper, we fit our model to data and find that the conditions for both weak and strong duality hold in practice.

The conversion of markup distortions into tariff-equivalent distortions has two notable implications. First, it reveals that the monopolistic pricing behavior of firms can be viewed as a *decentralized form of terms of trade manipulation*, resembling tariffs imposed by a central government. This insight suggests that governments seeking to manipulate the terms of trade, but constrained by international commitments, may choose to refrain from regulating anti-competitive practices in order to maintain the implicit terms of trade benefits. Second, by converting rent-shifting externalities into equivalent tariff measures, we can identify policy solutions that are enforceable under existing trade agreements, as these

[13] State formally, $\Delta\mathscr{D}_i = \mathscr{D}_i(\mathbf{t}) - \mathscr{D}_i(\mathbf{t}' \to \infty)$, where $\mathscr{D}_i(\mathbf{t}' \to \infty)$ represents the deadweight loss in country $i$ under prohibitive tariffs $\mathbf{t}'$ that effectively render country $i$ a closed economy. It is important to note that the fact that $\Delta\mathscr{D}_i$ captures only rent-shifting effects holds true specifically in our semi-parametric model and may not be the case in general. In a more general setting, trade can also influence the degree of markup dispersion, as measured byMLD$_{e_i}\left(\frac{1}{\mu}\right)$.

agreements are designed to discipline explicit border policy measures. For instance, under the World Trade Organization (WTO), tariffs must adhere to the principle of reciprocity (Bagwell and Staiger (1999)). Proposition 3 implies that unilateral tariff concessions could effectively neutralize rent-shifting externalities by simply invoking the reciprocity principle within the WTO framework.

# 8 Quantitative Implementation

To calculate the deadweight loss of markup distortions and associated rent-shifting effects using our formulas, we need the following sufficient statistics: sales-weighted average markups by industry ($\widetilde{\mathbb{E}}_{\rho_k}[\mu]$), industry-level expenditure shares per country ($e_{i,k}$), and industry-level output shares per country ($y_{i,k}$). In more complex environments, we also need multi-national profits ownership shares ($\pi_{in}$) and national-level input-output shares. Among these statistics, markups must be estimated, while the rest are directly observable shares. We source the output, expenditure, and input-output share data from the OECD Inter-Country Input-Output (ICIO) Tables, which cover 64 major countries and 36 sectors from 2005 to 2015. We construct original data on profit ownership shares using Orbis, which we will detail in the following section.

To investigate whether rent-shifting externalities have disproportionate effects on high-income and low/middle income countries, we classify the countries in our sample based on the United Nations Country Classification. Our sample consists of 64 countries, each of which is categorized as either *low/middle income* or *high-income*. Table A4 presents the complete list of countries in our sample along with their respective income status. It is important to note that our sample also includes an aggregate of the rest of the world, which mostly represents low-income countries and is classified accordingly.

## 8.1 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 Berry et al. (1995)). 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 at scale across a wide range of countries and industries. To the best of our knowledge, this is one of the first attempts to compare the macro-level implication of markups estimated using the demand- and cost-based approaches.

### 8.1.1 Cost-Based Markup Estimation.

Our cost-based approach to markup estimation closely follows the method proposed by De Loecker and Warzynski (2012). We rely on their observation that the firm markup can be calculated based on cost minimization as follows:

$$\mu_{kt}(\omega) = \alpha_{kt}(\omega) \frac{p_{kt}(\omega) q_{kt}(\omega)}{C_{kt}(\omega)},$$

where $\mathcal{C}_{kt}(\omega)$ denotes the variable inputs cost and $\alpha_{kt}(\omega)$ is the firm-level output elasticity with respect to variable inputs. Since estimating the output elasticity at the firm level is practically infeasible, the standard approach to markup estimation recovers the output elasticity under the simplifying restriction that all firms within product category $k$ use the same production function. Under this restriction, we can estimate the industry-wide output elasticity, $\alpha_{kt}(\omega)$, using the control function approach in Olley and Pakes (1996). In the first stage, we purge output of measurement error and unanticipated shocks by regressing it on a second-order polynomial of inputs and investment. In the second stage, we estimate the output elasticities by fitting an AR(1) process for productivity and leveraging moment conditions that impose orthogonality between implied productivity and lagged variable inputs and current capital inputs.[14] Since balance sheet data record expenditure and sales rather than physical quantities, the structural error term in the production function is contaminated with unobserved prices shifter such as markups. Following De Loecker et al. (2020), we control for unobserved markups using firms' sales shares within industries.

We conduct the production function estimation per industry using firm-level financial accounts data from Compustat North America, which provides rich coverage regarding capital inputs and firm-level investments for the US and Canadian firms. The data is reported based on the SIC industry classification. So, we concord SIC industries into the 36 ICIO industries for which we have macro-level trade and production data. For each industry and year during the 2005-2015 period, we separately estimate the output elasticity using the control function method described above. Since panel data are required for the control function estimation, we employ 5-year rolling windows, assigning the elasticity estimates derived from data in years $t-2$ to $t+2$ the central year $t$.

We then compute firm-level markups using internationally-representative data from the Worldscope Global Database. The data reports the cost of variable inputs $\mathcal{C}_{kt}(\omega)$ and sales $p_{kt}(\omega)\, q_{kt}(\omega)$ across 71,546 publicly traded firms from 134 countries during the 2005 -2015 period. Some firms in this database operate in more than one industry, but we do not observe the breakdown of firm-level sales and costs by industry. To handle this, we assume that sales and costs are equally spread across different products.[15] Following De Loecker and Eeckhout (2018), we assume that the output elasticity is the same across countries. Letting $\hat{\alpha}_{kt}$ denote the output elasticity estimated previously, we calculate the markup charged by firm $\omega$ on industry $k$ goods as $\mu_{kt}(\omega) = \hat{\alpha}_{kt} p_{kt}(\omega)\, q_{kt}(\omega) / \mathcal{C}_{kt}(\omega)$. We then compute the harmonic sales-weighted average markup in industry $k$ as $\widetilde{\mathbb{E}}_{\rho_{kt}}[\mu] = \left[\sum_{\omega \in \Omega_k} \rho_{kt}(\omega) / \mu_{kt}(\omega)\right]^{-1}$, where $\rho_{kt}(\omega)$ is firm $\omega$'s sales share within the sample of firms operating in industry $k$. Recall that

[14]The estimating equation can be formally expressed as follows

$$\ln q_{kt}(\omega) = \alpha_{kt} \ln \mathcal{C}_{kt}(\omega) + \kappa_{kt} D_{kt}(\omega) + \varepsilon_{kt}(\omega)$$

where $\mathcal{C}_{kt}(\omega)$ denotes the variable input cost and $\kappa_{kt} D_{kt}(\omega)$ represents nonlinear controls like quasi-fixed inputs and investment. We could alternatively estimate $\alpha_{kt}$ using a simple non-linear regression relying on Ackerberg et al.'s (2015) identifying assumption that the variable input bundle is non-dynamic and chosen prior to the investment decision in $t$, allowing for productivity shocks to hit the firm between these two sub-periods.

[15]Specially, if firm $\omega$ is recorded as operating in $n_t(\omega)$ different ICIO industries in year $t$, one of which is $k$, we set $p_{kt}(\omega)\, q_{kt}(\omega) = \frac{1}{n_t(\omega)} p_t(\omega)\, q_t(\omega)$ and $\mathcal{C}_{kt}(\omega) = \frac{1}{n_t(\omega)} \mathcal{C}_t(\omega)$

$\widetilde{\mathbb{E}}_{\rho_{kt}}[\mu]$ is the sufficient markup statistics for calculating the trade-induced change in the dead-weight loss of markups. Figure A7 in the appendix reports $\widetilde{\mathbb{E}}_{\rho_{kt}}[\mu]$ derived from our cost-based markup estimates across various ICIO industries.

The cost-based approach, despite its advantages, has several known shortcomings. One issue is the aggregation bias due to the presumption that all firms have identical production functions. Foster et al. (2022) use granular data to estimate output elasticities that vary across establishments and time, highlighting the drawbacks of the common elasticity assumption. Another drawback is that the production function estimation relies on revenue and input expenditure data rather than physical quantities, effectively estimating the revenue elasticity. Bond et al. (2021) argue that the revenue elasticity reveals no meaningful information about markups and the price-to-marginal product ratios implied by this elasticity are more indicative of input market power. Meanwhile, De Ridder et al. (2022) utilize quantity data for French firms and find that while the level of markup estimates from revenue data is biased, it correlates with true markups. Given these potential limitations, we also estimate markups using the demand-based approach, which is described in the following section.

#### 8.1.2 Demand-Based Markup Estimation.

Markups can be alternatively derived from demand parameters, but demand estimation at scale presents several challenges. First, we must impose parametric assumptions to make progress. To navigate this issue without loss of generality, we estimate a mixed multinomial logit model (MMNL) which can approximate our semi-parametric demand system as closely as possible.[16] Second, the conventional approach to estimating the MMNL model, introduced by Berry et al. (1995, BLP hereafter), is computationally demanding, making it impractical to perform over thousands of product categories. To tackle this issue, we employ a log-linear approximation of the MMNL model proposed by Salanié and Wolak (2019), which is considerably simpler to estimate. The final difficulty lies in the data requirements for large-scale demand estimation. The standard BLP approach leverages data on observable product characteristics to achieve identification, but globally representative data on observed product characteristics is unavailable. We overcome this obstacle by leveraging high-frequency customs and exchange rate data to guide identification, eliminating the need for explicit data on product characteristics.

Before diving into our estimation strategy, let us provide a high-level overview of the MMNL model, which forms the foundation of our estimation. Consider a market populated by an infinite number of households, each of which chooses one product variety from the set $\Omega_{kt}$ of products available in industry $k$ in year $t$. There is also an outside good, the indirect utility of which is normalized to 1. Assuming that the idiosyncratic taste for product varieties is distributed *iid* according to a type-I Extreme Value

[16] This claim follows from Thisse and Ushchev (2016), who show that the homothetic with an aggregator demand system can be alternatively derived from a random utility model; and from McFadden and Train (2000) who establish that any random utility model can be approximated as closely as needed by the MMNL model.

distribution with scale parameter 1, the market share of variety $\omega \in \Omega_{kt}$ can be specified as

$$\lambda_{kt}(\omega) = \mathbb{E}_{\epsilon}\left[\frac{\exp\left(\left(\overline{\boldsymbol{\beta}}_{kt} + \boldsymbol{\epsilon}\right) \cdot \mathbf{X}_{kt}(\omega) + \xi_{kt}(\omega)\right)}{1 + \sum_{\omega' \in \Omega_{i,k}} \exp\left(\left(\overline{\boldsymbol{\beta}}_{kt} + \boldsymbol{\epsilon}\right) \cdot \mathbf{X}_{kt}(\omega') + \xi_{kt}(\omega')\right)}\right],$$

In this equation, $\mathbf{X}$ represents a vector of *observed* product characteristics, such as prices, and $\overline{\boldsymbol{\beta}}$ denotes the mean coefficients on these characteristics. $\boldsymbol{\epsilon}$ is a random coefficient that follows an *iid* distribution $N(0, \boldsymbol{\Sigma}_{kt})$, where $\boldsymbol{\Sigma}_{kt}$ is a diagonal variance matrix.[17] The demand shifter, $\xi$, captures *unobserved* product characteristics, such as perceived product quality at the market level. The BLP approach to estimating demand recovers $\boldsymbol{\xi}$ by inverting the market share equation and using the recovered values to enforce the moment condition $\mathbb{E}[\Delta\boldsymbol{\xi} \mid \mathbf{z}] = 0$, where $\mathbf{z}$ represents a set of price instruments. The inversion approach, however, is computationally challenging, particularly for large-scale applications. To overcome these computational hurdles, Salanié and Wolak (2019) propose an alternative approach that approximates $\boldsymbol{\xi}$ using the following equation:

$$\xi_{kt}(\omega) = \ln\left(\lambda_{kt}(\omega) / \lambda_{kt,0}\right) - \overline{\boldsymbol{\beta}}_{kt} \cdot \mathbf{X}_{kt}(\omega) - \tilde{\boldsymbol{\Sigma}}_{kt} \cdot \mathbf{K}_{kt}(\omega) + O\left(\| \boldsymbol{\Sigma}_{kt} \|^2\right),$$

where $\tilde{\boldsymbol{\Sigma}}_{kt} = \mathrm{Tr}\left[\boldsymbol{\Sigma}_{kt}\right]$ and $\mathbf{K}$ is an *artificial* regressor whose elements are calculated as

$$K_{kt}(\omega) \equiv X_{kt}(\omega)\left[\frac{1}{2} X_{kt}(\omega) - \sum_{\omega' \in \Omega_{i,k}} \lambda_{kt}(\omega') X_{kt}(\omega')\right].$$

Following this approach and omitting higher-order terms, we obtain an approximated version of $\xi$, denoted by $\tilde{\xi}$. We then estimate the demand parameters by exploiting the moment condition $\mathbb{E}\left[\Delta\tilde{\boldsymbol{\xi}} \mid \mathbf{z}\right] = 0$, which is similar to running a linear 2SLS regression. Given that $\ln p \subset \mathbf{X}$, markups can be recovered as $\mu = \frac{\partial \ln \lambda}{\partial \ln p}\left(1 + \frac{\partial \ln \lambda}{\partial \ln p}\right)^{-1}$, assuming single-product and profit-maximizing firms.

The next challenge is finding a valid instrument to guide identification with limited data on observed product characteristics. Our dataset reports three observable characteristics: the country of origin, the product classification used by the statistical agency, and the unit price ($p$). The demand residual *conditional* on these characteristics, $\tilde{\xi}$, is presumably contaminated with omitted variables correlated with $p$—unlike small-scale estimations like BLP, where $\xi$ is purged from a wider range of observable product characteristics using richer data. To overcome this identification challenge, we leverage high-frequency transaction data and interact it with high-frequency exchange rate data to construct a granular shift-share instrument for $\ln p$ that measures the exposure to exchange rate fluctuations at the variety level and is uncorrelated with $\tilde{\xi}$. We begin with the observation that the year-over-year change in the unit price of variety $\omega$ can be approximated by the sales-weighted average of monthly price changes:

[17] More specifically, the utility of household $h$ derives from purchasing variety $\omega$ is $\left(\overline{\boldsymbol{\beta}}_{kt} + \boldsymbol{\epsilon}_{h,kt}\right) \cdot \mathbf{X}_{kt}(\omega) + \xi_{kt}(\omega) + u_{h,kt}(\omega)$, where $u$ accounts for idiosyncratic heterogeneity in taste for product varieties, which is distributed *iid* according to a type-I Extreme Value distribution with scale parameter 1.

$\Delta \ln p_{kt}(\omega) = \sum_{m \in \mathbb{M}_t} \rho_{kt}(\omega, m) \Delta \ln p_{kt}(\omega, m)$, where $\rho_{kt}(\omega, m)$ and $p_{kt}(\omega, m)$ denote month $m$'s share of export sales and the year-over-year change in export prices in month $m$ of year $t$ (i.e., $m \in \mathbb{M}_t$). Since $p_{kt}(\omega, m)$ is denominated in the destination market's currency, it varies with the year-over-year change in the exchange rate between firm $\omega$'s origin country and the destination market it serves in month $m$, denoted as $\mathscr{E}_t(\omega; m)$. Motivated by this accounting relationship, we construct the shift-share instrument: $z_{kt}(\omega) = \sum_{m \in \mathbb{M}_t} \rho_{kt-1}(\omega; m) \Delta \ln \mathscr{E}_t(\omega; m)$. This instrument interacts the lagged export share with the concurrent exchange rate change per month to measure variety-level exposure to aggregate exchange rate fluctuations. The exposure measure $z$ is uncorrelated with $\tilde{\xi}$ under the identifying assumption that aggregate exchange rate fluctuations and past export composition are independent of *unobserved* concurrent demand shocks.

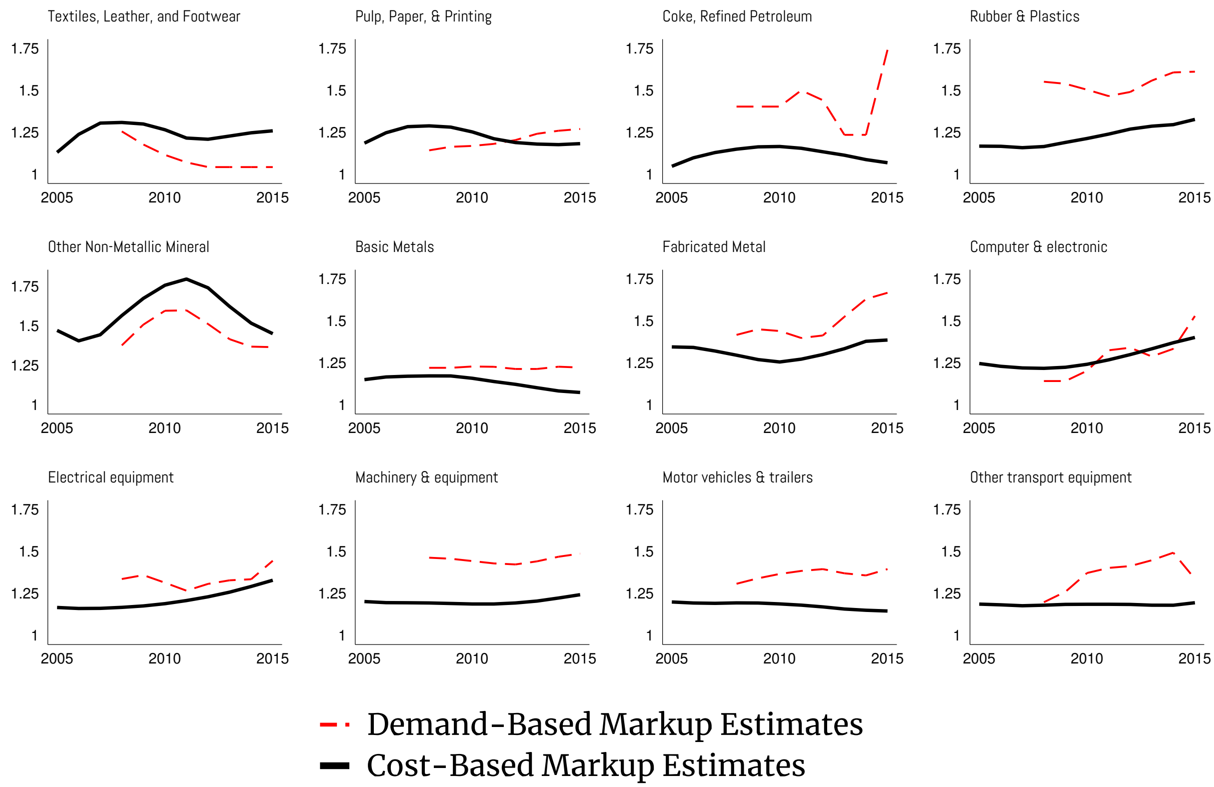
Our estimation uses the universe of import transactions for Colombia from 2007 to 2016. The dataset encompasses over 93,000 firms from 251 different countries and reports high-frequency transaction-level sales and quantities for individual firms exporting to Colombia at the Harmonized System 10-digit product level. We complement this data with matching high-frequency exchange rate data from the Bank of Canada for the same time period. To fully leverage the granularity of our data, we conduct our estimation using market share and price data for 10-digit product categories. However, to ensure compatibility between our estimated markups and the level of aggregation in the ICIO data, we pool all 10-digit product categories and estimate demand parameters at the ICIO industry level. Appendixes A and K provide further details about our data and estimation methodology.

Figure 6 displays the estimated markups for select manufacturing industries during 2005-2015, based on both demand-based and cost-based approaches. The graph displays the sales-weighted average markup for each industry in a given year. Since 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. As anticipated, there are some discrepancies between the demand-based and cost-based markup values. However, in many industries, the 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 deadweight loss of markups and its response to trade via rent-shifting.

## 8.2 Measuring Cross-Country Profit Payments

We assemble data on profit ownership shares, $\{\pi_{in,t}\}$, using the Orbis Database provided by Bureau van Dijk (BvD). We first clean and refine the data using the algorithm described in Appendix A. The cleaned dataset forms a panel consisting of 3,075,899 firms globally from 2005 to 2015. For each firm $\omega$ in this sample, we have information on its gross profits, denoted as $\varpi_{it}(\omega)$, in year $t$, where the subscript $i$ represents the country in which the firm's operation is based. Additionally, we observe the firm's equity share associated with shareholders located in country $n$, denoted as $\kappa_n(\omega) \in (0, 1]$. Using this information, we calculate the share of country $i$'s profits repatriated to country $n$ in year $t$ via equity

Figure 6: The variation in estimated markups over time: *manufacturing industries*




*Note*: Cost-Based markups are estimated using the WORLDSCOPE and COMPUSTAT data. Demand-based markups are estimated using Colombian import transactions from DATAMYNE (see Appendix A). Industry classifications are based on the ICIO TABLES.

financing using the following formula:

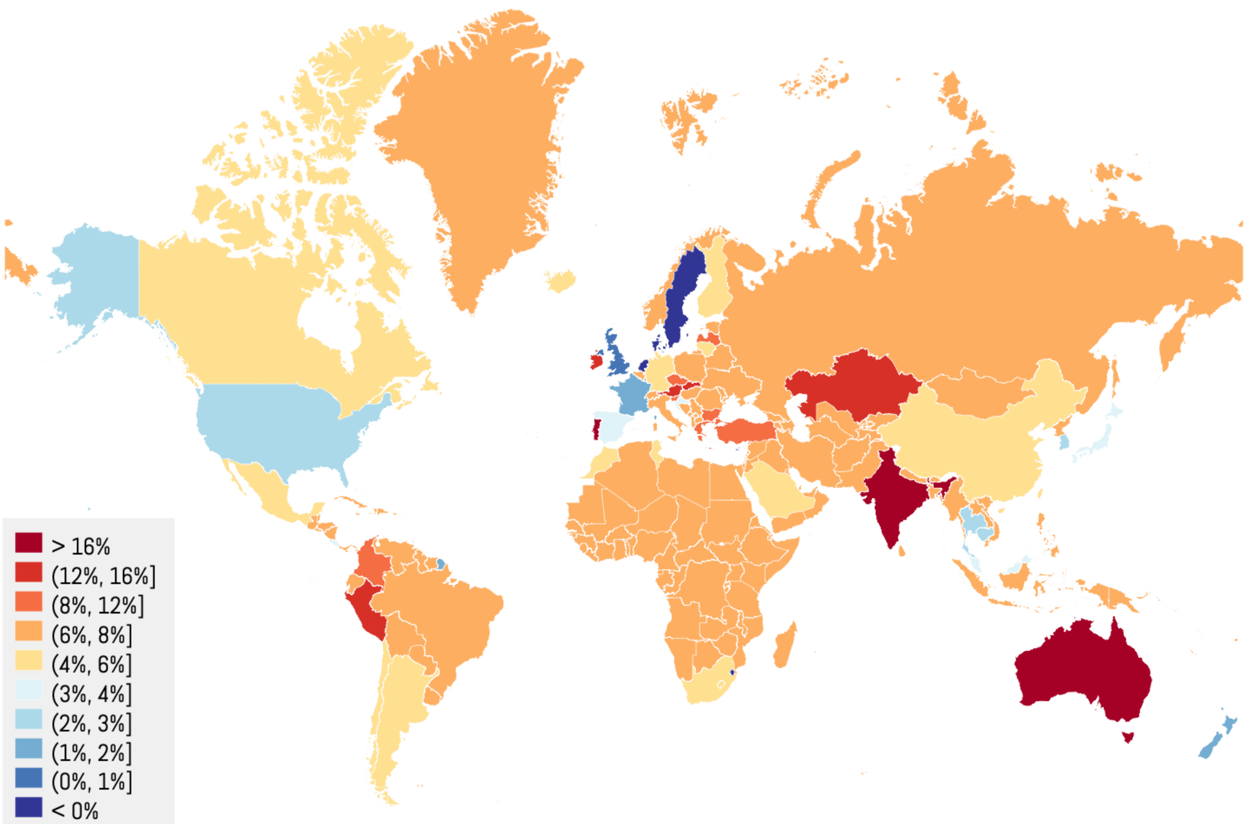
$$\pi_{in,t} = \frac{\sum_{\omega \in \Omega_{i,t}} \varpi_{i,t}(\omega)\ \kappa_n(\omega)}{\sum_{\omega \in \Omega_{i,t}} \varpi_{i,t}(\omega)},$$

where $\Omega_{i,t}$ denotes the set of firms operating in country $i$ in year $t$ in our sample. By applying this formula for each triplet $(i, n, t)$, we obtain square matrices of bilateral profit ownership shares for each year in 2005-2015 that are compatible with ICIO tables. Table A2 in the appendix provides an overview of multinational profit ownership. For each country in the ICIO sample, it reports the share of profits retained in the country of origin, repatriated to high-income countries, and repatriated to low/middle-income countries.

## 8.3 The Global Rise in the Deadweight Loss of Markups

In this section, we report the deadweight loss of markups for various countries, which is the sum of the welfare loss due to markup dispersion and the country's exposure to international rent-shifting externalities. We compute the deadweight loss by plugging our estimated markup values and share data into our semi-parametric formula for $\mathscr{D}_i$. Figure 7 presents the results when multi-national profit payments are accounted for. The deadweight loss of monopoly distortions is noticeably higher in low-income regions. Remarkably, some high-income countries, such as the Netherlands, actually benefit from markup distortions.[18] This indicates that for these countries, the positive gains from rent-shifting more than offset the loss from markup dispersion. However, as noted earlier, rent-shifting effects are zero-sum, meaning that these benefits come at the expense of other nations, primarily low-income ones.

Figure 7: The deadweight loss from markup distortions across different countries



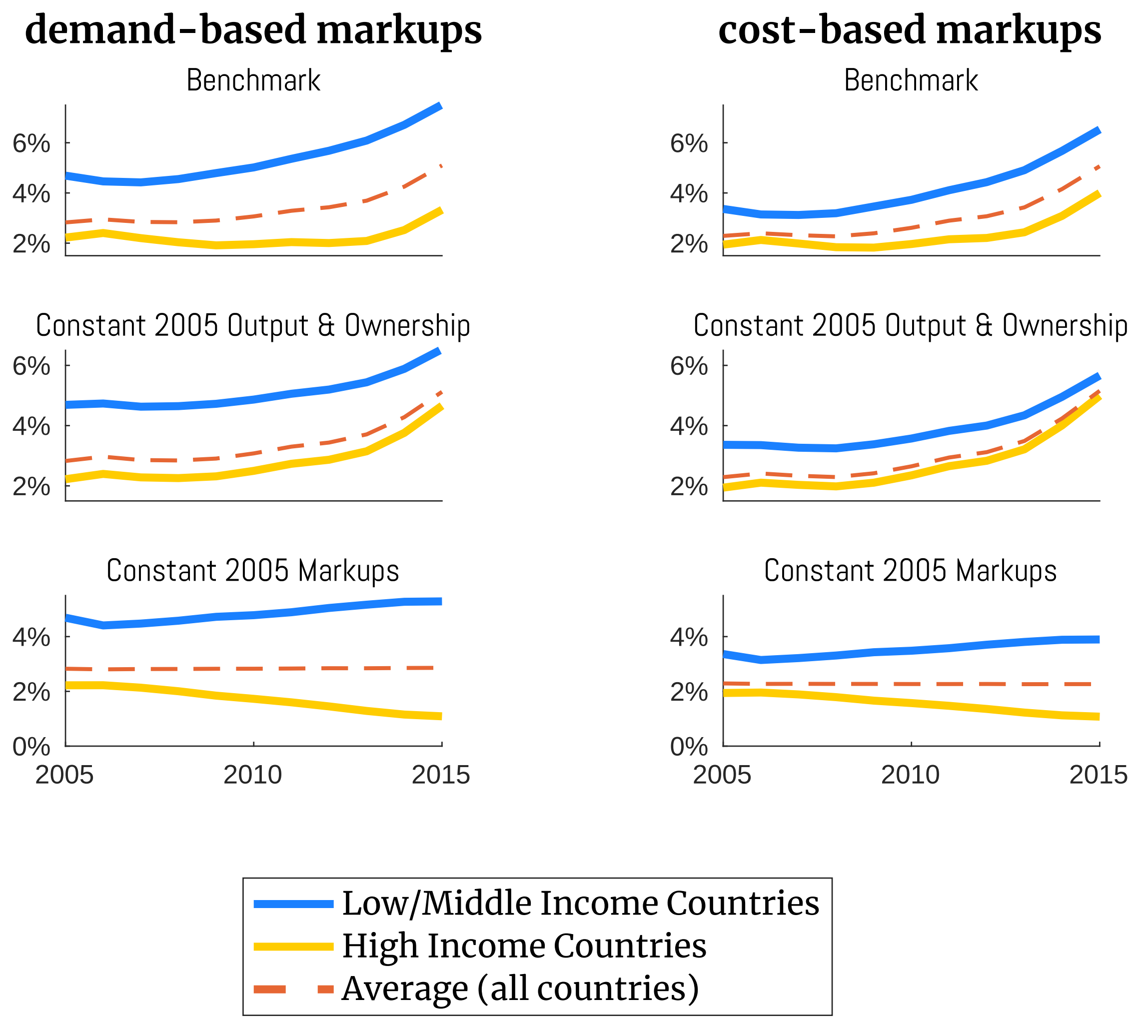

*Note*: This map reports the deadweight loss (DWL) of markups, measured as the per-cent loss in real consumption due to monopolistic markup distortions. The DWL is calculated using demand-based markup estimates in 2015 and accounts for multi-national ownership. Data on industry-level expenditure and production are from the Icio. Data on multinational ownership are from Orbis.

We next examine whether the deadweight loss of markups has increased over time. Figure 8 presents the findings, depicting the change in the deadweight loss of markups between 2005 and 2015. The y-axis denotes the deadweight loss, quantified as the percentage loss in real consumption due to markup distortions. The figure presents GDP-weighted averages for high-income and low/middle-income groups, categorized according to the classification outlined in Table A4. The left panel showcases the deadweight loss calculated using demand-based markup estimates, while the right panel displays results derived from cost-based markup estimates. The results in Figure 8 point to substantial welfare losses, but it is important to recognize that these estimates may still understate the true extent of the loss, as they do not take

[18]It is important to emphasize that without trade, all countries would have experienced losses from markup distortions.

into account the amplification from input-output linkages demonstrated in Figure A8 in the appendix.[19]

Figure 8: The rising deadweight loss of markups and its drivers over time




*Note*: The above graph reports the deadweight loss of markups and its change over time for high versus low/middle income countries. A 5% deadweight loss implies that markups lower real national consumption by 5% relative to its efficient level. The deadweight loss is calculated by plugging our estimated markup values and output/expenditure share data into our formula for $\mathcal{D}_i$, accounting for multinational profit payments. The figures in the middle panel are computed by assuming that output/expenditure shares and cross-country profit payments 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 shares are from the Icio.

Figure 8 clearly illustrates that markups result in a greater deadweight loss for low-income countries compared to high-income nations. Moreover, the deadweight loss has been steadily increasing over time, with the trend being particularly acute among low-income nations. While these results are consistent with the existing literature on the *rise of market power*, there are two noteworthy aspects that set our analysis apart. First, rather than focusing on the sales-weighted average markup as De Loecker et al.

[19]A higher substitution elasticity between aggregate industries also amplifies the deadweight loss (DWL) of markup distortions. Figure A12 in the appendix shows that the DWL of markups rises significantly with a higher cross-industry elasticity of substitution. For the US, increasing elasticity from 1 to 2 more than doubles the DWL of markups. This is because resource allocation across industries becomes more sensitive to markup distortions as substitutability increases.

(2020), we directly quantify the deadweight loss of markups based on theoretical foundation. This distinction is crucial, as the average markup can, in principle, increase without necessarily leading to a greater deadweight loss for the economy. Second, prior studies have primarily relied on cost-based markup estimates to document the rise in market power. However, Figure 8 suggest that the same pattern emerges even when demand-based markup estimates are employed.

The lower panels in Figure 8 decompose the change in markup distortions into changes driven by (1) adjustments to output composition and multinational profit payment over time, and (2) adjustments to markup levels over time. The middle panel takes a ceteris paribus approach, holding output and expenditure shares constant at their 2005 levels. This allows us to isolate the effect of changing markup levels on the deadweight loss measure, $\mathscr{D}_i$. Conversely, the bottom panel keeps markups fixed at their 2005 levels and tracks the change in $\mathscr{D}_i$ that can be attributed to shifts in the economic composition. The results from these lower panels are quite revealing. They demonstrate that the global increase in the deadweight loss of markups is entirely driven by changes in markup levels over time. Meanwhile, the change in output composition and multinational profit payment has merely redistributed the burden of markup distortions, shifting it from high-income countries to their low-income counterparts. These findings provide an initial glimpse into the zero-sum nature of rent-shifting effects, which are formally quantified in the following section.

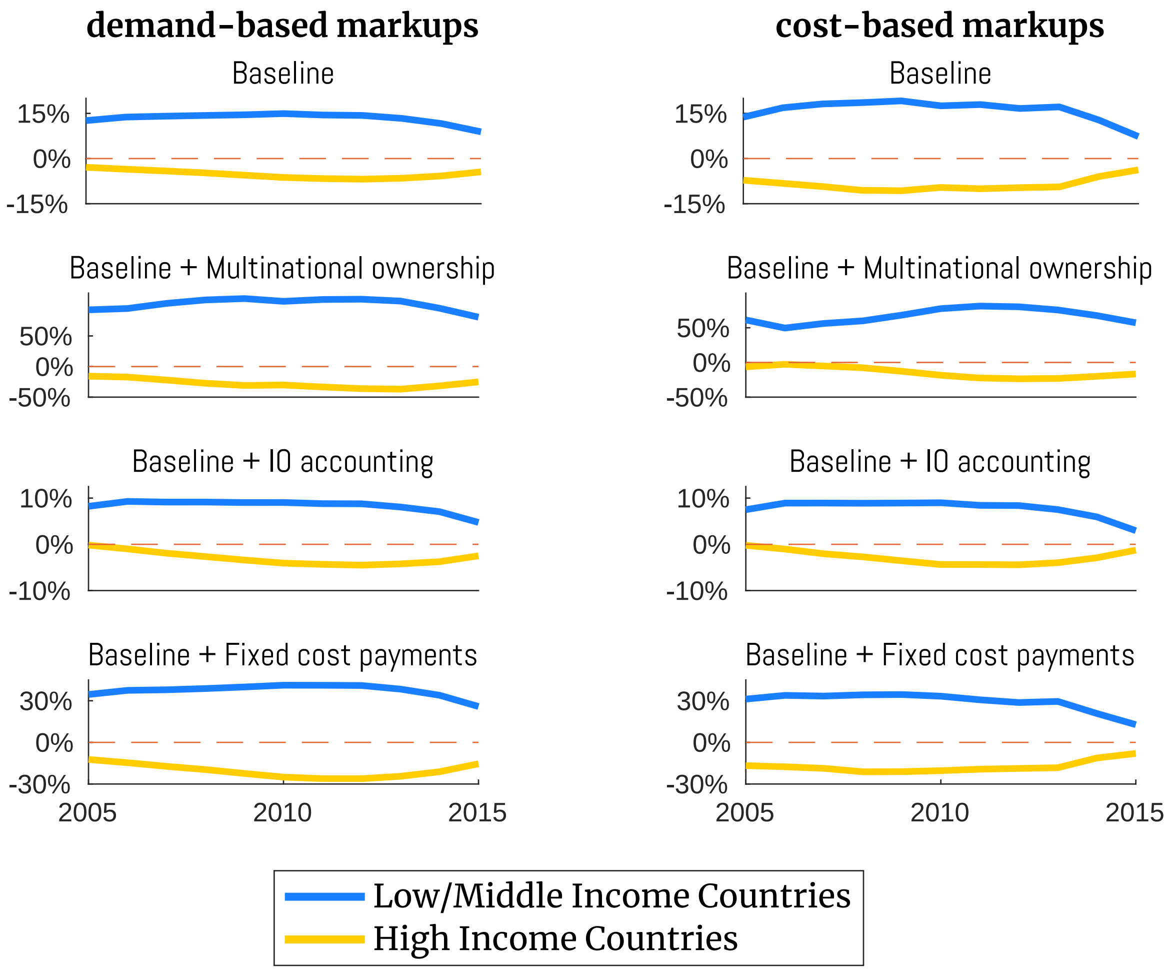
### 8.4 The Pro-Rich Bias of International Rent-Shifting Externalities

We invoke Proposition 2 to isolate how trade integration has shifted the burden of markup distortions internationally. For completeness, we measure the impact of trade under various considerations such as multi-national ownership, global input-output linkages, and fixed overhead costs that trim profits. It is important to note that from the lens of our semi-parametric model, trade modifies the burden of markup distortions solely through rent-shifting effects. The logic is that trade integration prompts specialization based on comparative advantage, dampening the deadweight loss of markups for countries that specialize in high-markup product categories, while amplifying it for others through rent-shifting.

Figure 9 displays the change in the deadweight loss of markups due to trade integration, reporting average effects across low- and high-income country groups. The results reveal that through international rent-shifting, trade has shifted the burden of markups from high-income nations to low-income countries. This finding is also robust to the method used for markup estimation and persists even after accounting for multi-national ownership, global input-output networks, and fixed overhead costs.

Our findings, averaged across all years and specifications (such as demand-based and cost-based markup estimation), indicate that trade integration has had a significant impact on the deadweight loss of markups for low- and middle-income countries. On average, it has increased the deadweight loss for these countries by 44% while simultaneously reducing it by 15% among high-income nations. These effects represent substantial transfers between countries that occur solely through international rent-shifting, a phenomenon that has been largely overlooked in previous literature. The existing literature has mainly focused on the pro-competitive effects of trade, which reduce markup dispersion and are

Figure 9: Trade-induced change in the deadweight loss of markups through international rent-shifting




*Note*: The above graph reports the percent change in the deadweight loss (DWL) of markups due to trade openness. For example, a 5% change represents to a 5% increase in the DWL of markups defined in Section 5. To top row reported results obtained from our baseline model suing Proposition 2. The next row accounts for multi-national ownership (Equation 10). the third account for global input-output linkages as explained in Appendix G. The last row accounts for fixed overhead costs that trim profits as detailed in Appendix H. Data on industry-level expenditure, production and input-output shares are from the ICIO. Data on global profit ownership are from ORBIS. Cost-Based markups are estimated using data from WORLD SCOPE and demand-based markups are estimated using transaction level import data from Colombia provided by DATAMYNE.

internationally symmetric, with some studies finding these effects to be relatively small.

The asymmetric effects of trade on the deadweight loss of markups become even more pronounced when considering the role of multi-national ownership and the repatriation of profits to foreign shareholders. This finding is consistent with our previous empirical observation that profits earned by multinational corporations are primarily repatriated to shareholders in high-income countries. When global input-output (IO) linkages are accounted for, the impact of trade is somewhat attenuated, although the directionality of the effects remains the same. This attenuation occurs because IO linkages amplify the deadweight loss of markup dispersion while diluting the extent of rent-shifting, making the rent-shifting component of the deadweight loss less consequential. When fixed overhead costs are considered, the asymmetric effects of trade are amplified, suggesting that fixed cost payments incurred in foreign markets contribute to rent-shifting, as low-income countries paying net quasi-rents to high-income partners in the form of fixed cost payments.

It is important to note that the results emerging from Figure 9 are not apparent a priori. While rent-shifting effects are zero-sum by nature, there is no inherent reason to believe that they favor high-income countries. That being said, Figure 9 masks the heterogeneity in exposure to rent-shifting within income groups. To delve deeper into this aspect, Figures A9 and A10 in the appendix provide a more granular visualization of the impacts of rent-shifting effects, highlighting heterogeneous effects even within low and middle-income groups. For example, rent-shifting is less detrimental for the Chinese economy but extremely costly for African countries.[20] Relatedly, Figure A11 in the appendix visualizes the flow of markup rents on a bilateral basis, although interpreting these flows is more intricate due to issues related to balanced trade and the fact that these flows do not directly translate to welfare effects without proper normalization.

### 8.5 Discussion of Results and Limitations

This section explores potential reasons behind the pro-rich bias of international rent-shifting effects and examines why these effects have diminished over time. We also discuss important limitations that should be taken into account when interpreting our findings.

*The Deep Origins of International Rent-Shifting.* Based on our formulas, a country's pattern of specialization across low- and high-markup industries determines its exposure to international rent-shifting, with countries specializing in high-markup industries benefiting at the expense of others. Following this logic, the results in Figure 9 imply that high-income countries tend to have a comparative advantage in high-markup industries, which raises the question: What factors contribute to this pattern of comparative advantage? Theoretically, several papers show that high-income countries have a generic comparative advantage in high-markup product categories. Fajgelbaum et al. (2011) attribute this pattern to the home-market effect, where strong local demand for high-quality, high-markup goods and economies of scale prompt local specialization in these product categories. As a result of this specialization, high-income countries become net exporters of high-markup goods to low-income regions. Lashkaripour (2020) attributes this pattern of specialization to the fact that firms in high-income countries operate with higher input costs, and as a result, have a revealed comparative advantage in high-markup product segments where demand is less sensitive to the high production cost. Alternatively, this pattern of specialization can be rooted in the different factor and institutional endowments of low- and high-income countries. We explore this possibility in Appendix L. Figure A1 (in this appendix) plots the relationship between the national-level loss from rent-shifting ($\Delta\mathscr{D}_i$) and national institutions and resource endowments across countries. The findings suggest that countries with better legal and credit market institutions tend to specialize in high-markup industries and benefit from rent-shifting.

[20] For most countries, the impacts of international rent-shifting have a stable sign across specifications. In a few instances, however, accounting for input-output linkages or firm-selection reverses our baseline predictions.

Conversely, countries with abundant natural resources are more likely to experience adverse exposure to rent-shifting effects, echoing the resource curse argument. What is apparent from this exercise is that institutions that foster specialization in high-markup industries are correlated with income per capita, hence the pro-rich bias of international rent-shifting effects as documented in Figure 9.

*The Evolution of Rent-Shifting Effects Over Time.* Figure 9 indicates that rent-shifting from low- to high-income nations may have dampened over time. This trend can be attributed to two possible factors. First, middle-income nations may have become more specialized in high-markup industries between 2005 and 2015. Second, markup levels may have evolved in a way that dampens rent-shifting from low-to-high-income nations. We examine these two possibilities in Appendix M. Our analysis reveals that the dampening effect is almost entirely explained by changes in North-South specialization patterns. That is, low and middle income nations have become increasingly specialized in sophisticated, high-markup industries, dampening the extent to which markup rents flow out of these economies to high-income trading partners—demonstrated by the bottom panel of Figure A3 in Appendix M.[21]

*Discussion of Limitations.* Our results provide fresh insights into a largely-overlooked aspect of globalization, but we must highlight some limitations. A basic limitation, shared among macroeconomic studies of market power, is that our markup estimates may not be representative. Our cost-based markup estimates are derived from the balance sheets of publicly traded firms. Facing the same limitation, De Loecker et al. (2020) re-weight industries using weights from the United States' Bureau of Economic Analysis and find that their results are robust to the re-weighting. While this exercise increases confidence in the use of publicly traded firms' balance sheet data, the representativeness of cost-based markup estimates remains a subject of debate. In contrast, our demand-based markup estimates encompass all firms serving the Colombian market. Through the lens of our theoretical model, we can infer the distribution of firm-level markups (for each product) in other markets based on the Colombian data. In fact, our theory allows for pricing-to-market, but the markup distribution is not sensitive to this consideration under the Pareto assumption. Therefore, our demand-side markup estimates are globally representative to the extent that the Pareto assumption about firm productivity holds in the data. Despite these limitations, the alignment of results derived from demand-based and cost-based markup estimates is an encouraging sign. Another potential limitation is that our measurement focuses solely on output market power, which can be problematic for two reasons. First, we estimate the output-side markup using price-to-marginal-product ratios derived from revenue elasticities. However, Bond et al. (2021) argue that these ratios may not inform us about the output-side markup but rather the input-side markdown due to monopsony power in input markets. Second, if we replace output-side markups with input-side markdowns, the burden of distortion shifts. In that case, trade amplifies the burden of monopsony distortions on workers in countries whose wages are subject to greater markdowns, which, based on Bond et al. (2021)'s arguments, would be workers in high-income nations. Despite this limitation, two

[21] We use a static classification of countries from the United Nations when dividing our sample into low- and high-income countries. An alternative approach is to classify countries according to their real GDP per capita. The reduced-form analysis conducted using this alternative classification in Section 3 does not suggest a dampening of North-South rent-shifting effects.

factors could instill trust in our result: *(a)* we find the same pattern of rent-shifting using demand-side markup estimates that are immune to this critique, and *(b)* De Ridder et al. (2022) demonstrate that while reliance on revenue data may bias the estimated level of markups, the correlation between the estimated and actual markup remains strong.

### 8.6 Global Policy Remedies for Rent-Shifting Externalities

To explore potential global policy solutions for rent-shifting externalities, it is important to understand their nature. When firms engage in monopolistic pricing, they generate monopoly rents by distorting prices and reducing consumer surplus. As shown in Section 7, this behavior is essentially a form of *decentralized* terms of trade manipulation. Monopolistic pricing creates rent-shifting externalities that function like implicit tariffs, tilting the terms of trade to favor the countries that receive monopoly rents from the rest of the world. We quantitatively illustrate this by calculating these implicit tariffs for all countries using the methodology outlined in Appendix N. We calibrate the industry-level trade elasticity values to those estimated by Caliendo and Parro (2015). Country $i$'s effective tariff is then determined as $t_i + \widetilde{t}_i$, where $t_i$ is the applied (or explicit) tariff and $\widetilde{t}_i$ is the implicit tariff resulting from rent-shifting externalities.

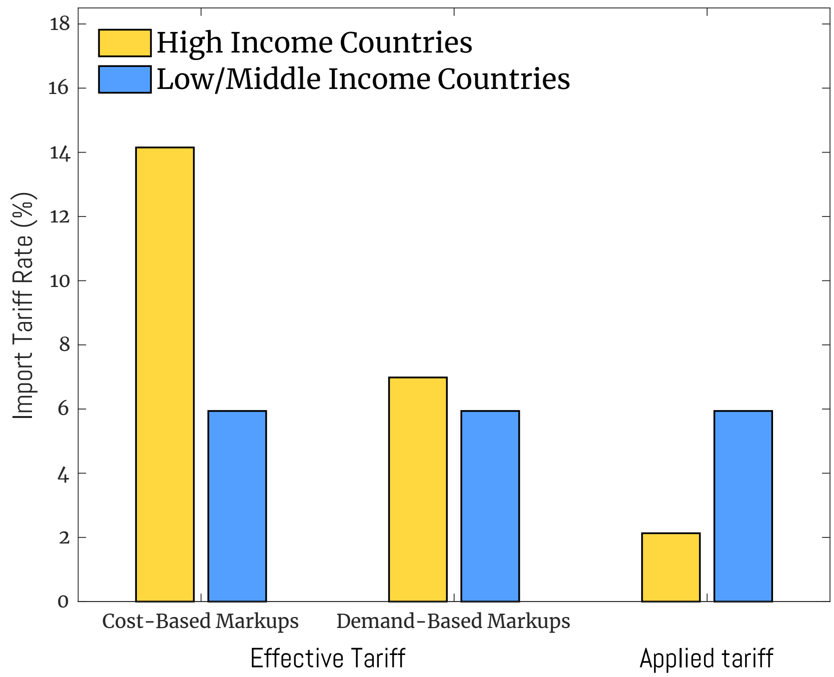
Figure 10 illustrates the effective tariffs and compares them to the applied tariffs. On the surface, high-income countries appear to levy lower tariffs than low-income countries, consistent with the WTO's generalized system of preferences (GSP). In 2015, high-income countries had a weighted average applied tariff of only 2.3%, while low and middle-income countries had a higher rate of 5.9%. This suggests that high-income countries have made more tariff concessions under the WTO framework. However, when accounting for rent-shifting externalities, the effective tariff landscape changes significantly. By incorporating implicit tariffs arising from rent-shifting, the effective tariff for high-income countries reaches 10.5%, based on the average of cost-based and demand-based markup estimates.[22] This finding indicates that high-income countries not only fail to provide additional tariff concessions under the GSP but also effectively impose an excess 7% tariff on imports from their low-income trading partners.

These results state that the rent-seeking behavior of firms in high-income countries aligns with the government's desire to manipulate the terms of trade—possibly, discouraging regulation of anti-competitive behaviors that would otherwise be addressed in a closed economy. Accordingly, while monopolistic pricing practices reflect a domestic policy failure in a closed economy, they amount to a negative international externality in a global setting. Shallow cooperation, thus, entails that governments tackle the rent-shifting externality to prevent adverse impacts on their trading partners—at least based on the basic principles underlying the WTO.

The first-best policy to address rent-shifting externalities is internationally coordinated markup cor-

[22] As explained in Section 7, Figure 10 illustrates *weak* duality between tariffs and markup distortons. Figure A13 depicts *strong* duality between tariffs and markup distortions, whereby there exists a vector of tariffs that could replicate not only the rent-shifting externalities associated with markups but the entire welfare loss from markup distortions.

Figure 10: The applied tariff vs effective tariff under international rent-shifting externalities




Note: This figure reports the effective tariffs (applied tariffs + implicit tariffs due to rent-shifting effects) and compares them to applied tariffs for each country group. The tariff rates for each country group are the GDP-weight average rates and calculated using the algorithm described in Appendix N. Our calculation uses expenditure and output data from the ICIO tables and our estimated markups (described in Section (8)) in year 2015. Industry-level trade elasticity values are from Caliendo and Parro (2015).

rection. However, implementing this solution is challenging within the WTO's current framework, which focuses on regulating and coordinating border policies rather than domestic policy measures. Therefore, we propose two alternative policy solutions. The first leverages existing mechanisms within the WTO, advocating for a revised interpretation of the reciprocity principle. The second solution could be integrated into the evolving global minimum tax agreement.

*WTO-based Remedy for International Rent-Shifting.* The World Trade Organization (WTO) operates on the principle of reciprocity, requiring member countries to make balanced tariff concessions that benefit all parties involved. However, there are exceptions to this rule, such as the Generalized System of Preferences (GSP), which allows high-income countries to grant preferential market access to disadvantaged economies. Our findings indicates that to achieve *strict* reciprocity, the GSP should be further utilized by high-income nations. Specifically, we find that these countries must make an additional 6% tariff concession under the GSP scheme, which could be implemented by eliminating their tariffs on low-income nations entirely and permitting them to increase the tariff cap by 3.7%.

The logic behind this result can be understood as follows: Suppose the trade elasticity is uniform across industries and all countries as small open economies. Moreover, assume that all firms in country $i$ impose a uniform markup $\mu_i$, which is higher the higher-income country $i$. Non-cooperative Nash tariff in this scenario would be $t_i^N = 1/\theta$, as implied by the optimal tariffs formulas in Lashkaripour (2021). However, the effective Nash tariff would be $t_i^N = 1/\theta + \tilde{t}_i(\mu_i)$, where $\tilde{t}_i(.)$ represents the implicit tariff due to rent-shifting, which increases with $\mu_i$. Under the WTO, countries typically exercise *first difference* reciprocity, which allows them to eliminate the systematic non-cooperative tariff term, $1/\theta$. But this approach would result in high-income countries effectively charging a higher tariff after the tariff liberalization episode. To ensure strict reciprocity, high-income countries must make additional concessions based on $\tilde{t}_i\left(\mu_i\right)$, which our calculations estimate to be around 6%.

*Remedy Based on Global Tax Agreement.* The previous policy proposal could effectively address rent-shifting externalities, but implementing it may be politically challenging in the current environment. We propose an alternative policy that, although only partially effective, can be readily incorporated into Pillar One of the global minimum tax agreement under the Base Erosion and Profit Shifting (BEPS) project. Pillar One seeks to shift taxing rights, enabling the taxation of profits where multinational companies have significant consumer-facing activities and generate profits. Let $\tau^{global}$ represent the global destination tax rate on profits, with revenues collected at the point of sale. This tax scheme could potentially be integrated into Pillar One of the BEPS project. The exposure to rent-shifting externalities under this global tax scheme can be expressed as follows:

$$\Delta\mathscr{D}_i^{(\tau)} = \ln\left(\tau^{global} + \left(1-\tau^{global}\right)\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}\left[\mu\right]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}\left[\mu\right]^{-1}}\right),$$

indicating that a global destination tax on profits mitigates rent-shifting externalities. In the extreme case where $\tau^{global}$ approaches 1, rent-shifting externalities are entirely eliminated. Table 1 reports the effectiveness of a global destination profit tax in reducing rent-shifting externalities at various rates (15%, 30%, and 45%). The first column presents the magnitude of rent-shifting effects in the absence of a global tax, averaged across all specifications and years in our sample. The remaining columns demonstrate that rent-shifting effects are reduced with a global destination tax on profits. With a 45% tax, the deadweight loss of markups for low-income countries is only amplified by 28% through trade relations, compared to 44% without such a tax. It is clear for these result that this taxing scheme is not as effective as unilateral tariff concessions in mitigating rent-shifting, but it could serve as a politically viable alternative—one that can be incorporated into Pillar One of the evolving global tax agreement.

Table 1: Mitigating Rent-Shifting Externalities through a Global Destination Tax on Profits

| | no global tax | $\tau^{global}=15\%$ | $\tau^{global}=30\%$ | $\tau^{global}=45\%$ |
|---|---|---|---|---|
| **$\Delta\mathscr{D}$ (low-income)** | 43.9% | 39.0% | 33.7% | 27.9% |
| **$\Delta\mathscr{D}$ (high-income)** | -14.8% | -12.3% | -9.9% | -7.7% |

*Note*: This table reports the trade-induced change in the DWL of markups ( $\Delta\mathscr{D}_i^{(\tau)}$) under various rates of an internationally coordinated destination tax on profits. The data on expenditure and output shares are from the ICIO. Markups are estimated using demand-based and cost-based methods, with the reported results representing the average effects across the two estimation methods.

## 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 have led to substantial welfare transfers from low-income to high-income countries through international rent-shifting externalities. These effects are akin to implicit tariffs that distort the terms of trade in favor of high-income countries. This observation suggests that, contrary to prevailing wisdom, low-income countries have made greater concessions under the current system of global trade agreements. To create a more level playing field, we propose two policy reforms that can mitigate the burden of international rent-shifting on low-income countries. The first reform involves high-income countries making unilateral tariff concessions under the WTO's Generalized System of Preferences (GSP) mechanism. The second reform entails the implementation of a destination tax on profits, which, while only partially effective, may be more viable from a political economy standpoint. These policy solutions require international coordination among cooperative governments. In the absence of global cooperation, the rent-shifting externalities highlighted in this paper can be addressed unilaterally. Characterizing unilaterally optimal policy remedies and evaluating their effectiveness presents a promising direction for future research.

## 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.

Adão, R., C. Arkolakis, and S. Ganapati (2020). Aggregate implications of firm heterogeneity: A nonparametric analysis of monopolistic competition trade models. Technical report, National Bureau of Economic Research.

Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review 107*(3), 633–689.

Adhami, M., J.-F. Brouillette, and E. Rockall (2024). Markups, firm scale, and distorted economic growth.

Aghion, P., A. Bergeaud, T. Boppart, P. Klenow, and H. Li (2022). Good rents versus bad rents: R&d misallocation and growth. Technical report, Working Paper.

Alviarez, V., M. Fioretti, K. Kikkawa, and M. Morlacco (2023, May). *Two-Sided Market Power in Firm-to-Firm Trade*.

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.

Ardelean, A. and V. Lugovskyy (2023). It pays to be big: Price discrimination in maritime shipping. *European Economic Review 153*, 104403.

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.

Baqaee, D. and E. Farhi (2019). Networks, barriers, and trade. 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.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics 121*, 1–14.

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.

Bown, C. P., F. Parro, R. W. Staiger, and A. O. Sykes (2023). Reciprocity and the china shock.

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.

Coşar, A. K., P. L. Grieco, S. Li, and F. Tintelnot (2018). What drives home market advantage? *Journal of international economics 110*, 135–150.

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.

De Ridder, M., B. Grassi, G. Morzenti, et al. (2022). The hitchhiker's guide to markup estimation.

DellaVigna, S. and M. Gentzkow (2019). Uniform pricing in us retail chains. *The Quarterly Journal of Economics 134*(4), 2011–2084.

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

Dhyne, E., A. K. Kikkawa, and G. Magerman (2022). Imperfect competition in firm-to-firm trade. *Journal of the European Economic Association 20*(5), 1933–1970.

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

Dix-Carneiro, R., P. Goldberg, C. Meghir, and G. Ulyssea (2021). Trade and informality in the presence of labor market frictions and regulations. *Economic Research Initiatives at Duke (ERID) Working Paper* (302).

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 (2023). How costly are markups? *Journal of Political Economy 131*(7), 1619–1675.

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

Errico, M. and D. Lashkari (2022). The quality of us imports and the consumption gains from globalization. *Boston College*.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2011). Income distribution, product quality, and international trade. *Journal of Political Economy 119*(4), 721–765.

Farrokhi, F., A. Lashkaripour, and H. S. Pellegrina (2024). Trade and technology adoption in distorted economies. *Journal of International Economics 150*, 103922.

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 (2023). Import and export markups, production relocation, and the gains from trade. *Production Relocation, and the Gains from Trade (November 25, 2020).*

Foster, L. S., J. C. Haltiwanger, and C. Tuttle (2022). Rising markups or changing technology? Technical report, National Bureau of Economic Research.

Ganapati, S. (2021). Growing oligopolies, prices, output, and productivity. *American Economic Journal: Microeconomics 13*(3), 309–327.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review 110*(8), 2586–2624.

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.

Hottman, C. J. (2017). Retail markups, misallocation, and store variety across us cities.

Hsieh, C.-T. and E. Rossi-Hansberg (2023). The industrial revolution in services. *Journal of Political Economy Macroeconomics 1*(1), 3–42.

Hummels, D., V. Lugovskyy, and A. Skiba (2009). The trade reducing effects of market power in international shipping. *Journal of Development Economics 89*(1), 84–97.

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

Keller, W. and S. R. Yeaple (2020). Multinationals, markets, and markups. Technical report, Working Paper.

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. (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 (2023). Grounded by gravity: A well-behaved trade model with industry-level economies of scale. *American Economic Journal: Macroeconomics 15*(2), 372–412.

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

Lashkaripour, A. (2020). Within-industry specialization and global market power. *American Economic Journal: Microeconomics 12*(1), 75–124.

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 (2023). Profits, scale economies, and the gains from trade and industrial policy. *American Economic Review 113*(10), 2759–2808.

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

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.

McFadden, D. and K. Train (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics 15*(5), 447–470.

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

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The review of economic studies 75*(1), 295–316.

Morlacco, M. (2019). Market power in input markets: Theory and evidence from french manufacturing. *Unpublished, March 20*, 2019.

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.

Święcki, T. (2017). Intersectoral distortions and the welfare gains from trade. *Journal of International Economics 104*, 138–156.

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

Thisse, J.-F. and P. Ushchev (2016). When can a demand system be described by a multinomial logit with income effect? *Higher School of Economics Research Paper No. WP BRP 139*.

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

Voronina, M. (2022). Endogenous growth and optimal market power.

# Appendix (for online publication)

## A Data Sources

### A.1 Unido-Indstat

This dataset is provided by the United Nations Industrial Development Organization (UNIDO), and is accessible through the UNIDO Data Portal. The data can be downloaded after registering for access on the UNIDO website, and it includes comprehensive industrial statistics covering a wide range of countries, years, and industries. We use the subsample corresponding to the 1980-2015 period, covering 196 countries and 23 ISIC rev.3 industries. For each industry and country in a given year, the data reports output, value added, wages and salary payments, number of employees, number of establishments, and gross fixed capital formation, among other variables. We use these variables to calculated aggregate accounting profits margin for each *industry-country-year* triad. We supplement this data with its derivative, the TradeProd database, developed and maintained by the Cepii. Users can access the TradeProd database through the official CEPII web portal after registration, which requires basic information and agreement to the terms of use. The TradeProd database only covers manufacturing industries, which are more traded and spans fewer years than the Unido-Indstat data. However, it reports domestic absorption measures per industry, allowing us to calculate net exports for each *industry-country-year* triad.

### A.2 The OECD Inter-Country Input-Output (ICIO) Tables.

The ICIO Tables (2018 edition) provides comprehensive information on international trade and production across major global economies. This dataset includes a sample of 64 major countries, covering 36 industries that span the entire economy from 2005 to 2015.[23] The dataset reports extensive information on trade flows across various origin-destination pairs and national level input-output tables disaggregated at the ICIO sector level. Users can access the ICIO data through the official OECD web portal after registration which requires basic user information and agreement to the terms of use. We use the ICIO data to construct country and industry-level output and expenditure shares ($y_{i,k}$ and $e_{i,k}$) as well as national-level input-output shares, $\alpha_{i,gk}$. In particular, from the ICIO tables, we can observe $X_{ni,k}$, which is the total flows of industry $k$ goods from origin country $n$ to destination country $i$. The expenditure share $e_{i,k}$ and output shares $y_{i,k}$ of country $i$ in industry $k$ in the baseline model are constructed

[23] ICIO tables include 64 countries (i.e. 36 OECD countries and 28 non-OECD economies), the Rest of the World and split tables for China and Mexico. In our analysis, we exclude the split tables for China and Mexico (i.e. CN1, CN2, MX1, and MX2).

as follows with this information:

$$e_{i,k} = \frac{\sum\limits_{n=1}^{65} (X_{ni,k})}{\sum\limits_{g=1}^{36}\sum\limits_{n=1}^{65} (X_{ni,g})}, \qquad\qquad y_{i,k} = \frac{\sum\limits_{n=1}^{65} (X_{in,k})}{\sum\limits_{g=1}^{36}\sum\limits_{j=1}^{65} (X_{in,g})}$$

### A.3 Compustat – North America

We conduct the production function estimation per industry using firm-level financial accounts data from Compustat – North America, which provides rich coverage regarding capital inputs and firm-level investments for publicly held companies in the United States and Canada. This database provided by S&P Global Market Intelligence and can be accessed through the Wharton Research Data Services (WRDS) platform, which requires an institutional subscription. Users affiliated with subscribing institutions can download the data after logging into the WRDS platform and navigating to the Compustat – North America database. The data is reported based on the SIC industry classification. So, we concord SIC industries into the 36 ICIO industries for which we have macro-level trade and production data using the following steps:

**(a)** we obtain the full sample of companies during the period 2003-2017 from the Compustat – North America database.

**(b)** we deflate firms' sales, cost of good sold, capital expense, staff expense, and general administrative expanse by U.S. GDP.

**(c)** we drop observations with negative sales, cost of good sold, capital expense, staff expense, general administrative expanse, and sales-to-cost ratio.

**(d)** since the database only reported the SIC industry classification, we have to concord SIC industries into the 36 ICIO industries for which we have macro-level trade and production data discussed in Section A.2. Unfortunately, we do not have official correspondence table mapping SIC to ICIO industry classification. Therefore, we concord SIC to ICIO by the following steps:

$$SIC \overset{(a)}{\rightarrow} ISIC\ rev.3 \overset{(b)}{\rightarrow} ISIC\ rev.3.1 \overset{(c)}{\rightarrow} ISIC\ rev.4 \overset{(d)}{\rightarrow} ICIO$$

The official correspondence tables of steps (a)-(c) can be found on the website of United Nation; and the correspondence table of step (d) can be found on the data descriptions from OECD Inter-Country Input-Output (ICIO) Tables. In addition, we also check and correct the correspondence tables manually by the verbal descriptions of each industry classifications to make sure we have the best match from SIC to ICIO. This step gives us 21,386 firms operating across 36 ICIO industries in United States and Canada from 2003 to 2017. After complaining the final data, we separately estimate the output elasticity for each industry and year during the 2005-2015 period using the control function method described in the main text.

### A.4 Worldscope

Worldscope is a database provided by Thomson Reuters, containing financial statement data and other financial information for publicly traded companies worldwide. The database can be accessed through the Thomson Reuters Eikon platform or Datastream, both of which require a subscription. Users with access to these platforms can download the data by searching for the desired firms and variables within the Worldscope database. In this paper, we process the data by the following steps:

**(a)** we download firm-level data of sales and costs of good sold from the Worldscope database during the 2005 - 2015 period.

**(b)** we drop observations with negative values on sales or costs of good sold. This steps give us 71,546 publicly traded firms operating across 987 SIC industries in 134 countries.

**(c)** some firms in this database operate in more than one industry, but we do not observe the breakdown of firm-level sales ($y$) and input costs ($c$) by industry. We treat a particular firm that operates in $n$ SIC-industries as $n$ different single product firms, and each firm is assumed to have sales as $y/n$ and cost as $c/n$. Then we can calculate firm's cost-sales ratio that will be used in the cost-based markup estimation process.

Table A1 reports the summary statistics including the average number of unique firms per country, the average number of industries served per firm, the average sales per firm, and the average input cost per firm . We report statistics for 63 main countries/regions in the ICIO database and the rest of the world.

### A.5 Orbis

Bureau van Dijk's Orbis database is the most comprehensive global resource on private firms. The dataset reports financial information on more than 489 million companies across regions and countries, which is originally collected from local registries and companies' annual reports. The database can be accessed through the Wharton Research Data Services (WRDS) platform, which requires a subscription. By paying a subscription fee, a user can search any firms if it exists in the database, and download the detailed information such as firm profile, consolidated and unconsolidated balance sheets, income statements, and the information of shareholders and subsidiaries. For our purpose, we first download the gross profits of all available firms during 2005-2015 (including very large, large, medium, and small companies) from the sub-dataset called "Financials for Industrial Companies" on the Orbis' online portal.[24] To clean the data of gross profits, we take the following steps:

[24]It should be noted that the Orbis' online portal updates company data when new data becomes available, however, it only provides 10 years of financial information for a company. Therefore, the available years of coverage depends on the last available year for a company's financial data. For example, when the latest financial data of a company becomes available in 2016, the Orbis'

Table A1: Worldscope database: Summary Statistics

| Country | number of firms | avg number of industries operated per firm | sales per firm (local currency) | input cost per firm (local currency) |
|---|---|---|---|---|
| Argentina | 107 | 4 | 3,253.1 | 1,813.2 |
| Australia | 2,041 | 3 | 462.9 | 259.7 |
| Austria | 102 | 4 | 1,446.5 | 823.8 |
| Belgium | 158 | 3 | 1,667.7 | 656.9 |
| Brazil | 429 | 4 | 4,750.4 | 2,392.2 |
| Bulgaria | 263 | 4 | 49.0 | 25.6 |
| Cambodia | 2 | 2 | 489,229.6 | 202,954.6 |
| Canada | 3,404 | 2 | 423.4 | 239.0 |
| Chile | 259 | 4 | 441,967.3 | 303,846.5 |
| China | 3,276 | 4 | 6,680.0 | 4,165.4 |
| Colombia | 79 | 5 | 2,139,877.0 | 1,089,354.0 |
| Costa Rica | 8 | 5 | 96,564.8 | 69,893.0 |
| Croatia | 114 | 5 | 1,127.9 | 798.9 |
| Cyprus | 128 | 3 | 75.4 | 43.3 |
| Czechia | 22 | 4 | 23,636.5 | 13,260.9 |
| Denmark | 373 | 3 | 3,571.3 | 3,264.2 |
| Estonia | 17 | 5 | 145.3 | 106.9 |
| Finland | 149 | 4 | 1,232.5 | 897.4 |
| France | 900 | 3 | 2,456.7 | 1,439.0 |
| Germany | 970 | 3 | 2,522.1 | 1,546.9 |
| Greece | 300 | 4 | 346.9 | 192.4 |
| Hong Kong (China) | 1,337 | 5 | 6,034.8 | 3,392.7 |
| Hungary | 47 | 3 | 164,571.6 | 108,459.8 |
| Iceland | 20 | 4 | 46,048.8 | 31,444.5 |
| India | 2,757 | 3 | 16,401.4 | 11,923.3 |
| Indonesia | 494 | 3 | 3,392,747.0 | 2,217,032.0 |
| Ireland | 79 | 3 | 1,286.8 | 832.1 |
| Israel | 546 | 3 | 1,330.9 | 785.4 |
| Italy | 336 | 4 | 2,605.7 | 1,213.4 |
| Japan | 4,064 | 5 | 177,394.3 | 122,260.1 |
| Kazakhstan | 62 | 3 | 72,720.4 | 37,474.9 |
| Korea | 1,870 | 4 | 1,050,410.0 | 758,646.7 |
| Latvia | 32 | 3 | 46.7 | 34.3 |
| Lithuania | 37 | 4 | 106.9 | 76.2 |
| Luxembourg | 65 | 3 | 2,156.2 | 1,860.3 |
| Malaysia | 1,059 | 5 | 959.4 | 611.1 |
| Malta | 22 | 3 | 57.0 | 13.1 |
| Mexico | 160 | 5 | 36,555.2 | 24,466.9 |
| Morocco | 76 | 3 | 3,248.9 | 2,145.1 |
| Netherlands | 208 | 4 | 4,730.3 | 3,059.1 |
| New Zealand | 176 | 3 | 510.5 | 342.4 |
| Norway | 270 | 3 | 6,537.8 | 3,950.7 |
| Peru | 182 | 4 | 877.8 | 481.2 |
| Philippines | 269 | 3 | 14,483.7 | 10,044.4 |
| Poland | 531 | 4 | 965.8 | 661.6 |
| Portugal | 59 | 5 | 1,426.4 | 1,035.8 |
| Rest of World | 2,822 | 3 | 13,700,000.0 | 728,454.4 |
| Romania | 162 | 4 | 406.3 | 240.1 |
| Russian | 992 | 3 | 34,532.2 | 16,526.5 |
| Saudi Arabia | 153 | 5 | 3,252.6 | 2,208.9 |
| Singapore | 714 | 4 | 587.1 | 418.0 |
| Slovakia | 25 | 4 | 298.1 | 248.2 |
| Slovenia | 54 | 5 | 275.5 | 206.4 |
| South Africa | 388 | 4 | 8,433.9 | 4,943.3 |
| Spain | 192 | 5 | 3,030.8 | 1,650.8 |
| Sweden | 602 | 3 | 5,804.3 | 3,539.6 |
| Switzerland | 302 | 4 | 3,126.6 | 1,294.5 |
| Chinese Taipei | 1,839 | 3 | 14,360.8 | 11,022.7 |
| Thailand | 666 | 4 | 13,478.8 | 10,678.3 |
| Tunisia | 65 | 3 | 154.9 | 90.5 |
| Turkey | 385 | 3 | 1,362.4 | 955.4 |
| United Kingdom | 2,367 | 2 | 875.6 | 614.6 |
| United States | 10,145 | 3 | 3,235.0 | 1,999.4 |
| Viet Nam | 698 | 7 | 1,166,592.0 | 795,627.7 |

*Note*: this table reports firm-level characteristics per country the average averaged across years 2005 to 2015. The average number of industries per firm is the average number of SIC industries served by firms in each country across years. The average sales and average input cost per firm are denominated in 1000,000 units of the local currency. The source of the data is Worldscope.

**(i)** For firms with multiple sources of gross profits in the same year, we first keep data with the filing type of "annual report" instead of "local registry filing". If we still observe multiple sources of gross profits for a particular firm, we only keep data with the consolidation type of "C2", which indicates that the financial statement is consolidated;

**(ii)** We assume that there's no cross-country profit payment by equity financing when a company is in deficit, so, we drop observations with negative gross profits in our dataset;

We then download the time-invariant shareholder information of all available firms from the sub-dataset called *"All Current Shareholders First Level"* on the Orbis' online portal.[25] This data contains information on all current shareholders of each firm in the database, which enables us to build links between a firm and its shareholders in different countries. With the reported information of equity shares for each shareholder, we can calculate the share of firm's profits that could be claimed by other countries through equity financing. We clean the data of shareholders by the following steps:

**(i)** Since Orbis only reports the firm's latest shareholder information without providing any information on the changes of ownership structure, we make an assumption that the firm's ownership structure is rarely changed over time.

**(ii)** We use the variable of *"shareholder – direct %"* as our primary measures for the equity shares of a particular shareholder, and we use *"shareholder – total %"* as supplement in the case the value of *"shareholder – direct %"* is missing.[26] Since Orbis Database may not have information for *all* shareholders of a company, we also assume the rest of missing equity shares are owned by the home country. For example, the Orbis Database reports that 30% equity shares of firm $A$ in country $i$ is owned by firm $B$ in country $j$; and 10% equity shares of firm $A$ is owned by firm $C$ in country $n$, then the rest of 60% equity shares are assumed to be owned by country $i$.

After merging firms' gross profits with the shareholders' information, we obtain a panel data of 3,075,899 firms from 2005 to 2015.[27] For each company $\omega$, we derive the share of country $i$'s profits repatriated to country $n$ at time $t$ via equity financing as: $\pi_{in,t} = \sum_{\omega\in\Omega_i} \left[\varpi_{it}(\omega)\,\kappa_n(\omega)\right] / \sum_{\omega\in\Omega_i}\left[\varpi_{it}(\omega)\right]$ ,where $\varpi_{it}(\omega)$ is the gross profits of firm $\omega$ operating in country $i$ at time $t$; and $\kappa_n(\omega) \in (0,1]$ is the equity share of firm $\omega$'s shareholders located in country $n$. By applying this formula for each triplet $(i,n,t)$, we get three-dimensional $63\times 63$ matrices of bilateral profit payments year between 2005 to 2015. Table A2 displays the average shares of profits rebated in the country of origin, repatriated to high-income

online portal will drop all data of this company before 2007 and we will only access to the data of this company from 2007 to 2016. In this paper, the data of gross profits was downloaded in May, 2024.

[25] In this paper, the data of shareholders was downloaded in May, 2024.

[26] The variable of "shareholder – direct %" represents the direct percentage owned by the shareholder in the company, while "shareholder – total %" represents the summation of direct and indirect percentages owned by the shareholder in the company. Since the variable of "shareholder – total %" has much more missing observations, we take "shareholder – direct %" as our primary measure.

[27] This dataset is an highly unbalance panel with 121,031 firms in 2005; 228,732 firms in 2006; 268,338 firms in 2007; 340,434 firms in 2008; 308,097 firms in 2009; 199,051 firms in 2010; 535,852 firms in 2011; 970,966 firms in 2012; 1,564,952 firms in 2013; 2,149,469 firms in 2014; and 1,860,007 firms in 2015.

countries, and repatriated to low/middle-income countries. Consistent with Figure 5 in the main text, the majority of profits are rebated in the firm's country of operation with repatriated profits accruing primarily to high-income shareholder.

### A.6 Transaction-Level Trade Data from Datamyne

We conduct our demand estimation using transaction-level trade records for Colombia purchased from Datamyne Inc. Access to the data was originally purchased from Datamyne in May 2014 and then again in June 2017. The data were available for manual online download in segments of five thousand observations per download. Each observation uniquely identifies the exporting firm and its country of origin, the 10-digit Harmonized System (HS10) product code under which the transacted goods are classified, and the exact time of the transaction. For each transaction, we observe the quantity and value of the goods imported, from which we construct data on market shares ($\lambda$), and unit prices ($p$). We supplement this data with daily exchange rate data between international currencies and the Colombian Peso as well as the US dollar provided by the Bank of Canada. We collect this data by manually downloading historical daily exchange rate data for various international currencies from the Bank of Canada web portal. The underlying data for exchange rates is sourced from Refinitiv (formerly Thomson Reuters).

## B The Deadweight Loss of Markup Distortions

**Characterizing the Efficient Frontier:** We equivalently formulate the planning problem described in Section 5 as one where the planner selects after-tax prices, $\tilde{\mathbf{p}}$, and each country's share of global income $\alpha_i$, subject to the adding up constraint, $\sum_i \alpha_i = 1$. Notice that that choice of $\alpha_i$ determines the optimal schedule of lump-sum transfers. More formally, the planner solves

$$\max_{\tilde{\mathbf{p}},\boldsymbol{\alpha}} \sum_i \delta_i \ln V_i \left(\alpha_i Y, \tilde{\mathbf{p}}_i\right),$$

subject to equilibrium constraints including the global budget constraint whereby global income $Y$ satisfies

$$Y = \sum_i w_i L_i + \int_{\mathcal{M}} \sum_{n,i} \left[ \left( \tilde{p}_{ni}\left(\mu\right) - \frac{1}{\mu} p_{ni}\left(\mu\right) \right) q_{ni}\left(\mu\right) d\mu \right],$$

Table A2: Summary of multinational profit ownership: Orbis database

| Country | Income Group | Retained in the Origin Country | Repatriated to Foreign Shareholders | |
|---|---|---|---|---|
| | | | High-Income Shareholders | Low-Income Shareholders |
| Argentina | Low/Middle Income | 88.4% | 8.6% | 3.1% |
| Bulgaria | Low/Middle Income | 77.7% | 13.8% | 8.5% |
| Brazil | Low/Middle Income | 88.4% | 10.4% | 1.2% |
| China | Low/Middle Income | 91.3% | 8.1% | 0.6% |
| Colombia | Low/Middle Income | 80.0% | 11.9% | 8.0% |
| Costa Rica | Low/Middle Income | 86.3% | 13.7% | 0.0% |
| Hungary | Low/Middle Income | 73.3% | 21.2% | 5.6% |
| Indonesia | Low/Middle Income | 95.6% | 3.7% | 0.7% |
| India | Low/Middle Income | 56.4% | 38.3% | 5.3% |
| Kazakhstan | Low/Middle Income | 71.9% | 17.6% | 10.4% |
| Morocco | Low/Middle Income | 84.3% | 10.6% | 5.2% |
| Mexico | Low/Middle Income | 96.6% | 3.4% | 0.1% |
| Malaysia | Low/Middle Income | 92.0% | 7.1% | 0.9% |
| Peru | Low/Middle Income | 67.3% | 19.7% | 13.1% |
| Philippines | Low/Middle Income | 86.7% | 11.7% | 1.6% |
| Romania | Low/Middle Income | 67.9% | 31.4% | 0.7% |
| Rest of the World | Low/Middle Income | 97.4% | 2.2% | 0.4% |
| Russia | Low/Middle Income | 87.8% | 9.7% | 2.4% |
| Thailand | Low/Middle Income | 84.8% | 13.9% | 1.3% |
| Tunisia | Low/Middle Income | 91.2% | 8.2% | 0.6% |
| Turkey | Low/Middle Income | 84.1% | 13.5% | 2.4% |
| Viet Nam | Low/Middle Income | 89.7% | 8.0% | 2.4% |
| South Africa | Low/Middle Income | 87.0% | 12.6% | 0.5% |
| Austria | High Income | 58.4% | 29.5% | 12.1% |
| Australia | High Income | 68.7% | 29.9% | 1.4% |
| Belgium | High Income | 65.7% | 33.8% | 0.5% |
| Canada | High Income | 87.8% | 9.7% | 2.5% |
| Switzerland | High Income | 79.3% | 17.6% | 3.1% |
| Chile | High Income | 84.7% | 14.7% | 0.6% |
| Cyprus | High Income | 47.5% | 25.0% | 27.5% |
| Czech Republic | High Income | 66.5% | 32.7% | 0.8% |
| Germany | High Income | 77.3% | 18.4% | 4.4% |
| Denmark | High Income | 91.3% | 8.7% | 0.0% |
| Estonia | High Income | 66.5% | 30.7% | 2.8% |
| Spain | High Income | 65.6% | 28.7% | 5.8% |
| Finland | High Income | 81.7% | 17.2% | 1.1% |
| France | High Income | 84.1% | 14.8% | 1.1% |
| United Kingdom | High Income | 77.5% | 17.7% | 4.8% |
| Greece | High Income | 76.2% | 22.8% | 1.0% |
| Hong Kong (China) | High Income | 47.2% | 4.1% | 48.6% |
| Croatia | High Income | 84.0% | 13.3% | 2.8% |
| Ireland | High Income | 66.2% | 32.4% | 1.5% |
| Israel | High Income | 82.9% | 15.0% | 2.1% |
| Iceland | High Income | 83.6% | 16.1% | 0.3% |
| Italy | High Income | 81.2% | 17.7% | 1.1% |
| Japan | High Income | 92.9% | 7.0% | 0.1% |
| Korea | High Income | 94.6% | 4.0% | 1.4% |
| Lithuania | High Income | 72.1% | 26.7% | 1.2% |
| Luxembourg | High Income | 73.2% | 25.0% | 1.8% |
| Latvia | High Income | 66.2% | 31.3% | 2.5% |
| Malta | High Income | 54.5% | 28.4% | 17.1% |
| Netherlands | High Income | 68.1% | 28.3% | 3.7% |
| Norway | High Income | 73.0% | 26.7% | 0.3% |
| New Zealand | High Income | 86.9% | 12.8% | 0.4% |
| Poland | High Income | 79.0% | 20.7% | 0.3% |
| Portugal | High Income | 57.7% | 34.1% | 8.2% |
| Saudi Arabia | High Income | 93.7% | 1.1% | 5.2% |
| Sweden | High Income | 85.5% | 12.9% | 1.7% |
| Singapore | High Income | 64.1% | 21.0% | 14.8% |
| Slovenia | High Income | 59.7% | 39.5% | 0.7% |
| Slovak Republic | High Income | 59.2% | 32.3% | 8.5% |
| Chinese Taipei | High Income | 96.3% | 2.7% | 1.0% |
| United States | High Income | 96.5% | 2.6% | 0.8% |

*Note*: This table reports the share of profits rebated to shareholders in the domestic economy and repatriated to foreign shareholders. The data is from Orbis for the 2005-2015 period. We only report summary statistics for 62 main countries/regions which are also represented in the ICIO data, with the rest of the countries aggregated into the "Rest of the World."

where the summation on the right-hand side collects global income from profits and tax revenues. Noting that $\partial E_i / \partial Y = \alpha_i$, the first-order condition w.r.t. $\tilde{p}_{ni}(\mu) \in \tilde{\mathbf{p}}$ can be written as

$$\delta_i \frac{\partial \ln V_i(.)}{\partial \ln \tilde{p}_{ni}(\mu)} - \sum_{\ell} \left( \delta_\ell \alpha_\ell \frac{\partial \ln V_\ell(.)}{\partial E_\ell} \right) \tilde{p}_{ni}(\mu)\, q_{ni}(\mu)$$
$$+ \sum_{\ell} \left( \delta_\ell \alpha_\ell \frac{\partial \ln V_\ell}{\partial E_\ell} \right) \int_{\mathcal{M}} \sum_{j,\ell} \left[ \left( \tilde{p}_{j\ell}(\mu) - \frac{1}{\mu} p_{j\ell}(\mu) \right) q_{j\ell}(\mu) \frac{\mathrm{d} \ln q_{j\ell}(\mu)}{\mathrm{d} \ln \tilde{p}_{ni}(\mu)} d\mu \right]$$
$$+ \sum_{\ell} \left( \delta_\ell \alpha_\ell \frac{\partial \ln V_\ell}{\partial E_\ell} \right) \sum_j \left( \left[ w_n L_n - \int_{\mathcal{M}} \sum_{\ell} \frac{1}{\mu} \frac{\partial \ln p_{j\ell}(\mu)}{\partial \ln w_j} p_{j\ell}(\mu)\, q_{j\ell}(\mu)\, d\mu \right] \frac{\mathrm{d} \ln w_j}{\mathrm{d} \ln \tilde{p}_{ni}(\mu)} \right) = 0.$$

Per Roy's identity we can re-write the first term in first-order condition as

$$\text{[Roy's identity]} \qquad \frac{\partial \ln V_i(.)}{\partial \ln \tilde{p}_{ni}(\mu)} = -\frac{\partial \ln V_n(.)}{\partial E_n} \tilde{p}_{ni}(\mu)\, q_{ni}(\mu).$$

Also, per Shephard's lemma, $\partial \ln p_{i\ell}(\mu) / \partial \ln w_i = 1$, which considering the labor-market clearing condition, $w_i L_i - \int_{\mathcal{M}} \sum_{\ell} \frac{1}{\mu} p_{i\ell}(\mu)\, q_{i\ell}(\mu)\, d\mu = 0$, asserts that the last line in the first-order condition reduces to zero. Taking these point into account and noting that $\partial \ln V_n / \partial \ln E_n = 1$ (since preferences are homothetic) simplifies the first-order conditions as,

$$\frac{1}{Y} \left[ \frac{\delta_i}{\alpha_i} - 1 \right] \tilde{p}_{ni}(\mu)\, q_{ni}(\mu) + \frac{1}{Y} \int_{\mathcal{M}} \sum_{j,\ell} \left[ \left( \tilde{p}_{j\ell}(\mu) - \frac{1}{\mu} p_{j\ell}(\mu) \right) q_{j\ell}(\mu) \frac{\mathrm{d} \ln q_{j\ell}(\mu)}{\mathrm{d} \ln \tilde{p}_{ni}(\mu)} d\mu \right] = 0.$$

The trivial solution to the above equation requires marginal cost pricing for all varieties paired with lump-sum transfers ensure country $i$'s share from global income corresponds to its Pareto weight,

$$\alpha_i^* = \delta_i, \qquad \tilde{p}_{ni}^*(\mu) = \frac{1}{\mu} p_{ni}(\mu).$$

Note that the Pareto efficient frontier can be traced by varying the Pareto weights $\{\delta_i\}$. All points on the frontier exhibit marginal-cost-pricing but differ in the underlying transfers, as implicitly determined by $\alpha_i^*$.

**Distance to Efficient Frontier.** Following Section 5, the welfare change from implementing the efficient transfer is

$$\Delta \ln W_i \approx \Delta \ln E_i - \sum_n \int_{\mathcal{M}} \Delta \ln p_{ni}(\mu)\, \lambda_{ni}(\mu)\, e_n(\mu)\, d\mu,$$

where the $\Delta E_i = \Delta Y_i$, per the representative consumer's budget constraint, with $\Delta Y_i$ given by

$$\Delta \ln Y_i = \Delta \ln w_i + \Delta \ln \mathscr{T}_i + \ln \mathbb{E}_{y_i} \left[ \frac{1}{\mu} \right].$$

Moreover, since every point on the Pareto efficient frontier requires marginal-cost-pricing, the expenditure-weighted change in consumer prices for restoring marginal-cost-pricing is

$$\sum_n \int_{\mathcal{M}} \Delta \ln p_{ni}(\mu)\, \lambda_{ni}(\mu)\, e_n(\mu)\, d\mu = \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right] + \sum_n \lambda_{ni} \Delta \ln w_n.$$

To determine the pure loss from monopoly distortions, we are interested in a point on the Pareto efficient frontier where lump-sum transfers nullify the factoral terms of trade effects for restoring marginal cost pricing, i.e., $\Delta \ln \mathscr{T}_i = \sum_n \lambda_{ni} \Delta \ln w_n$ for all $i$. It is straightforward to verify that the noted transfer structure satisfies the adding up constraint, $\sum_i T_i = 0$ . Combining these intermediate results, we obtain the deadweight loss of monopoly distortions $\mathscr{D}_i = \Delta \ln W_i \mid_{\Delta \ln \mathscr{T}_i = \sum_n \lambda_{ni} \Delta \ln w_n}$ as follows:

$$\mathscr{D}_i \approx \ln \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right]$$

**Taylor Approximation.** In the closed economy case, $y_i(\mu) = e_i(\mu)$ for all $\mu$, which when plugged into our formula for $\mathscr{D}_i$ yields

$$\mathscr{D}_i^{closed} \approx \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] - \mathbb{E}_{e_i}\left[\ln \frac{1}{\mu}\right] \sim \text{MLD}_{e_i}\left(\frac{1}{\mu}\right).$$

We use Taylor's theorem to link $\mathscr{D}_i^{closed}$ to cross-industry markup dispersion. For a generic industry-level variable $x(\mu)$, The Taylor expansion of function $\mathbb{E}_{e_i}[\ln x] = \int_{\mathcal{M}} e_i(\mu) \ln x_i(\mu)\, d\mu$ around $x_0 \sim \mathbb{E}_{e_i}[x]$ can be expressed as

$$\mathbb{E}_{e_i}[\ln x] \approx \ln \mathbb{E}_{e_i}[x] + \left[\int_{\mathcal{M}} \frac{e_i(\mu)}{\mathbb{E}_{e_i}[x]} \left(x(\mu) - \mathbb{E}_{e_i}[x]\right) d\mu\right] + \frac{1}{2} \sum_k \left[\int_{\mathcal{M}} \frac{e_i(\mu)}{\mathbb{E}_{e_i}[x]^2} \left(x_k - \mathbb{E}_{e_i}[x]\right)^2 d\mu\right]$$

Note that, by definition, $\int_{\mathcal{M}} \frac{e_i(\mu)}{\mathbb{E}_{e_i}[x]} \left(x(\mu) - \mathbb{E}_{e_i}[x]\right) d\mu = \frac{1}{\mathbb{E}_{e_i}[x]} \left(\mathbb{E}_{e_i}[x] - \mathbb{E}_{e_i}[x]\right) = 0$, so the second term on the right-hand side collapses to zero. Rearranging the above equation, therefore, yields

$$\ln \mathbb{E}_{e_i}[x] - \mathbb{E}_{e_i}[\ln x] \approx \frac{1}{2} \mathbb{E}_{e_i}[x]^{-2}\, \mathbb{E}_{e_i}\left[\left(x - \mathbb{E}_{e_i}[x]\right)^2\right] = \frac{1}{2} \mathbb{E}_{e_i}[x]^{-2}\, \text{Var}_{e_i}(x),$$

Letting $x(\mu) = \frac{1}{\mu}$ into the above equation delivers $\mathscr{D}_i^{closed} \approx \text{Var}_{e_i}\left(\frac{1}{\mu}\right) \times \widetilde{\mathbb{E}}_{e_i}[\mu]^2$, were recall that $\widetilde{\mathbb{E}}_{e_i}[.]$ represented the harmonic mean operator.

# C Trade-Induced Change in the DWL of Markups (Proposition 2)

This appendix derives the exact formula for the trade-induced change in the deadweight loss of markups in our semi-parmateric model (Proposition 2). Section 6 the main text provides details on the parametric assumptions underlying the model.

**Characterizing the within-industry markup distribution.** We begin by characterizing the markup distribution associated with firm variaties supplied by each location, demonstrating the invariance of the distribution to the origin country and underlying trade costs. To economize on notation, we drop the subscript denoting the destination market to which firm variaties are supplied to. Firms' profit maximization implies the standard Lerner formula for the optimal markup, which depends on its competitiveness $\nu \equiv c/P_k$. In particular, the optimal markup for each variety implicitly solves

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

where $\varepsilon_k(x) \equiv \mid D_k'(x) \mid$. As noted by Arkolakis et al. (2019), is straightforward to check that $m_k\left(.\right)$ is a strictly increasing function provided that Marshall's Second Law of Demand is satisfied (i.e., $\varepsilon_k'(x) < 0$.. In addition to being monotone, the function $m_k\left(.\right)$ is independent of the origin country, and independent of the underlying vector of trade costs, $\boldsymbol{\tau} \equiv \left\{\tau_{ij,k}\right\}_{i,j,k}$. Let $\varphi_{in,k}^*$ denote the minimum productivity cut-off above which demand is non-zero, implying that for any firm variety with productivity $\varphi$, the competitiveness is given by $\nu = \varphi/\varphi_{in,k}^*$. The distribution of markups for goods sold from origin $i$ to destination $n$ in industry $k$ is, accordingly, given by

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

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

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

Since $\left(m_k^{-1}\left(\mu\right)\right)^{-\theta_k}$ is independent of the origin country and the underlying vector of trade costs, it follows immediately that the distribution of markups charged in destination $n$ is invariant to trade openness and the origin from which the goods are sources, i.e., $\mathcal{M}_{in,k}\left(\mu;\boldsymbol{\tau}\right) = \mathcal{M}_k\left(\mu\right)$. Leveraging this intermediate result, we now proceed to the derivation of $\Delta\mathscr{D}_i$.

**Deriving the exact formula for $\Delta\mathscr{D}$.** Notice that the gains from trade under the decentralized and efficient allocations are given by

$$GT_i = W_i - W_i^{(closed)}, \qquad GT_i^* = W_i^* - W_i^{*(closed)},$$

from which we can obtain the following representation for the trade-induced change in the deadweight loss (DWL) of markup distortions:

$$\Delta\mathscr{D}_i = \left(W_i^* - W_i\right) - \left(W_i^{*(closed)} - W_i^{(closed)}\right) = GT_i^* - GT_i.$$

One can immediately verify that the parametric model with efficient pricing satisfies restrictions R1-R3 in Arkolakis et al. (2012). Hence, The gains from trade under efficient-pricing are given by the ACR formula, $GT_i^* = \prod_k \lambda_{ii,k}^{-e_{i,k}}$. The gains from trade in the markup-distorted economy ($GT_i$) can be characterized by noting that the local welfare change due to an infinitesimal change in trade costs can be specified as

$$\mathrm{d}\ln W_i = \mathrm{d}\ln Y_i - \sum_k \sum_n \int_\omega e_{i,k}\lambda_{ni,k}(\omega)\,\mathrm{d}\ln p_{ni,k}(\omega)$$

Following Section 5, the change in nominal income consists if the change in the wage bill plus profit payments. Namely,

$$\mathrm{d}\ln Y_i = \mathrm{d}\ln\left(w_i L_i\right) - \mathrm{d}\ln\sum_k\left(y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}\left[\mu\right]^{-1}\right),$$

where $\widetilde{\mathbb{E}}_{\rho_k}\left[\mu\right] \sim \widetilde{\mathbb{E}}_{\lambda_k}\left[\mu\right]$ is invariant to trade costs. The welfare effects that channel through changes to consumer prices can be expanded as

$$\begin{aligned}\sum_n \int_\omega \lambda_{ni,k}(\omega)\,\mathrm{d}\ln p_{ni,k}(\omega)\,d\omega &= \sum_n \int_{\varphi_{ni,k}^*}^\infty \lambda_{ni,k}(\varphi)\left[\mathrm{d}\ln\mu_{ni,k}(\varphi) + \mathrm{d}\ln c_{ni,k}\right]dG_{n,k}(\varphi)\\ &= \sum_n \lambda_{ni,k}\left[\mathrm{d}\ln c_{ni,k} + \int_{\varphi_{ni,k}^*}^\infty \frac{\lambda_{ni,k}(\varphi)}{\lambda_{ni,k}}\mathrm{d}\ln\mu_{ni,k}(\varphi)\,dG_{n,k}(\varphi)\right]\end{aligned}$$

where given that $\mu_{ni,k}(\omega) = m_k\left(\varphi/\varphi_{ni,k}^*\right)$, then we can can specify $\mathrm{d}\ln\mu_{ni,k}(\varphi) = \frac{d\ln m_k\left(\varphi/\varphi_{ni,k}^*\right)}{d\ln\varphi_{ni,k}^*}\mathrm{d}\ln\varphi_{ni,k}^*$. Then, defining

$$\rho_{ni,k} \equiv \int_{\varphi_{ni,k}^*}^\infty \frac{\lambda_{ni,k}(\varphi)}{\lambda_{ni,k}}\frac{d\ln m_k\left(\varphi/\varphi_{ni,k}^*\right)}{d\ln\varphi_{ni,k}^*}dG_{n,k}(\varphi),$$

we can write the price effects as

$$\sum_n \int_\omega \lambda_{ni,k}(\omega)\,\mathrm{d}\ln p_{ni,k}(\omega)\,d\omega = \sum_n \lambda_{ni,k}\left[\mathrm{d}\ln c_{ni,k} + \rho_{ni,k}\mathrm{d}\ln\varphi_{ni,k}^*\right].$$

Following Arkolakis et al. (2019), we can show that the markup elasticity is invariant to trade costs and common for all origin-destination dyads. In particular,

$$\rho_{ni,k} = \rho_k = \int_1^\infty \frac{d\ln m_k(\nu)}{d\ln\nu}\frac{\left(m_k(\nu)/\nu\right)D_k\left(m_k(\nu)/\nu\right)\nu^{-\theta_k-1}}{\int_1^\infty\left(m_k(\nu')/\nu'\right)D_k\left(m_k(\nu')/\nu'\right)\nu'^{-\theta_k-1}d\nu'}d\nu.$$

Note that by definition, $\mathrm{d}\ln\varphi^*_{ni,k} = \mathrm{d}\ln c_{ni,k} - \mathrm{d}\ln P_{i,k}$. And when preferences are homothetic, $\mathrm{d}\ln P_{i,k} = \sum_n [\lambda_{ni,k}\mathrm{d}\ln c_{ni,k}]$. Consolidating these two points and invoking the uniformity and invariance of $\rho_k$ we obtain

$$\sum_n \lambda_{ni,k}\rho_{ni,k}\mathrm{d}\ln\varphi^*_{ni,k} = \rho_k\left(\sum_n [\lambda_{ni,k}\mathrm{d}\ln c_{ni,k}] - \mathrm{d}\ln P_{i,k}\right) = 0.$$

Leveraging the constant elasticity aggregate import demands system, $\lambda_{ni,k}/\lambda_{ii,k} = (c_{ni,k}/c_{ii,k})^{-\theta_k}$, one can show in the spirit of ACR that

$$\mathrm{d}\ln c_{ni,k} - \mathrm{d}\ln c_{ii,k} = -\frac{1}{\theta_k}\left(\mathrm{d}\ln\lambda_{ni,k} - \mathrm{d}\ln\lambda_{ii,k}\right),$$

where $\mathrm{d}\ln c_{ii,k} = \mathrm{d}\ln(\tau_{ii,k}w_i) = 0$ by choice of numeriare. Using the expression for $\mathrm{d}\ln c_{ni,k}$ from the above equation, yields

$$\sum_n \lambda_{ni,k}\mathrm{d}\ln c_{ni,k} = -\frac{1}{\theta_k}\underbrace{\sum_n(\lambda_{ni,k}\mathrm{d}\ln\lambda_{ni,k})}_{=0} + \frac{1}{\theta_k}\mathrm{d}\ln\lambda_{ii,k}.$$

Plugging the above expressions back into our initial expression for $\mathrm{d}\ln W_i$, delivers the following simplified expression

$$\mathrm{d}\ln W_i = \mathrm{d}\ln\sum_k\left(y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}\right) - \underbrace{\sum_k \frac{e_{i,k}}{\theta_k}\mathrm{d}\ln\lambda_{ii,k}}_{\mathrm{d}\ln W_i^*},$$

where $\mathrm{d}\ln W_i^* = \sum_k \frac{e_{i,k}}{\theta_k}\mathrm{d}\ln\lambda_{ii,k}$ follows for the ACR formula regarding the gains from trade in efficient economies with constant elasticity aggregate demand systems. The gains from trade can be obtained by performing an integration on the above equation, starting from the actual trade costs until trade costs approach infinity under autarky ($\tau \to \infty$). Doing so yields

$$GT_i = \ln\left(\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right) + \underbrace{\sum_k \frac{e_{i,k}}{\theta_k}\ln\lambda_{ii,k}}_{GT_i^*}.$$

Appealing to the above equation the impact of trade on the DWL of monopoly distortions can be recovered as $\Delta\mathscr{D}_i = GT_i^* - GT_i$, which delivers the expression under Proposition 2. Namely,

$$\Delta\mathscr{D}_i = \ln\left(\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right).$$

## C.1 Deriving the Approximate Formula for $\Delta\mathscr{D}_i$

Applying Taylor's Theorem to $f(\mathbf{y}) = \ln \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right] = \ln \int_{\mathcal{M}} \frac{1}{\mu} y_i(\mu)\, d\mu$, we can derive the following approximation around $y_i(\mu) = e_i(\mu)$, which corresponds to a small deviation from autarky,

$$\ln \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right] \approx \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] + \int_{\mathcal{M}} \left(\frac{1}{\mu}\left[y_i(\mu) - e_i(\mu)\right] / \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]\right) d\mu.$$

Noting that $\widetilde{\mathbb{E}}_{e_i}[\mu] = 1/\mathbb{E}_{e_i}\left[\frac{1}{\mu}\right]$, we can invoke our notation for covariance to rewrite the above equations as

$$\ln \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right] - \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu}\right] \approx \widetilde{\mathbb{E}}_{e_i}[\mu] \times \mathrm{Cov}\left(\frac{1}{\mu}, \frac{y_i(\mu)}{e_i(\mu)}\right).$$

## C.2 Accounting for Quasi-Rents

Suppose a fraction $\delta_i(\mu)$ of the markup in country $i$ generates quasi-rents that cover by fixed cost payments to primary production factors. In this case, the nominal income in country $i$ is given by:

$$Y_i = \mathbb{E}_{y_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right]^{-1} w_i L_i$$

where the fixed cost payments are now included in the aggregate wage bill $w_i L_i$, and the wage income multiplier is adjusted downwards to account for the dissipation of quasi-rents. As we will demonstrate shortly, wedges are efficient to the extent that they generate quasi-rents. Therefore, the prices that obtain the efficient allocation can be represented as:

$$p^*_{in}(\mu) = \left(\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right) p_{in}(\mu).$$

Absent quasi-rents ($\delta_i = 0$), the efficient price corresponds to marginal cost pricing. When markup wedges only generate quasi-rents, the efficient and decentralized prices are exactly the same. Extrapolating from our baseline derivation, it immediately follows that:

$$\mathscr{D}_i \approx \underbrace{\ln \mathbb{E}_{e_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right] - \mathbb{E}_{e_i}\left[\ln\left(\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right)\right]}_{\mathrm{MLD}_{e_i}\left(\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right)} + \ln\left(\frac{\mathbb{E}_{y_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right]}{\mathbb{E}_{e_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right]}\right).$$

Notice that the above expression includes our baseline formula as a special case where $\delta_i = 0$. Accordingly, the pure impact of trade on the DWL of monopoly distortions becomes:

$$\Delta\mathscr{D}_i = \Delta\mathrm{MLD}_{e_i}\left(\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right) + \ln\left(\frac{\mathbb{E}_{y_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right]}{\mathbb{E}_{e_i}\left[\frac{1-\delta_i(\mu)}{\mu} + \delta_i(\mu)\right]}\right).$$

It is important to note that accounting for quasi-rents does not necessarily reduce the DWL from markup wedges. For example, suppose markups are nearly uniform across different categories of goods, but $\delta_i(\mu)$ exhibits significant heterogeneity. In this case, $\delta_i(\mu)$ contributes to the dispersion in non-quasi-rent-generating markups, thereby amplifying the DWL of markup distortions compared to when quasi-rents are not accounted for.

**The Constrained-Efficiency of Quasi-Rent-Generating Wedges.** Below, we prove that when the rents associated with wedges leave the economy, the decentralized economy is constrained-efficient. In other words, 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 The Deadweight Loss of Markups under 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

$$P_{in}(\mu) = \mu\tau_{in}(\mu)\, w_i M_i(\mu)^{1-\mu},$$

where $M_i(\mu)$ denotes the mass of firms supplying markup $\mu$ from country $i$. Let $f_i^e(\mu)$ denote the constant unit labor cost of entry into markup segment $\mu$ in country $i$, and $L_i(\mu)$ denote the number of workers employed by firms producing the product with markup $\mu$, either for entry and production purposes. The number of firms per markup per good is determined by free entry condition, which equates total profits to the total entry cost payments. Namely:

$$\Pi_i(\mu) = \frac{\mu-1}{\mu} w_i L_i(\mu) = M_i(\mu)\, w_i f_i^e(\mu)$$

Following Lashkaripour and Lugovskyy (2023), the efficient allocation under free entry is implementable if the social planner implements a good-specific subsidy that equals the inverses markup, i.e., $\tau_i^*(\mu) = 1/\mu$. Our goal is to characterize the welfare gains from implementing efficient subsidies in closed economy; in particular, $\mathscr{D}_i^{closed} = \ln \hat{Y}_i^a - \ln \hat{P}_i^a$, where the superscript $a$ denotes autarky variables. After the implementation of tax, $\tau_i^*(\mu)$, total income is

$$\begin{aligned} Y_i^* &= w_i^* L_i + T_i^* = w_i^* L_i + \sum_{\mu\in\mathcal{M}} \left[\left(1 - \frac{1}{\tau_i^*(\mu)}\right) Y_i^*(\mu)\right] \\ &= w_i^* L_i + \sum_{\mu\in\mathcal{M}} \left[(1-\mu)\, y_i^*(\mu)\right] Y_i^* = w_i^* L_i + \left(1 - \mathbb{E}_{y_i^*}[\mu]\right) Y_i^* \end{aligned}$$

Note that in a closed economy operating under autarky, $y_i^{*a}(\mu) = e_i(\mu)$, which based on the above equation implies $Y_i^{a*} = w_i^{*a} L_i / \mathbb{E}_{e_i}[\mu_k]$. Moreover, $Y_i^a = w_i^a L_i$ and $w_i^a = w_i^{*a} = 1$ based on the choice of numeraire. Capitalizing on these points and rearranging the above equation yields

$$Y_i^{a*} = \frac{w_i^a L_i}{\mathbb{E}_{e_i}[\mu]}, \qquad \hat{Y}_i^a = \frac{1}{\mathbb{E}_{e_i}[\mu]}$$

Considering that $\hat{w}_i^a = 1$ by choice of numeraie, the change in the good-specific price index under autarky is $\hat{P}_i^a = \hat{\tau}_i(\mu)\, \hat{M}_i^a(\mu)^{\mu-1}$. We can calculate $\hat{M}_i^a(\mu)$ using the free entry condition, whereby $\hat{M}_i^a(\mu) = \hat{\Pi}_i^a(\mu)$. Under the efficient policy, this condition can be stated as

$$\Pi_i^{*a}(\mu) = \left(1 - \frac{1}{\mu}\right) e_i(\mu)\, Y_i^{*a} / \tau^*(\mu) = (\mu - 1)\, e_i(\mu)\, Y_i^{*a}.$$

The profits in the decentralized equilibrium are, meanwhile, given by $\Pi_i^a(\mu) = \left(1 - \frac{1}{\mu}\right) e_i(\mu) Y_i^a$, which yields, $\hat{\Pi}_i^a(\mu) = \mu \hat{Y}_i^a$. This in turn implies that $\hat{M}_i^a(\mu) = \hat{\Pi}_i^a(\mu) = \mu \hat{Y}_i^a$. Appealing to this expression, we can write the change in the consumer prices index as

$$\hat{P}_i^a = \sum_{\mu \in \mathcal{M}} e_i(\mu) \ln \hat{P}_i^a(\mu) = \sum_{\mu \in \mathcal{M}} \left[ e_i(\mu) \ln \left( \frac{1}{\mu} \hat{M}_i^a(\mu)^{1-\mu} \right) \right]$$
$$= \sum_{\mu \in \mathcal{M}} \left[ e_i(\mu) \ln \left( \frac{1}{\mu} \left(\mu \hat{Y}_i^a\right)^{1-\mu} \right) \right] = \sum_{\mu \in \mathcal{M}} \left[ e_i(\mu) \ln \left( \mu^{-\mu} \mathbb{E}_{e_i}[\mu]^{1-\mu} \right) \right].$$

Plugging the expressions for $\hat{Y}_i^a$ and $\hat{P}_i^a$ into $\mathscr{D}_i^{closed} = \ln\left(\hat{Y}_i^a / \hat{P}_i^a\right)$ yields

$$\mathscr{D}_i^{closed} = \mathbb{E}_{e_i}[\mu \ln \mu] - \mathbb{E}_{e_i}[\mu] \ln \mathbb{E}_{e_i}[\mu].$$

To assess the impact of trade on the DWL of distortions, we can compare the gains from trade under both the decentralized and efficient allocations. This is possible due to the design of the study. It can be easily verified that the gains from trade, starting from an initial allocation $\{y_i(\mu), e_i(\mu), \lambda_{ii}(\mu)\}_\mu$, are given by the following equation:

$$\Delta \ln W_i = \sum_{\mu \in \mathcal{M}} \left[ -\frac{e_i(\mu)}{\epsilon(\mu)} \ln \lambda_{ii}(\mu) + e_i(\mu)(\mu - 1) \ln \left( \frac{y_i(\mu)}{e_i(\mu)} \right) \right],$$

where $\Delta \ln W_i \equiv \ln W_i - \ln W_i^a$. In this equation, $e_i(\mu)$ remains unchanged by trade, based on the assumption that the utility aggregator across markup categories or industries has a Cobb-Douglas specification. The efficient allocation of interest, recall, corresponds to a point on the efficient frontier (denoted by $*$) where wages align with their factual values, implying that $\lambda_{ii}(\mu) \approx \lambda_{ii}^*(\mu)$. With this in mind, we can calculate $\Delta \mathscr{D}_i = \Delta W_i^* - \Delta W_i$ as follows:

$$\Delta \mathscr{D}_i \approx \sum_{\mu \in \mathcal{M}} e_i(\mu)(\mu - 1) \ln \left( \frac{y_i^*(\mu)}{y_i(\mu)} \right) = \mathbb{E}_{e_i}\left[ (\mu - 1) \frac{y_i^*(\mu)}{y_i(\mu)} \right].$$

As shown in Figure A6, $\Delta \mathscr{D}_i$ under free entry has similar properties to the restricted entry case emphasized in our baseline model. Specifically, international exposure to entry distortions has an international zero-sum structure, which is comparable to international rent-shifting effects under restricted entry.

# E The Deadweight Loss of Markups under Multinational Ownership

Let $\pi_{ni}$ represent the share of country $n$'s profits repatriated to households in country $i$. Given that country $n$'s aggregate profits are $\Pi_n = \left(\widetilde{\mathbb{E}}_{y_n}[\mu] - 1\right) w_n L_n$, the income of the representative consumer in country $i$ can be expressed as the sum of wage income and both domestic and international profit payments:

$$E_i = w_i L_i + \sum_{n=1}^{N} \left[ \pi_{ni} \left( \widetilde{\mathbb{E}}_{y_n} [\mu] - 1 \right) w_n L_n \right],$$

In this expression,$\mathbb{E}_{y_n} [\overline{\mu}]$ denotes the sales-weighted average markup charged by firms operating in country $n$ from the the lens of our semi-parmateric model. More specifically, considering that our semi-parametric has the same aggregate representation as a a model with a constant industry-wide markup, $\widetilde{\mathbb{E}}_{\rho_k} [\mu]$, we get

$$\widetilde{\mathbb{E}}_{y_n} [\mu] = \mathbb{E}_{y_n} \left[ \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right]^{-1} = \sum_k \left( y_{i,k} \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right)^{-1}$$

where $\widetilde{\mathbb{E}}_{\rho_k} [\mu]$ denotes the sales-weighted average markup in industry $k$, which is common across countries in our semi-parametric model. The change in country $i$'s expendable income after markup correction is, accordingly,

$$\hat{E}_i = \frac{w_i L_i}{w_i L_i + \sum_n \pi_{ni} \left( \widetilde{\mathbb{E}}_{y_n} [\mu] - 1 \right) w_n L_n} = \frac{1}{1 + \sum_n \pi_{ni} \left( \widetilde{\mathbb{E}}_{y_n} [\mu] - 1 \right) \frac{w_n L_n}{w_i L_i}}$$

Noting that country $i$'s output-side income or GDP is $Y_i = \widetilde{\mathbb{E}}_{y_i} [\mu] w_i L_i$, we can rewrite the above expression as

$$\hat{E}_i = \frac{1}{1 + \sum_n \pi_{ni} \left( \widetilde{\mathbb{E}}_{y_n} [\mu] - 1 \right) \frac{Y_n / \widetilde{\mathbb{E}}_{y_n} [\mu]}{Y_i / \widetilde{\mathbb{E}}_{y_i} [\mu]}} = \frac{1}{1 + \widetilde{\mathbb{E}}_{y_i} [\mu] \sum_n \pi_{ni} \left( 1 - \widetilde{\mathbb{E}}_{y_n} [\mu]^{-1} \right) \frac{Y_n}{Y_i}}.$$

We can unpack and rewrite the above expression as follows:

$$\hat{E}_i = \frac{1}{\widetilde{\mathbb{E}}_{y_i} [\mu]} \times \frac{1}{1 - (1 - \pi_{ii}) \left( 1 - \widetilde{\mathbb{E}}_{y_i} [\mu]^{-1} \right) + \sum_{n \neq i} \pi_{ni} \left( 1 - \widetilde{\mathbb{E}}_{y_n} [\mu]^{-1} \right) \frac{Y_n}{Y_i}},$$

Since by definition, $1 - \pi_{ii} = \sum_{n \neq i} \pi_{in}$, we can rearrange and rewrite the above expression as follows:

$$\ln \hat{E}_i = -\ln \widetilde{\mathbb{E}}_{y_i} [\mu] - \ln \left( 1 + \sum_{n \neq i} \left[ \pi_{ni} \frac{Y_n}{Y_i} \left( 1 - \sum_k y_{n,k} \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right) - \pi_{in} \left( 1 - \sum y_{i,k} \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right) \right] \right).$$

From here, we can extrapolate from our baseline derivation to obtain the following formula for the trade-led change in the DWL of monopoly distortions:

$$\Delta \mathscr{D}_i = \ln \left( \frac{\sum_k y_{i,k} \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1}}{\sum_k e_{i,k} \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1}} \right) + \ln \left( 1 + \sum_{n \neq i} \left[ \pi_{ni} \frac{Y_n}{Y_i} \sum_k y_{n,k} \left( 1 - \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right) - \pi_{in} \sum_k y_{i,k} \left( 1 - \widetilde{\mathbb{E}}_{\rho_k} [\mu]^{-1} \right) \right] \right).$$

## F The Deadweight Loss of Markups under CES Preferences across Industries

Our baseline semi-parmateric model assumed that the utility aggregator across industries is Cobb-Douglas. Here, we relax this assumption and characterize the deadweight loss (DWL) of markup distortion under a more flexible CES demand aggregator across industries. Considering the isomorphism between our semi-parametric baseline model and a multi-industry model with a constant industry-wide markup, we henceforth focus on the latter for a clearer exposition. More specifically, there are several categories of goods or industries indexed by $k = 1, ..., K$. Each category is characterized by a constant markup. So, we can alternatively index goods based on their sales-weighted average markup $\mu \in \mathcal{M} = \{\mu_1, ..., \mu_K\}$.With this choice of notation in mind, we now specify the the demand and supply side of the economy. Suppose preferences across industries have a CES rather than Cobb-Douglas parameterization. Namely,

$$U_i = \left[\sum_{\preceq \in \mathcal{M}} b_i(\mu)^{\frac{1}{\eta}} Q_i(\mu)^{\frac{\eta-1}{\eta}}\right]^{\frac{1-\eta}{\eta}}, \qquad \text{where} \qquad Q_i(\mu) = \left(\sum_n b_{ni}(\mu)^{\frac{1}{\sigma(\mu)}} \tilde{q}_{ni}(\mu)^{\frac{\sigma(\mu)-1}{\sigma(\mu)}}\right)^{\frac{\sigma(\mu)}{\sigma(\mu)-1}}.$$

Under this formulation, $\eta \geq 1$ denotes the elasticity of substitution across industries, with the special case $\eta = 1$ coinciding with the baseline Cobb-Douglas specification. In the CES model, markup-specific expenditure shares are endogenous and respond to trade openness or corrective policies. Accordingly, $e_i(\mu)$ throughout this appendix denotes the *endogenous* expenditure share on goods with markup $\mu$. Despite this added layer of richness, we can still infer the autarky DWL of markups for economy $i$ from observable shares, markups, and substitution elasticities. The following lemma presents this result with a formal proof provided in the following subsection.

**Lemma 1.** *Suppose preferences across goods or industries with average markups are CES with substitution elasticity, $\eta$. The deadweight loss of markups for country i (under autarky) can be inferred from good-specific average markups, expenditure shares, and substitution elasticities, $\boldsymbol{X} = \{\lambda_{ii}(\mu), e_i(\mu), \mu, \sigma(\mu), \eta\}_{i,\mu}$, as*

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

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

Evaluating the DWL of markups under CES preferences requires three additional statistics, domestic expenditure shares, $\{\lambda_{ii}(\mu)\}_{i,\mu}$, substitution elasticities, $\sigma(\mu)$, and the cross-good substitutability parameter, $\eta$. 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 1 to derive the following approximation for the autarky DWL of markups:

$$\mathscr{D}_i^{closed} \approx \frac{\eta}{2} \times \left[\text{CV}\left(\frac{1}{\mu}\lambda_{ii}\left(\mu\right)^{\frac{1-\eta}{1-\sigma(\mu)}}\right)\right]^2.$$

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

**Proposition 4.** *Suppose preferences across goods or industries with average markups are CES with substitution elasticity, $\eta$. The trade-induced change in the deadweight loss of markups, $\Delta\mathscr{D}_i$, can be inferred from good-specific average markups, expenditure shares, and substitution elasticities, $\boldsymbol{X} = \{\lambda_{ii}(\mu), e_i(\mu), \mu, \sigma(\mu), \eta\}_{i,\mu}$, as*

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

The CES-compatible expression for $\Delta\mathscr{D}_i$ exhibits an additional term that accounts for the impact of trade on *markup dispersion*. Specifically, as elaborated under Equation ??, the DWL of markups 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 good-specific expenditure shares are constant. Under CES preferences, however, good-specific expenditure shares react to trade, which translates into a change in the *expenditure-weighted* markup dispersion.

## F.1 Proof of Lemma 1

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-good utility aggregator. Next, suppose country $i$ was operating under autarky. Extrapolating from Appendix B and treating $w_i$ as the numeraire, the change in income after restoring marginal cost-pricing is

$$\hat{Y}_i^a = \frac{\Pi_i^a + w_i^a L_i}{w_i^a L_i} = \sum_{\mu \in \mathcal{M}}\left[\frac{1}{\mu}e_i^a\left(\mu\right)\right] = \mathbb{E}_{e_i^a}\left[\frac{1}{\mu}\right]$$

where $e_i^a$ corresponds to the autarky expenditure share on markup $\mu$ goods in country $i$. Notice, the autarky expenditure share is strictly different from the factual expenditure share under CES preferences, i.e., $e_i^a(\mu) \neq e_i(\mu)$. 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 good-specific expenditure shares if we shut down trade is

$$\hat{e}_i(\mu) \equiv \frac{e_i^a(\mu)}{e_i(\mu)} = \frac{e_i(\mu)\,\hat{P}_i(\mu)^{1-\eta}}{\sum_{\preceq'} e_i(\mu')\,\hat{P}_i(\mu')^{1-\eta}},$$

where $\hat{P}_i(\mu) = P_i^a(\mu)/P_i(\mu)$ is the change in markup $\mu$'s price index after shutting down trade. Following Arkolakis et al. (2012), we know that $\hat{w}_i/\hat{P}_i(\mu) = \lambda_{ii}(\mu)^{\frac{1}{1-\sigma(\mu)}}$, where $\lambda_{ii}(\mu)$ is the domestic expenditure share on markup $\mu$ goods under the status quo. Rearranging the aforementioned expression delivers $\hat{P}_i(\mu) = \hat{w}_i\lambda_{ii}(\mu)^{\frac{1}{1-\sigma(\mu)}}$. Plugging the expression for $\hat{P}_i(\mu)$ into the equation describing $\hat{e}_i(\mu)$, yields

$$e_i^a(\mu) = \frac{\hat{w}_i\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\,e_i(\mu)}{\sum_{\preceq'}\hat{w}_i\lambda_{ii}(\mu')^{\frac{1-\eta}{1-\sigma(\mu')}}\,e_i(\mu')} = \frac{\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\,e_i(\mu)}{\sum_{\preceq'}\lambda_{ii}(\mu')^{\frac{1-\eta}{1-\sigma(\mu')}}\,e_i(\mu')} = \frac{\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}}{\mathbb{E}_{e_i}\left[\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\right]}e_i(\mu).$$

Stated verbally, we can infer the counterfactual autarky expenditure share on markup $\mu$ goods from factual expenditure shares $e_i(\mu)$ and $\lambda_{ii}(\mu)$, and substitution elasticities, $\sigma(\mu)$ and $\eta$. Plugging the above expression for $e_i^a(\mu)$ in the our original expression for $\hat{Y}_i^a$, delivers the following expression

$$\hat{Y}_i^a = \sum_{\mu\in\mathcal{M}}\left[\frac{\frac{1}{\mu}\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}}{\mathbb{E}_{e_i}\left[\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\right]}e_i(\mu)\right] = \mathbb{E}_{e_i}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\right]$$

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

$$\hat{P}_i^a = \left[\sum_{\mu\in\mathcal{M}}\left(\frac{1}{\mu}\right)^{1-\eta}e_i^a(\mu)\right]^{\frac{1}{1-\eta}} = \left[\sum_{\mu\in\mathcal{M}}\left(\frac{1}{\mu}\right)^{1-\eta}\frac{\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}}{\mathbb{E}_{e_i}\left[\lambda_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\right]}e_i(\mu)\right]^{\frac{1}{1-\eta}} = \left(\mathbb{E}_{e_i}\left[\frac{1}{\mu}\tilde{\lambda}_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}\right]\right)^{\frac{1}{1-\eta}}.$$

Plugging the expressions for $\hat{Y}_i^a$ and $\hat{P}_i^a$ into $\mathscr{D}_i^{closed} = \ln\hat{Y}_i^a - \ln\hat{P}_i^A$, we obtain an updated expression for the autarky DWL of markups under CES preferences

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

**Deriving the Approximate Formula for $\mathscr{D}_i^{closed}$.** Define the function $f(.)$ as follows

$$f\left(x\left(\mu_1\right)...,x\left(\mu_K\right)\right)=\frac{1}{1-\eta}\ln\mathbb{E}_\omega\left[x^{1-\eta}\right]=\frac{1}{1-\eta}\ln\sum_{\mu\in\mathcal{M}}\left[\omega\left(\mu\right)x\left(\mu\right)^{1-\eta}\right].$$

Our goal is to derive the Taylor expression for $f(.)$ around $\bar{\mathbf{x}}=\left(\mathbb{E}_\omega\left[x\left(\mu\right)\right],...,\mathbb{E}_\omega\left[x\left(\mu\right)\right]\right)$. For this, we appeal to the following first- and second-order derivative of function $f(.)$ using the compact notation $x_k\sim x\left(\mu_k\right)$

$$\frac{\partial f}{\partial x_k}=\frac{\omega\left(\mu_k\right)x_k^{-\eta}}{\sum_{k'}\omega\left(\mu_{k'}\right)x_{k'}^{1-\eta}};\qquad\frac{\partial^2 f}{\partial x_k\partial x_g}=\frac{-\eta\,\omega\left(\mu_k\right)x_k^{-\eta-1}}{\sum_{k'}\omega\left(\mu_{k'}\right)x_{k'}^{1-\eta}}\times 1_{g=k}-\frac{\left(1-\eta\right)\omega\left(\mu_k\right)\omega\left(\mu_g\right)x_k^{-\eta-1}x_g^{-\eta}}{\left(\sum_{k'}\omega\left(\mu_{k'}\right)x_{k'}^{1-\eta}\right)^2}.$$

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 $f(.)=\frac{1}{1-\eta}\ln\mathbb{E}_\omega\left[x^{1-\eta}\right]$:

$$f(x_1...,x_K)\approx f(\bar{\mathbf{x}})+\sum_k\left[\frac{\partial f\left(\bar{\mathbf{x}}\right)}{\partial x_k}\left(x_k-\mathbb{E}_\omega\left[x\right]\right)\right]+\frac{1}{2}\sum_k\sum_g\left[\frac{\partial^2 f\left(\bar{\mathbf{x}}\right)}{\partial x_k\partial x_g}\left(x_k-\mathbb{E}_\omega\left[x\right]\right)\left(x_k-\mathbb{E}_\omega\left[x\right]\right)\right]$$

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

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

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

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

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

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

Noting that $\mathscr{D}_i^{closed}=\ln\mathbb{E}_{e_i^a}\left[\frac{1}{\mu}\right]-\frac{1}{1-\eta}\ln\mathbb{E}_{e_i^a}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]$, we immediately arrive at the following approximation for the autarky DWL of markups in economy $i$

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

where the last line follows from our previous observation that $e_i^a(\mu) = e_i(\mu)\,\tilde{\lambda}_{ii}(\mu)^{\frac{1-\eta}{1-\sigma(\mu)}}$ for all $\mu \in \mathcal{M}$.

## F.2 Proof of Proposition 4

To characterize impact of trade on the deadweight loss (DWL) of markups, we follow the same logic underlying the proof of Proposition 1. We first determine the DWL of markups 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 $*$ denote the globally efficient equilibrium, $\mathscr{D}_i = \ln \hat{Y}_i - \ln \hat{P}_i$, where $\hat{Y}_i = Y_i^*/Y_i$ and $\hat{P}_i = P_i^*/P_i$. The change in open economy $i$'s consumer price index after restoring marginal cost pricing is given by

$$\hat{P}_i = \left[\sum_{\mu\in\mathcal{M}} \left(\frac{1}{\mu}\right)^{1-\eta} e_i(\mu)\right]^{\frac{1}{1-\eta}} = \mathbb{E}_{e_i}\left[\left(\frac{1}{\mu}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}.$$

Note that above equation differs from $\hat{P}_i^a$ in that it depends on the factual good-specific expenditure shares, $e_i(\mu)$, rather than the counterfactual autarky expenditure shares. Likewise the change in income is

$$\hat{Y}_i = \sum_{\mu\in\mathcal{M}} \left[\frac{1}{\mu} y_i(\mu)\right] = \mathbb{E}_{y_i}\left[\frac{1}{\mu}\right],$$

where $y_i(\mu)$ denotes the goods-specific output 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 the DWL of markups as $\Delta\mathscr{D}_i = \ln\left(\frac{\hat{Y}_i}{\hat{Y}_i^a}\right) - \ln\left(\frac{\hat{P}_i}{\hat{P}_i^a}\right)$. Specifically, appealing to the previously-derived expressions for $\hat{P}_i^a$ and $\hat{Y}_i^a$, we get

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

Plugging the above two equations into $\Delta\mathscr{D}_i = \ln\left(\frac{\hat{Y}_i}{\hat{Y}_i^a}\right) - \ln\left(\frac{\hat{P}_i}{\hat{P}_i^a}\right)$, yields the formula presented under Proposition 4:

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

# G The Deadweight Loss of Markups under Input-Output Linkages

Now we consider an extension of our baseline model where production in economy $i$ employs labor and traded intermediate inputs. Considering the isomorphism between our semi-parametric baseline model and a multi-industry model with a constant industry-wide markup, we henceforth focus on the latter for a clearer exposition. More specifically, there are several categories of goods or industries indexed by $k = 1, ..., K$. Each category is characterized by a constant markup. So, we can alternatively index goods based on their markup $\mu \in \mathcal{M} = \{\mu_1, ..., \mu_K\}$.With this choice of notation in mind, we now specify the demand and supply side of the economy.

The demand side of the economy has the same specification as the baseline model introduced in Section 4. The supply side is richer and modeled in a similar fashion to Caliendo and Parro (2015). That is, production of markup $\mu$ goods in origin $i$ combines labor with internationally-sourced intermediate inputs. Let $\nu_i(\mu)$ denote the constant share of labor in production, or the value added share associated with markup level $\mu$. Assuming constant-returns to scale, $1 - \nu_i(\mu)$ represents the overall share of intermediate inputs in production. Goods with markup $\mu$ use intermediate inputs from various markup tiers, with $[1 - \nu_i(\mu)]\, \alpha_i(\mu', \mu)$ denoting the share of markup $\mu'$ inputs in the production of goods with markup $\mu$, with the adding up constraint, $\sum_g \alpha_i(\mu', \mu) = 1$. The composite bundle of inputs with markup $\mu'$ (namely, $I_i(\mu', \mu)$) is an Armington aggregator of inputs from various origin countries. In particular,

$$I_i(\mu', \mu) = \left( I_{1i}(\mu', \mu)^{\frac{\sigma(\mu)-1}{\sigma(\mu)}} + ... + I_{Ni}(\mu', \mu)^{\frac{\sigma(\mu)-1}{\sigma(\mu)}} \right)^{\frac{\sigma(\mu)}{\sigma(\mu)-1}},$$

where $I_{ji}(\mu', \mu)$ denotes the quantity of markup $\mu'$ inputs sourced from origin $j$. The above production structure assumes that the CES input aggregator has the same parameterization as the cross-national CES aggregator across consumer goods with markup $\mu'$. Cost minimization subject to this production structure yields the following price for composite variety $(i, j, \mu)$ representing origin $i$–destination $j$–markup $\mu$,

$$P_{ij}(\mu) = \mu \times \left( \frac{\tau_{ij}(\mu)}{\overline{\varphi}_i(\mu)} \right) w_i^{\nu_i(\mu)} \prod_{\mu' \in \mathcal{M}} P_i(\mu')^{[1-\nu_i(\mu)]\alpha_i(\mu', \mu)}, \tag{11}$$

where $P_i(\mu) = \left( \sum_j P_{ji}(\mu)^{1-\sigma(\mu)} \right)^{\frac{1}{1-\sigma(\mu)}}$ is a CES price index associated with internationally sourced inputs with markup $\mu$. Note that $P_i(\mu)$ also represents 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 markup $\mu$ 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 markup $\mu$ goods is given by

$$E_i(\mu) = e_i(\mu)(w_i L_i + \Pi_i) + [1 - \nu_i(\mu)] \sum_{\mu' \in \mathcal{M}} \left[ \alpha_i(\mu, \mu') C_i(\mu) \right], \tag{12}$$

where $\mathcal{C}_i(\mu)$ is the total input cost bill in origin $i$ for the production of markup $\mu$ goods, which includes payments to labor and intermediate inputs. By definition, the total input cost associated with producing markup $\mu$ goods in origin $i$ is equal to gross value of sales net of the underlying markup. Namely,

$$\mathcal{C}_i(\mu) = \sum_{\mu \in \mathcal{M}} \left[ \frac{1}{\mu} \lambda_{ij}(\mu) E_j(\mu) \right], \qquad \text{where} \qquad \lambda_{ij}(\mu) = \frac{P_{ij}(\mu)^{1-\sigma(\mu)}}{\sum_n P_{nj}(\mu)^{1-\sigma(\mu)}}. \tag{13}$$

*General Equilibrium under IO Linkages.* For a given vector of parameters and exogenous variables, $\{\mu, \sigma(\mu), L_i, e_i(\mu), \tau_{ij}(\mu), \overline{\varphi}_i(\mu), \nu_i(\mu), \alpha_i(\mu', \mu)\}_{i,\mu,\mu'}$, equilibrium is a vector of wages, aggregate rents, price indexes per markup tier, gross expenditure, and input cost levels, $\{w_i, \Pi_i, P_i(\mu), E_i(\mu), \mathcal{C}_i(\mu)\}_{i,\mu}$, that satisfy Equations 11–13 as well as the market clearing conditions in each market $i$,

$$w_i \overline{L}_i = \sum_k \nu_i(\mu) \mathcal{C}_i(\mu); \qquad \Pi_i = \sum_{\mu \in \mathcal{M}} (\mu - 1) \mathcal{C}_i(\mu).$$

*Gross Expenditure Shares.* With IO linkages, the gross expenditure share on industry $\mu$ goods typically differs from the net (or final good) expenditure share, $e_i(\mu)$. Gross expenditure shares encapsulate both intermediate and final good expenditure. This difference plays a prominent role in our analysis, so we use

$$\tilde{e}_i(\mu) \equiv \frac{E_i(\mu)}{\sum_{\mu'} E_i(\mu')} \sim \text{gross expenditure share}$$

to denote the gross expenditure share on markup $\mu$, where the gross expenditure level, $E_i(\mu)$, is described by Equation 12. Generally speaking, $\tilde{e}_i(\mu)$ will be greater than the net expenditure share, $e_i(\mu)$, for goods industries but lower for downstream goods. Moreover, unlike the net expenditure share, the gross expenditure is not *invariant* to trade. That is, we cannot readily determine the counterfactual autarky share, $\tilde{e}_i^a(\mu)$, based on its factual value, $\tilde{e}_i(\mu)$. We can, nonetheless, infer country $i$'s *autarky* gross expenditure shares from constant net expenditure shares, $\mathbf{e}_i = [e_i(\mu)]_\mu$, and the markup-adjusted input-output matrix, $\mathbf{\Phi}_i \equiv \left[ \frac{1}{\mu} [1 - \nu_i(\mu)] \alpha_i(\mu, \mu') \right]_{\mu,\mu'}$. In particular,

$$\tilde{\mathbf{e}}_i^a = (\mathbf{I} - \mathbf{\Phi}_i (\mathbf{I} - \mathbf{e}_i \otimes \mathbf{1}))^{-1} \mathbf{e}_i. \tag{14}$$

where $\mathbf{I}$ is an $K \times K$ identity matrix and $\mathbf{1}$ is a column vector of ones (see Appendix G for derivation details). Considering this point, we hereafter treat $\tilde{e}_i^a(.)$ as an observable statistic.

*The Deadweight Loss of Markups in a Closed Economy.* As in the baseline model, the DWL of markups is measured as the the welfare distance between the factual equilibrium and the efficient marginal cost-pricing equilibrium, $\mathscr{D}_i = \ln W_i^* - \ln W_i$. It is well known that IO linkages typically amplify the cost associated with a given set of markup wedges, as the markup on one type of good distorts production for other goods using that good as an input in production. To account for these ripple effects in a closed economy, suppose markups are eliminated for all goods. Let $P_i^*(\mu)$ denote the efficient price index after eliminating markups and $\hat{P}_i(\mu) = P_i^*(\mu)/P_i(\mu)$ denote the resulting change in the price index. Normalizing $w_i$ to one by choice of numeraire, the change in the price index of goods with markup, $\mu$, is the product the associated markup reduction ($\frac{1}{\mu}$) and the compounded reduction in input markups. In particular,

$$\hat{P}_i(\mu) = \frac{1}{\mu} \times \prod_{\mu' \in \mathcal{M}} \hat{P}_i(\mu)^{[1-\nu_i(\mu)]\alpha_i(\mu',\mu)}.$$

We can invert the above system to obtain $\hat{P}_i(\mu) = \prod_{\mu'} \left(\frac{1}{\mu'}\right)^{a_i(\mu,\mu')}$, where $a_i(\mu,\mu')$ denotes the $(\mu,\mu')$ entry of economy $i$'s inverse Leontief matrix.[28] The change in the consumer price index, $\hat{P}_i = \prod_{\mu} \hat{P}_i(\mu)^{e_i(\mu)}$, can thus be written as a compounded reduction in good-specific markups:[29]

$$\ln \hat{P}_i = \sum_{\mu \in \mathcal{M}} \left[\beta_i(\mu) \ln \frac{1}{\mu}\right], \quad \text{where} \quad \beta_i(\mu) \equiv \sum_{\mu'} \left[e_i(\mu')\, a_i(\mu',\mu)\right] \tag{15}$$

Weight $\beta_i(\mu)$ can be interpreted as the compounded weight of goods with markup $\mu$ in the consumer price index (CPI)—it reflects how a reduction in the good-specific markup translates to a reduction in CPI with ripple effects. Accordingly, for a strictly downstream good, $\beta_i(\mu)$ simply equals $e_i(\mu)$ which is the Cobb-Douglas share of industry $k$ in the consumption basket. Using the above observation and extrapolating the logic outlined in Section 5, we can produce an IO-adjusted sufficient statistics formula for the DWL of markups in a closed economy.

**Lemma 2.** *The deadweight loss of monopolistic markups for closed economy $i$ under IO linkages can be inferred from markups and observable shares, $\boldsymbol{X} = \{\mu, \tilde{e}_i(\mu), \nu_i(\mu), \alpha_i(\mu',\mu)\}_{\mu,\mu'}$, as$\mathscr{D}$where $\tilde{e}_i^a(\mu)$ and $\beta_i(\mu)$ are respectively given by Equations 14 and 15.*

Let us connect the above lemma to our baseline result. Absent input output linkages, which corresponds to $\nu_i(\mu) = 1$ and $\beta_i(\mu) = e_i(\mu) = e_i^a(\mu)$ for all $\mu$, the IO-adjusted formula for $\mathscr{D}_i^{closed}$ collapses to the baseline formula presented in Section 4. Beyond this special case, the DWL of markups depends crucially on the economy-wide input-output table, $\mathbf{A}_i \equiv [[1-\nu_i(\mu)]\,\alpha_i(\mu,\mu')]_{\mu,\mu'}$, which is directly observable. Also worth highlighting is that the IO-adjusted formula for $\mathscr{D}_i^{closed}$ uses information on both gross and net expenditure shares, $\tilde{e}_i(\mu)$, and $e_i(\mu)$.

*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 DWL of markups. Derivation details are presented in Appendix G and follow the same logic as our baseline Propositions 1. The resulting formula is presented below.

**Proposition 5.** *Suppose production employs traded intermediate inputs. The trade-induced change in the deadweight loss of markups, $\Delta\mathscr{D}_i$, can be inferred from industry-level markup observable shares, $\mathbf{X} = \{\mu_k, \nu_{i,k}, y_{i,k}, \tilde{e}_{i,k}, \alpha_{i,gk}$*

[28] 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.

[29] It is easy to check that $\sum_{\mu} \beta_i(\mu) = 1$, so $\beta_i(\mu)$ satisfies the condition to serve as a weight in the mean operator, $\mathbb{E}[.]$.

*as*

$$\Delta\mathscr{D}_i(\mathbf{X}) = \ln\left(\frac{\mathbb{E}_{y_i}\left[\frac{\nu_i(\mu)}{\mu}\right]}{1-\mathbb{E}_{y_i}\left[\frac{1-\nu_i(\mu)}{\mu}\right]}\right) - \ln\left(\frac{\mathbb{E}_{\tilde{e}_i^a}\left[\frac{\nu_i(\mu)}{\mu}\right]}{1-\mathbb{E}_{\tilde{e}_i^a}\left[\frac{1-\nu_i(\mu)}{\mu}\right]}\right),$$

*where* $\tilde{e}_i^a(\mu)$ *is given, in closed form, by Equation 14, as a function of the observable shares in set* $\mathbf{X}$*.*

When interpreting the above proposition, observe that gross expenditure and output shares coincide under autarky: $\tilde{e}_i^a = y_i^a$. Accordingly, the last term on the right-hand side of the above formula can be interpreted as consisting of averages weighted by autarky revenue shares. The formula for $\Delta\mathscr{D}_i$, thus, contrasts the allocation of resources across low- and high-markup goods under trade and autarky—suggesting that the impact of trade on the DWL of markups still channels primarily through rent-shifting effects, which are adjusted for input-output linkages.

### G.1 Proof of Lemma 2

As before, let superscript $*$ denote the globally efficient equilibrium and $a$ denote counterfactual values under autarky. The DWL of markups for a closed economy can be calculated as $\mathscr{D}_i^{closed} = \ln W_i^{a*} - \ln W_i^a$, 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 DWL of markups for a closed economy $i$ can, accordingly, we specified as

$$\mathscr{D}_i^{closed} = \ln W_i^{a*} - \ln W_i^a = \ln\left(\widehat{\pi}_i^a \widehat{w}_i^a\right) - \ln\left(\widehat{P}_i^a\right). \tag{16}$$

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 Appendix G that

$$\ln \hat{P}_i = \underbrace{-\sum_{\mu\in\mathcal{M}}\left[\beta_i(\mu)\ln\mu\right]}_{\mathbb{E}_{\beta_i}\left[\ln\frac{1}{\mu}\right]}, \quad \text{where} \quad \beta_i(\mu) \equiv \sum_{\mu'\in\mathcal{M}}\left[e_i(\mu')\, a_i(\mu',\mu)\right]. \tag{17}$$

where $a_i(\mu',\mu)$ denotes the $(\mu',\mu)$ entry of economy $i$'s inverse Leontief matrix and $e_i(\mu)$ denotes the net or final good expenditure share. What remains to be characterized in Equation 16 is the change in the economy-wide profit margins, $\widehat{\pi}_i^a$. To this end, we first specify $\pi_i^a$ as a function industry-level markups and shares. Recalling that $C_i(\mu)$ denotes total payment to production inputs when producing a goods with markup $\mu$ in origin $i$, we can write aggregate profits in a (closed or open) economy $i$ as

$$\Pi_i = \sum_{\mu \in \mathcal{M}} \left[ (\mu - 1) \, \mathcal{C}_i (\mu) \right]$$

Given our assumed production structure, total input costs are related to wage payments as $\mathcal{C}_i (\mu) = w_i L_i (\mu) / \nu_i (\mu)$, where $\nu_i (\mu)$ denotes the value-added share for goods produced with markup $\mu$ in origin $i$. We can, thus, rewrite total profits as a function of labor shares, $\ell_i (\mu) = L_i (\mu) / L_i$, and value-added shares as

$$\Pi_i = \sum_{\mu \in \mathcal{M}} \left[ \frac{\mu - 1}{\nu_i (\mu)} \frac{w_i L_i (\mu)}{w_i L_i} \right] w_i L_i = \sum_{\mu \in \mathcal{M}} \left[ \frac{\mu - 1}{\nu_i (\mu)} \ell_i (\mu) \right] w_i L_i. \tag{18}$$

Next, we need to write the labor shares as a function of output shares, $y_{i,k} = \frac{\sum_n P_{in}(\mu) Q_{in}(\mu)}{\sum_{\mu'} \sum_n P_{in}(\mu') Q_{in}(\mu')}$. This step relies on the observations that revenue shares, by definition, are related to labor shares as

$$\frac{\ell_i (\mu)}{\ell_i (\mu')} = \frac{\frac{\nu_i(\mu)}{\mu}}{\frac{\nu_i(\mu')}{\mu'}} \frac{y_i (\mu)}{y_i (\mu')} \quad \Longrightarrow \quad \ell_i (\mu) = \frac{\frac{\nu_i(\mu)}{\mu} y_i (\mu)}{\sum_{\mu'} \frac{\nu_i(\mu')}{\mu'} y_i (\mu')}.$$

The second line invokes the accounting property that labor and revenue shares add up to one: $\sum_\mu \ell_i (\mu) = \sum_\mu y_i (\mu) = 1$. Plugging the above expression for $\ell_i (\mu)$ back into Equation 18 yields

$$\Pi_i = \frac{\sum_\mu \left[ (\mu - 1) \frac{y_i(\mu)}{\mu} \right]}{\sum_\mu \left[ \nu_i (\mu) \frac{y_i(\mu)}{\mu} \right]} w_i L_i \quad \Longrightarrow \quad \pi_i \equiv 1 + \frac{\Pi_i}{w_i L_i} = \frac{\sum_\mu \left[ (\mu + \nu_i (\mu) - 1) \frac{y_i(\mu)}{\mu} \right]}{\sum_\mu \left[ \nu_i (\mu) \frac{y_i(\mu)}{\mu} \right]} = \frac{1 - \sum_\mu \left[ \frac{1 - \nu_i(\mu)}{\mu} y_i (\mu) \right]}{\sum_\mu \left[ \frac{\nu_i(\mu)}{\mu} y_i (\mu) \right]}.$$

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., $y^a_{i,k} = \tilde{e}^a_{i,k}$. Second, profits margins are zero under the efficient equilibrium, i.e., $\Pi^{a*}_i = 0$. Considering this, the above expression for $\pi_i$ implies the following change in profit margins after restoring efficiency:

$$\widehat{\pi}^a_i = \frac{\pi^{a*}_i}{\pi^a_i} = \frac{\sum_\mu \left[ \frac{\nu_i(\mu)}{\mu} \tilde{e}^A_i (\mu) \right]}{1 - \sum_\mu \left[ \frac{1 - \nu_i(\mu)}{\mu} \tilde{e}^A_i (\mu) \right]} = \frac{\mathbb{E}_{\tilde{e}^a_i} \left[ \frac{\nu_i(\mu)}{\mu} \right]}{1 - \mathbb{E}_{\tilde{e}^a_i} \left[ \frac{1 - \nu_i(\mu)}{\mu} \right]}. \tag{19}$$

Plugging Equations 17 and 19 back into our initial expression for $\mathscr{D}^{closed}_i$ (Equation 16) yields

$$\mathscr{D}^{closed}_i = \ln \left( \frac{\mathbb{E}_{\tilde{e}^a_i} \left[ \frac{\nu_i(\mu)}{\mu} \right]}{1 - \mathbb{E}_{\tilde{e}^a_i} \left[ \frac{1 - \nu_i(\mu)}{\mu} \right]} \right) - \mathbb{E}_{\beta_i} \left[ \ln \frac{1}{\mu} \right].$$

## G.2 Proof of Proposition 5

It is straightforward to check that the logic of Proposition 2 extends to an economy with input-output linkages, at least to an approximation. That is, for a common input-output structure, restoring marginal cost pricing retains the wage vector, $\mathbf{w}$, and the terms-of-trade. Hence, letting superscript $*$ denote the globally efficient equilibrium, the DWL of markups can be calculated as $\mathscr{D}_i = \ln W_i^* - \ln W_i$. With this background, we wish to characterize $\Delta\mathscr{D}_i$, which is the change in DWL of markups 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

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

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

$$\ln\widehat{W}_i = \ln\left(\widehat{w}_i / \widehat{P}_i\right) + \ln\left(\widehat{\pi}_i\right). \tag{20}$$

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 Costinot and Rodríguez-Clare (2014), we can derive the following relationship between real production cost, $C_i(\mu)/P_i(\mu)$, and the domestic expenditure share, $\lambda_{ii}(\mu)$,

$$\lambda_{ii}(\mu) = \left(\frac{P_{ii}(\mu)}{P_i(\mu)}\right)^{1-\sigma(\mu)} = \left(\frac{\mu\tau_{ii}(\mu)\, C_i(\mu)}{P_i(\mu)}\right)^{1-\sigma(\mu)} \qquad \Longrightarrow \qquad \frac{\widehat{C}_i(\mu)}{\widehat{P}_i(\mu)} = \lambda_{ii}(\mu)^{\frac{1}{1-\sigma(\mu)}}.$$

Appealing to the expression for $C_i(\mu)$ we can derive an equation relating the real wage in each industry to real production cost. Namely,

$$\widehat{C}_i(\mu) = \widehat{w}_i^{\nu_i(\mu)} \prod_{\mu'} \widehat{P}_i(\mu')^{[1-\nu_i(\mu)]\alpha_i(\mu',\mu)} \qquad \Longrightarrow \qquad = \ln\frac{\hat{w}_i}{\hat{P}_i(\mu)} = \ln\frac{\widehat{C}_i(\mu)}{\hat{P}_i(\mu)} + \sum_{\mu'}\left[(1-\nu_i(\mu))\,\alpha_i(\mu',\mu)\ln\frac{\hat{w}_i}{\hat{P}_i(\mu)}\right]$$

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

$$\frac{\hat{w}_i}{\widehat{P}_i(\mu)} = \prod_{\mu'}\left(\frac{\widehat{C}_i(\mu')}{\hat{P}_i(\mu')}\right)^{a_{i,kg}} = \prod_{\mu'}\left(\lambda_{ii}(\mu)^{\frac{1}{1-\sigma(\mu')}}\right)^{a_i(\mu,\mu')} \qquad \Longrightarrow \qquad \frac{\hat{w}_i}{\widehat{P}_i} = \prod_{\mu}\prod_{\mu'}\left(\lambda_{ii}(\mu)^{\frac{a_i(\mu,\mu')e_i(\mu)}{1-\sigma(\mu)}}\right).$$

where $a_i(\mu,\mu')$ is the $(\mu,\mu')$ entry of economy $i$'s inverse Leontief matrix and the last line follows from the fact that $\widehat{P}_i = \prod_\mu \widehat{P}_i(\mu)^{e_i(\mu)}$. The last line clearly indicates that $\hat{w}_i/\widehat{P}_i$ is independent of the underlying vector of markups. Accordingly, $\ln\left(\widehat{w}_i/\widehat{P}_i\right)$ on the right-hand side of Equation 20 corresponds to the gains from trade in an efficient economy. That is, $\ln\left(\widehat{w}_i/\widehat{P}_i\right) = \ln W_i^* - \ln W_i^{*a}$, which immediately implies that $\Delta\mathscr{D}_i = \ln(\widehat{\pi}_i)$. Considering this intermediate point, our goal herein is to derive a formula for $\ln(\widehat{\pi}_i)$ under input-output linkages. Recall from earlier that the profits margin in economy $i$ is

$$\pi_i \equiv 1 + \frac{\Pi_i}{w_i L_i} = \frac{1 - \sum_\mu \left[\frac{1-\nu_i(\mu)}{\mu} y_i(\mu)\right]}{\sum_\mu \left[\frac{\nu_i(\mu)}{\mu} y_i(\mu)\right]}.$$

Note that under autarky the gross revenue and expenditure shares are exactly the same, i.e., $y_i^a(\mu) = \tilde{e}_i^a(\mu)$. We can, thus, produce the following expression for $\widehat{\pi}_i = \pi_i/\pi_i^a$,

$$\widehat{\pi}_i = \frac{\left(1 - \sum_\mu \frac{1-\nu_i(\mu)}{\mu} y_i(\mu)\right) \sum_\mu \left[\frac{\nu_i(\mu)}{\mu} e_i^a(\mu)\right]}{\left(1 - \sum_\mu \frac{1-\nu_i(\mu)}{\mu} \tilde{e}_i^a(\mu)\right) \sum_\mu \left[\frac{\nu_i(\mu)}{\mu} y_i(\mu)\right]}.$$

Taking logs from the above equations and using the expectation notation introducer earlier, we arrive that following expression for the effect of trade on the DWL of markups as specified by Proposition 5:

$$\Delta\mathscr{D}_i = \ln(\widehat{\pi}_i) = \ln\left(\frac{\mathbb{E}_{y_i}\left[\frac{\nu_i(\mu)}{\mu}\right]}{1 - \mathbb{E}_{y_i}\left[\frac{1-\nu_i(\mu)}{\mu}\right]}\right) - \ln\left(\frac{\mathbb{E}_{\tilde{e}_i^a}\left[\frac{\nu_i(\mu)}{\mu}\right]}{1 - \mathbb{E}_{\tilde{e}_i^a}\left[\frac{1-\nu_i(\mu)}{\mu}\right]}\right)$$

### G.3 Inferring Autarky Gross Expenditure Shares from Observable Shares

The formulas for $\mathscr{D}_i^{closed}$ and $\Delta\mathscr{D}_i$ depend on gross expenditure shares under autarky, $\tilde{e}_i^a(\mu)$. So, to complete the proofs of Proposition 5 and Lemma 2 , we must characterize $\tilde{e}_i(\mu)$ as a function of observables. To this end, we use two relationships: First, that under autarky, $\mu\, \mathcal{C}_i^a(\mu) = E_i^a(\mu)$—that is, total revenues in from markup $\mu$, which are $\mu \mathcal{C}_i^a(\mu)$, equal counterfactual expenditure on that industry, $E_i^a(\mu)$, under autarky. Second, the accounting identity,

$$E_i(\mu) = e_i(\mu) Y_i + \sum_{\mu'} \left([1 - \nu_i(\mu')]\, \alpha_i(\mu, \mu')\, \mathcal{C}_i(\mu')\right),$$

which states that gross expenditure on markup $\mu$ 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

$$\begin{aligned}\tilde{e}_i^a(\mu) \equiv \frac{E_i^a(\mu)}{E_i^a} &= \frac{e_i(\mu) Y_i^a + \sum_{\mu'} \left([1 - \nu_i(\mu')]\, \alpha_i(\mu, \mu')\, \mathcal{C}_i^a(\mu')\right)}{E_i^a} \\ &= e_i(\mu) \frac{Y_i^a}{E_i^a} + \sum_\mu \left[\frac{[1 - \nu_i(\mu')]\, \alpha_i(\mu, \mu')}{\mu'} \tilde{e}_i^a(\mu')\right].\end{aligned} \tag{21}$$

We can, furthermore, express $Y_i^a/E_i^a$ (in the above equation) in terms the vector of gross expenditures shares $\{\tilde{e}_i^a(\mu)\}_\mu$ and observables. For this, we use the accounting identity, $Y_i = E_i - \sum_\mu [(1 - \nu_i(\mu))\, C_i(\mu)]$, which states that aggregate consumption expenditure across all industries equals aggregate gross expenditure minus input expenditure. Plugging $C_i^a(\mu) = E_i^a(\mu)/\mu$ in the aforementioned identity yields

$$\frac{Y_i^a}{E_i^a} = 1 - \sum_{\mu}\left[\frac{1-\nu_i(\mu)}{\mu}\tilde{e}_i^a(\mu)\right].$$

Plugging the above expression back into Equation 21 deliveries the following equation which implicitly characterizes $\tilde{e}_i^a(\mu)$ as a function of parameters and observable shares,

$$\tilde{e}_i^a(\mu) = e_i(\mu)\left[1-\sum_{\mu'}\left(\frac{1-\nu_i(\mu')}{\mu}\tilde{e}_i^a(\mu')\right)\right] + [1-\nu_i(\mu)]\sum_{\mu'}\left[\frac{\alpha_i(\mu,\mu')}{\mu'}\tilde{e}_i^a(\mu')\right] = e_i(\mu) + \sum_{\mu'}\left[\frac{1-\nu_i(\mu')}{\mu'}(\alpha_i(\mu',\mu) - e_i(\mu))\right.$$

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

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

where $\mathbf{\Phi}_i \equiv \left[\frac{[1-\nu_i(\mu')]\alpha_i(\mu,\mu')}{\mu'}\right]_{\mu',\mu}$ is the $K\times K$ markup-adjusted input-output matrix in country $i$ and $\tilde{\mathbf{e}}_i^a \equiv \left[\tilde{e}_i^a(\mu)\right]_\mu$ and $\mathbf{e}_i \equiv [e_i(\mu)]_\mu$ are $K\times 1$ column vectors. Inverting the above system yields the following closed-form expression for $\tilde{\mathbf{e}}_i^a \equiv [\tilde{e}_i(\mu)]_\mu$ as a function of observables,

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

# H The Deadweight Loss of Markups under Fixed Overhead Costs

Now we consider an extension of our baseline model where serving individual market requires a fixed overhead cost that consumes a fraction of the monopolistic rents. Considering the isomorphism between our semi-parametric baseline model and a multi-industry model with a constant industry-wide markup, we henceforth focus on the latter for a clearer exposition. More specifically, there are several categories of goods or industries indexed by $k = 1,...,K$. Each category is characterized by a constant markup. So, we can alternatively index goods based on their markup $\mu \in \mathcal{M} = \{\mu_1,...,\mu_K\}$.With this choice of notation in mind, we now specify the general equilibrium in this economy economy.

Firm $\omega \in \Omega_i(\mu)$ from origin $i$ with markup $\mu$ has to pay a fixed marketing cost, $w_j f_j(\mu)$, 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 in product category with markup $\mu$ independently draw their productivity, $\varphi$, from a Pareto distribution that has a product-specific shape parameter $\theta(\mu) > \gamma(\mu) - 1$. Under these assumptions, one can show that fixed marketing costs exhaust a constant fraction, $\rho(\mu)$, of origin $i$'s sales to destination. Namely,

$$M_{ij}(\mu)\, w_j f_j(\mu) = \rho(\mu)\; P_{ij}(\mu)\, Q_{ij}(\mu), \qquad \text{where} \qquad \rho(\mu) \equiv 1 - \frac{1+\theta(\mu)}{\mu\theta(\mu)}.$$

To be clear, $M_{ij}(\mu)$ denotes the mass of firms with markup $\mu$ that can profitably serve destination $j$ from origin $i$, which is a fraction of the total number of firms, $M_i(\mu)$. We can use this equation to derive a *firm-selection-adjusted* sufficient statistics formula for the impact of trade on the DWL (deadweight loss) 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, the DWL 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-described index of revealed comparative advantage, $\delta_i(\mu) \equiv \frac{y_i(\mu)}{e_i(\mu)} - 1$ 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, $\Delta\mathscr{D}_i$ depends not only on the pattern of specialization between low- and high-markup industries but also the pattern of specialization across low- and high-$\rho$ industries. The following proposition formalizes this point.

**Proposition 6.** *The effect of trade on the DWL of markups after accounting international fixed cost payments is given by*

$$\Delta\mathscr{D}_i(\mathbf{X}) = \ln\left(\mathbb{E}_{y_i}\left[\frac{1}{\mu}+\rho(\mu)\right] - \left(1+\frac{\mathbb{E}_{y_i}\left[\frac{1}{\mu}\right]}{1-\mathbb{E}_{e_i}\left[\rho(\mu)\right]}\right)\mathrm{Cov}_{e_i}\left(\rho(\mu),\delta_i(\mu)\right)\right) - \ln\mathbb{E}_{e_i}\left[\frac{1}{\mu}+\rho(\mu)\right]$$

*where $\delta_i(\mu) \equiv \frac{y_i(\mu)}{e_i(\mu)} - 1$ is a general index of revealed comparative advantage. The above equation can be evaluated given sales-weight average markups per industry, shape of the firm-size distribution, and observable shares, i.e,* $\mathbf{X} = \{\mu, \theta(\mu), y_i(\mu), e_i(\mu)\}$.

The above proposition is proven in following sub-appendix. To give some intuition, $\frac{1}{\mu} + \rho(\mu)$ can be interpreted as the inverse of the excess markup. As note earlier, the heterogeneity in excess markups determines the DWL of markups in the present setup. Accordingly, when $\rho(\mu) = 0$, which corresponds to the limiting case with zero fixed cost payments, the above formula reduces to the baseline formula specified under Proposition 2.

The adjustment, $\mathrm{Cov}_{e_i}(\rho(\mu), \delta_i(\mu))$, accounts for fixed cost payments transferring rents from one country to another. This term balances fixed cost payments paid to foreign workers against the corresponding payments received by domestic workers from foreign firms. The following example may help understand the crucial role of fixed marketing costs. Proposition 7 indicates that—unlike the baseline model—trade can amplify the DWL of markups even if markups are uniform across industries. In particular, one can easily check from Proposition 7 that if markups are uniform across all goods, then $\Delta\mathscr{D}_i \neq 0$ unless $\rho(\mu)$ or $\theta(\mu)$ are also uniform. Intuitively, if the degree of firm heterogeneity, $\theta(\mu)$, 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 alleviate the DWL of markups depending on whether resources are relocated to high- or low-$\rho$ industries.

## H.1 Characterizing Net Profit Margins

As an intermediate step, we characterize the aggregate profit margin ($\pi_i$) 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 $\pi_i$. In what follows we characterize $\pi_i$ in the presence of fixed costs for an economy that is either closed or open. Recall from Section 6.1 that fixed marketing costs in our model account for a constant fraction of origin $i$'s sales to destination $j$. Namely,

$$M_{ij}(\mu)\, w_j f_j(\mu) = \left(1 - \frac{1+\theta(\mu)}{\mu\theta(\mu)}\right) P_{ij}(\mu)\, Q_{ij}(\mu).$$

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

$$\Pi_i = \left[1 - \sum_{\mu\in\mathcal{M}} \left(\frac{1}{\mu} + \rho(\mu)\right) y_i(\mu)\right] \mathcal{R}_i, \tag{22}$$

where $y_i(\mu)$, 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,

$$w_i L_i = \sum_{\mu\in\mathcal{M}} \left[\frac{1}{\mu} y_i(\mu)\right] \mathcal{R}_i + \sum_{\mu\in\mathcal{M}} \left[\rho(\mu)\, e_i(\mu)\right] E_i, \tag{23}$$

where $E_i = \sum_j \sum_\mu P_{ji}(\mu)\, Q_{ji}(\mu)$ 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 22 and 23 yields the following relationship between national-level revenues and expenditure in country $i$

$$\mathcal{R}_i = \frac{1 - \sum_\mu \rho(\mu)\, e_i(\mu)}{1 - \sum_\mu \rho(\mu)\, y_i(\mu)} E_i.$$

Plugging the above expression back into Equations 22 and 23, yields the following formula for the aggregate profit margin in country $i$:

$$\pi_i \equiv \frac{w_i L_i + \Pi_i}{w_i L_i} = \frac{E_i}{\left[\sum_\mu \frac{1}{\mu} y_i(\mu)\right] \mathcal{R}_i + \sum_\mu \left[\rho(\mu)\, e_i(\mu)\right] E_i} = \frac{1}{\frac{1-\sum_\mu \rho(\mu) e_i(\mu)}{1-\sum_\mu \rho(\mu) y_i(\mu)} \sum_\mu \left(\frac{1}{\mu} y_i(\mu)\right) + \sum_\mu \left(\rho(\mu)\, e_i(\mu)\right)}. \tag{24}$$

### H.2 The DWL of Markups in a Closed Economy

Following the logic presented earlier, the DWL of markups for closed economy $i$ is given by $\mathscr{D}_i^{closed} = \ln W_i^{a*} - \ln W_i^a$, where superscript "$*$" corresponds to efficient equilibrium and superscript "$a$" denotes autarky. Taking into account that $W_i = \pi_i w_i L_i / P_i$ and assigning labor in country $i$ as the numeraire (i.e., $w_i^a = w_i^{a*} = 1$), we get

$$\mathscr{D}_i^{closed} = \ln \widehat{\pi}_i^a - \ln \widehat{P}_i^a,$$

where $\widehat{\pi}_i^a = \pi_i^{a*}/\pi_i^a$ and $\widehat{P}_i^a = P_i^{a*}/P_i^a$. 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 $m_i(\varphi;\mu) < \mu$ denote the excess markup in market $i$, which depends on the firm productivity, $\varphi$. After eliminating the excess markup for the existing set of firms, the change in the product-level CES price index is given by

$$\hat{P}_i^{a*}(\mu) = \left[\int_{\varphi_{ii}^{\star}(\mu)}^{\infty} m_i(\varphi;\mu)^{\gamma(\mu)-1} \lambda_{ii}(\varphi;\mu)\, dG_i(\varphi;\mu)\right]^{\frac{1}{1-\gamma(\mu)}} = \left[\int_1^{\infty} m^{\gamma(\mu)-1} \lambda_{ii}(m;\mu)\, d\mathcal{G}_i(m;\mu)\right]^{\frac{1}{1-\gamma(\mu)}}$$

where $\varphi_{ii}^{\star}(\mu)$ denotes the zero profit productivity cut-off and $\mathcal{G}_{i,k}(.)$ denotes the distribution of excess markup in market $i$.

To economize on the notation, we use $\mathcal{M}_i(\mu) \equiv \left[\int_1^{\infty} m^{\gamma(\mu)-1}\lambda_{ii}(m;\mu)\, d\mathcal{G}_i(m;\mu)\right]^{\frac{1}{\gamma(\mu)-1}}$ to denote the CES average *excess* markup. The change in the aggregate consumer price index is, correspondingly, $\ln \widehat{P}_i^a = \ln\left[\prod_\mu \hat{P}_i^a(\mu)^{e_i(\mu)}\right] = -\mathbb{E}_{e_i}[\ln \mathcal{M}_i(\mu)]$. To determine $\widehat{\pi}_i^a$, we can appeal to Equation 24, noticing that expenditure and revenue shares coincide under autarky (i.e., $y_i^a(\mu) = e_i^a(\mu)$) and excess profit margins are zero under the efficient equilibrium, $\pi_i^{a*} = 1$. In particular,

$$\hat{\pi}_i^a = \sum_{\mu \in \mathcal{M}} \left(\left[\frac{1}{\mu} + \rho(\mu)\right] e_i(\mu)\right),$$

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

$$\mathscr{D}_i^{closed} = \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu} + \rho(\mu)\right] + \mathbb{E}_{e_i}[\ln \mathcal{M}_i(\mu)].$$

### H.3 Proof of Proposition 6

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., $\ln \widehat{P}_i^a = \ln \widehat{P}_i =$

$-\mathbb{E}_{e_i}[\ln \mathcal{M}_i(\mu)]$, where $\widehat{P}_i = P_i^*/P_i$ and $\mathcal{M}_i(\mu)$ is invariant to trade following the logic outlined in Section C. As such, the impact of trade on the DWL of markups is

$$\mathscr{D}_i - \mathscr{D}_i^{closed} = \ln\left(\widehat{\pi}_i/\widehat{\pi}_i^a\right) - \ln\left(\widehat{P}_i/\widehat{P}_i^a\right) = \ln\left(\widehat{\pi}_i/\widehat{\pi}_i^a\right),$$

where $\widehat{P}_i = P_i^*/P_i$ and $\widehat{\pi}_i = \pi_i^*/\pi_i$ 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 $\widehat{\pi}_i^a = -\ln \mathbb{E}_{e_i}\left[\frac{1}{\mu} + \rho(\mu)\right]$. Applying our notation for expectations to Equation 24, implies

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

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

$$\Delta\mathscr{D}_i = \ln\left(\mathbb{E}_{y_i}\left[\frac{1}{\mu} + \rho(\mu)\right] - \left(1 + \frac{\mathbb{E}_{y_i}\left[\frac{1}{\mu}\right]}{1-\mathbb{E}_{e_i}[\rho(\mu)]}\right)\text{Cov}_{e_i}(\rho(\mu), \delta_i(\mu))\right) - \ln \mathbb{E}_{e_i}\left[\frac{1}{\mu} + \rho(\mu)\right]$$

Recall that $\rho(\mu)$ 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 the DWL of markups—which was presented under Proposition 1.

# I Other Extensions

In this appendix we explore two additional extensions, one with capital as a primary production factor and another where labor is supplied elastically. We derive formulas for the deadweight loss of monopolistic distortions in both cases, characterizing how trade affects the deadweight loss. In both case, the zero-sum effects of trade on the incidence of monopolistic distortions continue to hold, albeit with some modification.

## I.1 Capital as a Primary Production Input

Suppose production employs labor and capital inputs. Whereas labor is perfectly mobile across the production of different products, capital inputs are poduct-specific, with $\bar{\mathcal{K}}(\mu)$ denoting the constant supply of capital for the production of goods with markup $\mu$. Suppose the production function has a Cobb-Douglas parametrization with $1-\varsigma_i(\mu)$ denoting the constant share of capital in production.

The monopolistically competitive price index of goods sold by origin $i$ to destination $j$ is

$$P_{ij}(\mu) = \mu \frac{\tau_{ij}(\mu)}{\bar{\varphi}_i(\mu)} w_i^{\varsigma(\mu)} r_i(\mu)^{1-\varsigma(\mu)},$$

where $r_i(\mu)$ denotes the rental rate of capital used in origin $i$ for the production of markup $\mu$ goods. Note that per cost minimization, $r_i(\mu)\bar{\mathcal{K}}_i(\mu) = \frac{1-\varsigma(\mu)}{\varsigma(\mu)} w_i L_i(\mu)$, where $L_i(\mu)$ denotes the labor employed for markup $\mu$ products in origin $i$, with $\sum_\mu L_i(\mu) = L_i$ per the labor market clearing condition. Letting $\ell_i(\mu) = \frac{L_i(\mu)}{L_i}$ denote the labor share per markup category, the rental-rate-adjusted price indexes associated with economy $i$ can be re-written as

$$P_{ij}(\mu) = \mu \bar{a}_{ij}(\mu) w_i \ell_i(\mu)^{1-\varsigma(\mu)}; \qquad P_i = \prod_\mu \left( \sum_n P_{ni}(\mu)^{-\theta(\mu)} \right)^{-\frac{e_i(\mu)}{\theta(\mu)}},$$

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

$$Y_i = w_i L_i + \sum_\mu \left[ r_i(\mu) \bar{\mathcal{K}}_i(\mu) \right] + \Pi_i.$$

Noting that $r_i(\mu)\bar{\mathcal{K}}_i(\mu) = \frac{1-\varsigma(\mu)}{\varsigma(\mu)} w_i \ell_i(\mu) L_i$ and $\Pi_i = \sum_\mu \left[ (\mu - 1)\left( w_i L_i(\mu) + r_i(\mu)\bar{\mathcal{K}}_i(\mu) \right) \right]$, we can re-write the expression for nominal income as

$$Y_i = \sum_k \left[ \frac{\mu}{\varsigma(\mu)} \ell_i(\mu) \right] w_i L_i = \mathbb{E}_{\ell_i} \left[ \frac{\mu}{\varsigma(\mu)} \right] w_i L_i. \tag{25}$$

With the above background, we are ready to characterize the DWL of markup, which corresponds to the change in welfare after markups are eliminated. More specifically, we wish to characterize $\mathscr{D}_i \equiv \ln \hat{W}_i = \ln \left( W_i^* / W_i \right)$, where $W_i = Y_i / P_i$ and the $(*)$ 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

$$\hat{P}_i = \prod_\mu \left( \frac{1}{\mu} \hat{w}_i \hat{\ell}_i(\mu)^{1-\varsigma(\mu)} \right)^{e_i(\mu)}.$$

We can set the change in the wage rate to one by choice of numeraire, i.e., $\hat{w}_i = 1$. To characterize $\hat{\ell}_i(\mu)$, note that $w_i L_i(\mu) = \frac{\varsigma(\mu)}{\mu} P_{ii}(\mu) Q_i(\mu)$, which yields the following relationship between labor and revenue shares:

$$\ell_i(\mu) = \frac{\frac{\varsigma(\mu)}{\mu} y_i(\mu)}{\sum_{\mu'} \frac{\varsigma(\mu')}{\mu'} y_i(\mu')}. \tag{26}$$

which given that under autarky, $y_i^{(a)}(\mu) = e_i(\mu)$ is constant, implies the following

$$\ell_i^{(a)}(\mu) = \frac{\frac{\varsigma(\mu)}{\mu} e_i(\mu)}{\sum_{\mu'} \frac{\varsigma(\mu')}{\mu'} e_i(\mu')}; \qquad \ell_i^{(a)*}(\mu) = \frac{\varsigma(\mu) e_i(\mu)}{\sum_{\mu'} \varsigma(\mu') e_i(\mu')}.$$

Combing the above expressions we can determine $\hat{\ell}_i(\mu) = \ell_i^{(a)*}(\mu) / \ell_i^{(a)}(\mu) = \mu \frac{\sum_{\mu'} \frac{1}{\mu'} \varsigma(\mu) e_i(\mu')}{\sum_{\mu'} \varsigma(\mu) e_i(\mu')} \sim \mu \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right]$, which when plugged into our earlier expression for $\ln \hat{P}_i$ deliverslnWe can characterize $\hat{Y}_i$ by appealing to Equations 25 and 26, which yields

$$\ln \hat{Y}_i = \ln\left(\frac{\sum_\mu \left(\frac{1}{\varsigma(\mu)} \ell_i^{(a)*}\right)}{\sum_\mu \left(\frac{\mu}{\varsigma(\mu)} \ell_i^{(a)}(\mu)\right)} \hat{w}_i L_i\right) = \ln\left(\frac{\sum_{\mu'} \frac{1}{\mu'} \varsigma(\mu) e_i(\mu')}{\sum_{\mu'} \varsigma(\mu) e_i(\mu')}\right) \sim \ln \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right],$$

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 DWL of markups in a closed economy, $\mathscr{D}_i = \ln \hat{W}_i = \ln(\hat{Y}_i / \hat{P}_i)$,

$$\mathscr{D}_i = \ln \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right] \mathbb{E}_{e_i}[\varsigma(\mu)] - \mathbb{E}_{e_i}\left[\varsigma(\mu) \ln \frac{1}{\mu}\right].$$

Based on the above formula we can infer the DWL 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 $e_i(\mu) = e_n(\mu) = e(\mu)$ for $i$ and $n$. Note, restoring of marginal cost pricing global and conditional on wages constant via transfers preserves relative prices internationally within markup categories, implying that $\hat{\ell}_i(\mu) = \hat{\ell}_n(\mu)$ and $\hat{w}_i = \hat{w}_n$ for all $i$ and $n$. Moreover, given Equation 26 and internationally symmetric Cobb-Douglas preferences across product categories, it is straightforward to check that

$$\hat{\ell}_i(\mu) = \hat{\ell}_n(\mu) = \mu \mathbb{E}_{\varsigma e}\left[\frac{1}{\mu}\right],$$

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

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

Likewise, as before, we can combine Equations 25 and 26 to derive the following expression for the change in nominal income

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

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

$$\Delta\mathscr{D}_i = \ln \mathbb{E}_{\varsigma y_i}\left[\frac{1}{\mu}\right] - \ln \mathbb{E}_{\varsigma e_i}\left[\frac{1}{\mu}\right].$$

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 $\varsigma(\mu) = 1$ for all $\mu$, the above equation reduces to our baseline formula for $\Delta\mathscr{D}_i$.

### I.2 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 $i$, $\Pi_i$ is profit income, 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 $\epsilon$. 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 $\ell_i$ 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 $\ln \hat{P}_i = \mathbb{E}_{e_i}\left[\ln\frac{1}{\mu}\right]$. Plugging the expression for $\hat{P}_i$ back into the above equation and rearranging the terms, yields

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

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

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

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 the DWL of markups can be, accordingly, measured as

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

The first term represents international rent-shifting, the impact of which is amplified by the labor supply elasticity, $\epsilon$. 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, $\epsilon$.

# J Duality Between Monopolistic Markups and Tariffs

**Weak Duality (Proposition 3).** Consider an initially-efficient economy with no markup distortions ($\boldsymbol{\mu} = \mathbf{1}$) and let country $i$ introduce a flat export tax, $x_i$. Total income in country $i$ after the introduction of the export tax can be written as

$$Y_i = \left[1 + \sum x_i \sum_{n\neq i}\sum_{k}\ell_{in,k}(x_i)\right] w_i L_i,$$

where $\ell_{in,k} = L_{in,k}(x_i)/L_i$, with$L_{in,k}(x_i)$ denoting the demand for country $i$'s labor services from location $n$ in industry $k$. The labor demand $L_{in,k}(x_i)$ is obtained by converting the demand for traded goods into the labor services required to produce those goods. The change in country $i$'s welfare can be represented as

$$\Delta W_i = \int_0^{x_i}\left[\frac{\partial Y_i}{\partial x}dx - \sum_k\sum_n \lambda_{ni,k}(x)\,\frac{\partial P_{ni,k}}{\partial x}dx\right],$$

which considering that $\frac{\partial P_{ni,k}}{\partial x} = 0$ if $n \neq i$ and $\frac{\partial P_{ii,k}}{\partial x} = dw_i$, yields the following welfare change formula,

$$\begin{aligned}\Delta\ln W_i(x_i) &= \ln\left[1 + x_i\sum_{n\neq i}\sum_k \ell_{in,k}(x_i)\right] + \left(1-\overline{\lambda}_{ii}\right)\Delta\ln w_i\\ &= \ln\left[1 + x_i\ell_{-ii}(x_i)\right] + \left(1-\overline{\lambda}_{ii}\right)\Delta\ln w_i,\end{aligned}$$

where the last invokes the adding up constraint, $\sum_{n\neq i}\sum_k \ell_{in,k}(x_i) = 1 - \ell_{-ii}(x_i)$. A well-behaved demand function entails that $\lim_{x\to\infty} x \cdot \ell(x) = 0$, which in turns implies $\lim_{x_i\to 0}\Delta\ln W_i(x_i) = 0$. Also, if the trade elasticity (which regulated the demand for labor services) is arbitrarily small, then $\partial \ln \ell_{-ii}(x_i)/\partial \ln x_i \to 0$, implying that $\lim_{x_i\to\infty}\Delta\ln W_i(x_i)$ is arbitrarily large while $\lim_{x_i\to-\infty}\Delta\ln W_i(x_i)$ is arbitrarily small. Hence, following the Intermediate Value Function Theorem, there exists a uniform export tax $\tilde{x}_i$ that reproduced the welfare effects of rent-shifting for country $i$

$$\lim_{x_i\to-\infty}\Delta\ln W_i(x_i) \le \Delta\ln W_i(\tilde{x}_i) = \ln\left(\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right) \le \lim_{x_i\to\infty}\Delta\ln W_i(x_i).$$

We can now extend this logic to the multi-variate case,

$$\mathbf{x} = (x_1, ..., x_N) \qquad \Delta\ln\mathbf{W}(\mathbf{x}) = (\Delta\ln W_1(\mathbf{x}), ..., \Delta\ln W_N(\mathbf{x})),$$

to concludes that there exists a $\tilde{\mathbf{x}} \in \mathbb{R}^N$ that can yield $\Delta\ln\overline{\mathbf{W}} = (\Delta\mathscr{D}_1, ..., \Delta\mathscr{D}_N)$, where $\Delta\mathscr{D}_i = \ln\left(\frac{\sum_k y_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}{\sum_k e_{i,k}\widetilde{\mathbb{E}}_{\rho_k}[\mu]^{-1}}\right)$ denotes the change in the deadweight loss of markups due to rent-shifting. Next, we can invoke the Lerner symmetry, which asserts that the vector of uniform tariffs $\tilde{\mathbf{t}} = \tilde{\mathbf{x}}$yields equivalent welfare outcomes to the export tax $\tilde{\mathbf{x}}$. Accordingly, there exists a vector of uniform tariffs $\tilde{\mathbf{t}} = \tilde{\mathbf{x}}$ that exactly reproduces the welfare effects associated with rent-shifting—hence, the weak duality between tariffs and monopolistic markups.

**Strong Duality under Sufficiently Uniform Trade Elasticities.** We prove the strong duality in the case where all firms within industry $k$ set a common markup, $\mu_k$. We can express welfare as an explicit function of tariffs and industry-level markups, $\boldsymbol{\mu} = \{\mu_1, ..., \mu_k\}$, as $W_i = \mathbb{W}_i(\mathbf{t}, \boldsymbol{\mu})$. 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,

$$\exists\, \mathbf{t}: \qquad \mathbb{W}_i(\mathbf{t}, \mathbf{0}) = \mathbb{W}_i(\mathtt{0}, \boldsymbol{\mu}).$$

Observe that markups are akin to industry-level production taxes that exhibit the same rate in all countries. Extrapolating from this observation, an internationally uniform but industry-specific tariff, $t^{(\mu)}_{ni,k} = \mu_k - 1$ for all $ni,k \neq ii,k$, is more distortionary than equal rate markups from a global standpoint. That is, with properly-scaled welfare functions, $\sum_i \mathbb{W}_i\left(\mathbf{t}^{(\mu)}, \mathbf{0}\right) \le \sum_i \mathbb{W}_i(\mathtt{0}, \boldsymbol{\mu})$. To elaborate, the welfare gains/losses from tariffs is the sum of tariff dispersion and wage-driven terms-of-trade effects—analogous to markup distortions as specified by Equation **??** of the main text. The dispersion in tariffs, $\mathbf{t}^{(\mu)}$, exerts a greater deadweight loss than $\boldsymbol{\mu}$, since the tariff rate on the domestically produced goods is zero, creating both cross-national and cross-industry wedge dispersion. However, there may be countries for which $\mathbb{W}_i\left(\mathbf{t}^{(\mu)}, \mathbf{0}\right) > \mathbb{W}_i(\mathtt{0}, \boldsymbol{\mu})$, since a country that experiences losses from rent-shifting under the markup schedule, $\boldsymbol{\mu}$, may experience terms-of-trade improvements under the tariff schedule, $\mathbf{t}^{(\mu)}$.

With the above background in mind, we establish three intermediate claims. First, following Lashkaripour and Lugovskyy (2023), the unilaterally optimal tariff in each country is uniform across industries if markups were eliminated to zero. Accordingly, for the vector of tariffs, $\mathbf{t}^{(\mu)}$, there exists a uniform (non-industry-specific) 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, which are akin to a production taxes, yield a strictly lower welfare for a country than the unilaterally optimal export tax, $x^*_{ij,k}$.[30] Also, following Lashkaripour and Lugovskyy (2023), the optimal export tax for an efficient small open economy is related to the trade elasticity as $x^*_{ij,k} = \frac{1}{\theta_k}$ for all $ij,k$. Third, if trade elasticities, $\theta_k$, are uniform, then the optimal import tax is welfare-equivalent to the optimal export tax by the Lerner symmetry (Costinot and Werning, 2019). Combining these results, we can apply the Intermediate Value Theorem to the continuous function, $\mathbb{W}_i(.)$ 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

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

where recall that $\bar{t}_i$ is the country $i$'s uniform tariff-equivalent of the tariff schedule, $\mathbf{t}^{(\mu)}$—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 real world) or sufficiently different from the unilaterally optimal schedule for individual countries.

## K Demand-Based Markup Estimation: *Details*

This appendix provides a more details about our demand-based markup estimation procedure. As explained in the main text we conduct our estimation with set $\mathbf{X} = \left\{\ln p, D_{\text{origin}\times\text{HS10}}\right\}$ of observed product characteristics, where $p$ denotes the unit price and $D_{\text{origin}\times\text{HS10}}$ is an identifier for origin country and 10-digit product code. Our identification strategy is borrowed from Lashkaripour and Lugovskyy (2023), and amended to account for heterogeneity in individual-level demand elasticities. As explained in the main text, Theorem 1 in Salanié and Wolak (2019) asserts that the mixed multinomial logit model of demand can be approximated by a the following log-linear equation:

$$\Delta \ln \lambda_{kt}(\omega) = -\beta_{kt}\Delta \ln p_{kt}(\omega) + \sigma^2_{kt}\Delta K_{kt}(\omega) + D_{\omega k} + \Delta \ln \lambda_{0t} + \Delta\tilde{\xi}_{kt}(\omega), \tag{27}$$

where $\Delta\tilde{\xi}_{kt}(\omega)$ represents the variety-specific demand shock based on the approximation in Salanié and Wolak (2019), $D_{\omega k}$ is an origin-product fixed effects, and $\Delta \ln \lambda_{0t}$ is absorbed by the product–year fixed effect. Annual changes in price and import shares, $\Delta \ln p_{kt}(\omega)$ and $\Delta \ln \lambda_{kt}(\omega)$ are directly observable

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

for each import variety. The artificial regressor, $K_{kt}(\omega)$, controls for hidden demand heterogeneity, with $\sigma_{kt}$ representing the standard deviation parameter that governs the heterogeneity in individual-level demand slopes. This artificial regressor is constructed as

$$K_{kt}(\omega) \equiv \left( \frac{1}{2} \ln p_{kt}(\omega) - \sum_{\Omega_{kt}} \lambda_{kt}(\omega') \ln p_{kt}(\omega') \right) \ln p_{kt}(\omega),$$

using variety-level unit price data. As explained in the main text, $K_{kt}(\omega)$ to a first-order approximation accounts for individual-level heterogeneity in demand slopes. Absent individual-level demand heterogeneity (i.e., $\sigma_{kt} \to 0$), Equation 27 reduces to a standard CES demand function estimated by DellaVigna and Gentzkow (2019).

In our estimation equation, $k$ indexes an HS10 product category. But to calibrate our sufficient statistics formulas, we must estimate average demand elasticities and markups for broadly-defined ICIO sectors for which we posses the necessary trade and production data.[31] We, therefore, pool together HS10 products belonging to the same ICIO sector, $\mathcal{S}$, and estimate Equation 27 on this pooled sample assuming that $\sigma_{kt}$ and $\epsilon_{kt}^{(D)}$ are uniform across products within the same industry. That is,

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

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.

We face an identification challenge in that the change in log price $\Delta \ln p_{kt}(\omega)$ is an endogenous variables that can covary with the demand shock $\tilde{\xi}_{kt}(\omega)$. While country-level import demand estimations often use tariff rates as instruments for prices, this approach doesn't work for our firm-level estimation since tariffs vary by country of origin but not by firm within a country. To address this issue, we use a shift-share research design based on two key observations. First, a given product variety is typically imported under multiple invoices across different months within a year. The annual price of a variety is the quantity-weighted average of its monthly prices: $p_{kt}(\omega) = \sum_{m \in \mathbb{M}_t} \rho_{kt}(\omega; m) p_{kt}(\omega; m)$ where $m$ is the month, $\rho_{kt}(\omega; m)$ is the quantity share, and $p_{kt}(\omega; m)$ is the price for month m. Second, a variety's monthly price in Colombian Pesos equals the product of its markup-plus-taxes, marginal input cost in local currency, and the month's exchange rate: $p_{kt}(\omega; m) = \mu_{kt}(\omega) \times \mathrm{MC}_{kt}(\omega) \times \mathcal{E}_t(\omega; m)$ where $\mu$ is markup-plus-tax, $MC$ is marginal cost, and $\mathcal{E}_t(\omega; m)$ is the exchange rate between the firm's origin country's currency and the Colombian Peso in month $m$ of year $t$. Approximating to the first order, the change in a variety's annual price due to monthly exchange rate shocks is:

$$\Delta \ln p_{kt}(\omega) \approx \sum_{m \in \mathbb{M}_t} \rho_{kt}(\omega; m)\, \Delta \ln \mathcal{E}_t(\omega; m)$$

[31] We conducted our demand estimation for 19 broadly-defined and traded ICIO sectors for which we have sufficient trade data to conduct the estimation.

where $\Delta \ln \mathcal{E}_t(\omega; m)$ is the year-over-year change in origin country's exchange rate in month $m$, and $\rho_{kt}(\omega; m)$ is month $m$'s share in the variety's annual export sales to Colombia. Our shift-share instrument is the inner product of *lagged* monthly export shares and monthly exchange rate shocks:

$$z_{kt}(\omega) = \sum_{m \in \mathbb{M}_t} \rho_{kt-1}(\omega; m) \Delta \ln \mathcal{E}_t(\omega; m).$$

In essence, $z_{kt}(\omega)$ captures a firm's exposure to exchange rate shocks at the *firm*×*origin*×*product*×*year* level, based on the idea that aggregate exchange rate shocks affect firms differently depending on the monthly distribution of their prior exports to Colombia. There is a strong, statistically significant correlation between $z$ and $\Delta \ln p$ supporting the relevance of this instrument, as verified by the first-stage F-statistics reported in Table A3, below. We also use the number of (alternative) product codes served by firm $\omega$ in year $t$ to instrument for $\Delta K_{kt}(\omega)$. This validity of this instruments follows the standard assumption in the literature that entry decisions are made prior to the realization of idiosyncratic demand shocks, $\tilde{\xi}_{kt}(\omega)$.

Our identifying assumption is the instrument $z$ is uncorrelated with the error term in our estimating equation, i.e., $\mathbb{E}\left[z_{kt}(\omega) \Delta\tilde{\xi}_{kt}(\omega)\right] = 0$. This assumption requires that two conditions be satisfied: $(a)$ Past pricing decisions (and thus, lagged export shares) are uncorrelated to current demand shocks: $\mathbb{E}\left[\Delta \ln p_{kt-1}(\omega)\, \tilde{\xi}_{kt}(\omega)\right] = 0$. $(b)$ Monthly country-level exchange rate fluctuations are unrelated to product-level demand shocks: $\mathbb{E}\left[\Delta \ln \mathcal{E}_t(\omega)\; \tilde{\xi}_{kt}(\omega)\right] = 0$. Because our sample of import transactions has many firms but only a few months, the first condition is enough to ensure our estimates are consistent (Goldsmith-Pinkham et al. (2020)). The second condition is more important for the finite sample properties of our estimator. Both conditions could be violated if there are connections between inventories or if a few export products make up a large share of a country's total exports to Colombia; but as Lashkaripour and Lugovskyy (2023) show these situations can be ruled out in the case of our data.

We estimate $\beta_{kt}$ as a moving average demand elasticity by fitting Equation 27 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 the pooled estimation results reported in Table A3. This table results data at the level of ICIO sectors for which we want to obtain markup estimates. Apart from the weighted average demand demand elasticity, the table reports 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.

## L 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

Table A3: Pooled Demand estimation results by ICIO sector

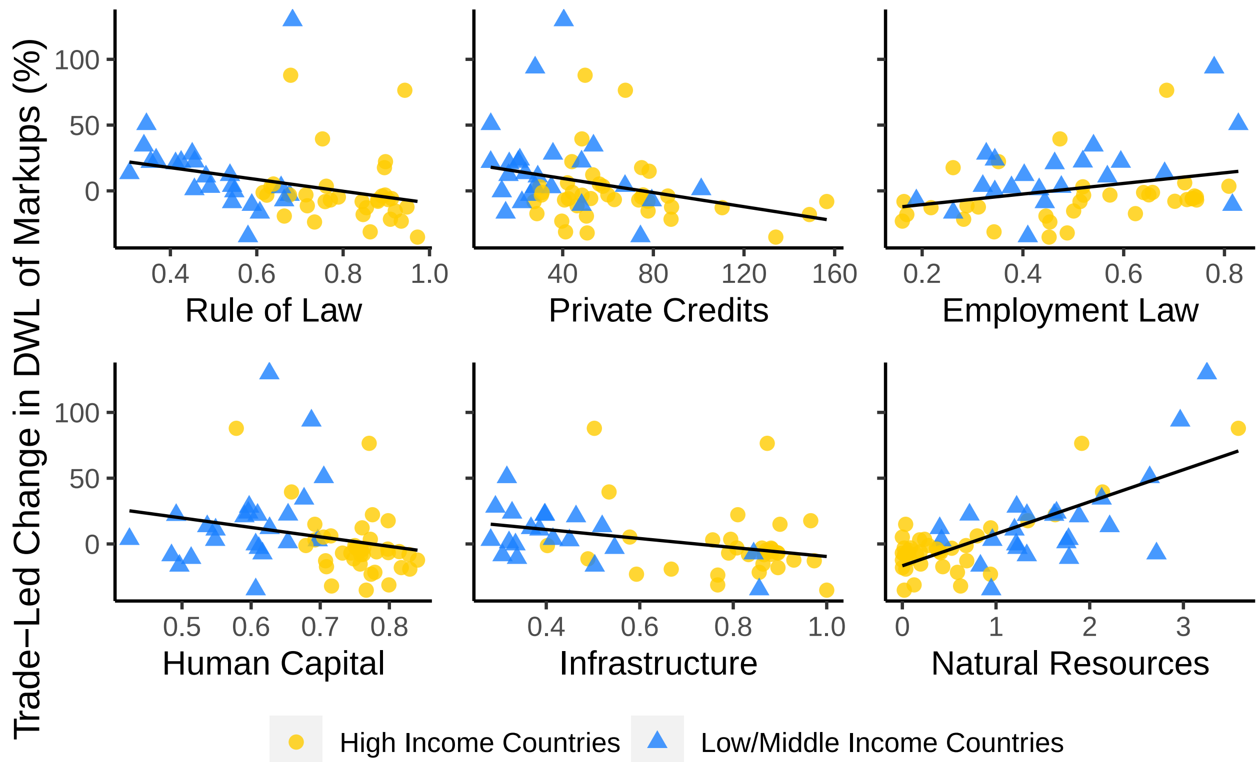
| | | | Estimated Parameter | | | |
|---|---|---|---|---|---|---|
| ICIO | Description | ISIC | $\beta_k$ | std error | F-stat | Observations |
| 1-3 | Agriculture & Mining | 1-9 | 5.26 | 3.45 | 16,331 | 3.72 |
| 4 | Food | 10-12 | 3.01 | 0.43 | 35,266 | 21.96 |
| 5 | Textiles, Leather & Footwear | 13-15 | 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 (27). 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).

and resource endowments. This appendix explores how these deep characteristics may explain the cross-national heterogeneity in $\Delta\mathscr{D}_i$ , which denotes the increase in the DWL of markups due to international rent-shifting.

Figure A1 displays the relation between $\Delta\mathscr{D}_i$ and some well-known determinants of comparative advantage. The upper panel of Figure A1 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 A1, countries that score better on the "rule of law" indicator are net beneficiaries from rent-shifting (i.e., exhibit a negative $\Delta\mathscr{D}_i$). 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.

Figure A1: Determinants of $\Delta\mathscr{D}$: national institutions and resource endowment




*Note*: The variable on the y-axis is the trade-induced change in the deadweight loss of markups ($\Delta\mathscr{D}_i$) as implied by our baseline model in 2015 with demand-based markups. 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.

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 A1 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 A1 displays the effect of infrastructure and factor endowment on $\Delta\mathscr{D}_i$. 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) and our measure of natural resource endowment is from the World Bank. Figure A1, in summary, suggests that better infrastructure and human capital contribute to specialization in high-markup industries and becoming net beneficiaries of rent-shifting (i.e., exhibiting a negative $\Delta\mathscr{D}_i$).

Among the many factors examined in Figure A1, natural resource abundance plays a pivotal role. Figure A2, therefore, explores the dependence of rent-shifting on natural resource-abundance in more depth. It displays the increase in the DWL 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).

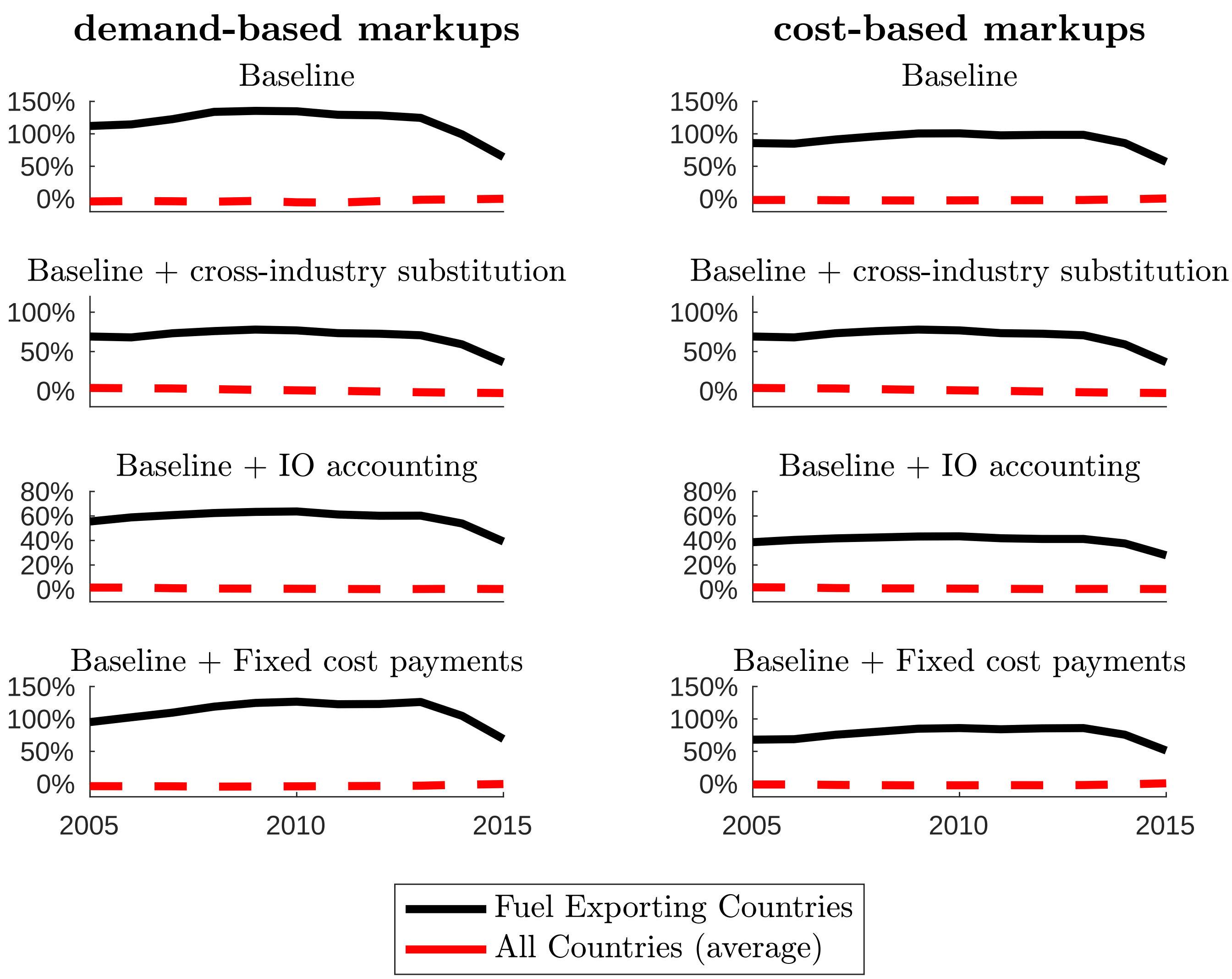
## M The Evolution of International Rent-Shifting Patterns

The results presented in Section 8 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 A3 examines these two possibilities by plotting the change in the DWL of markups due to

[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".

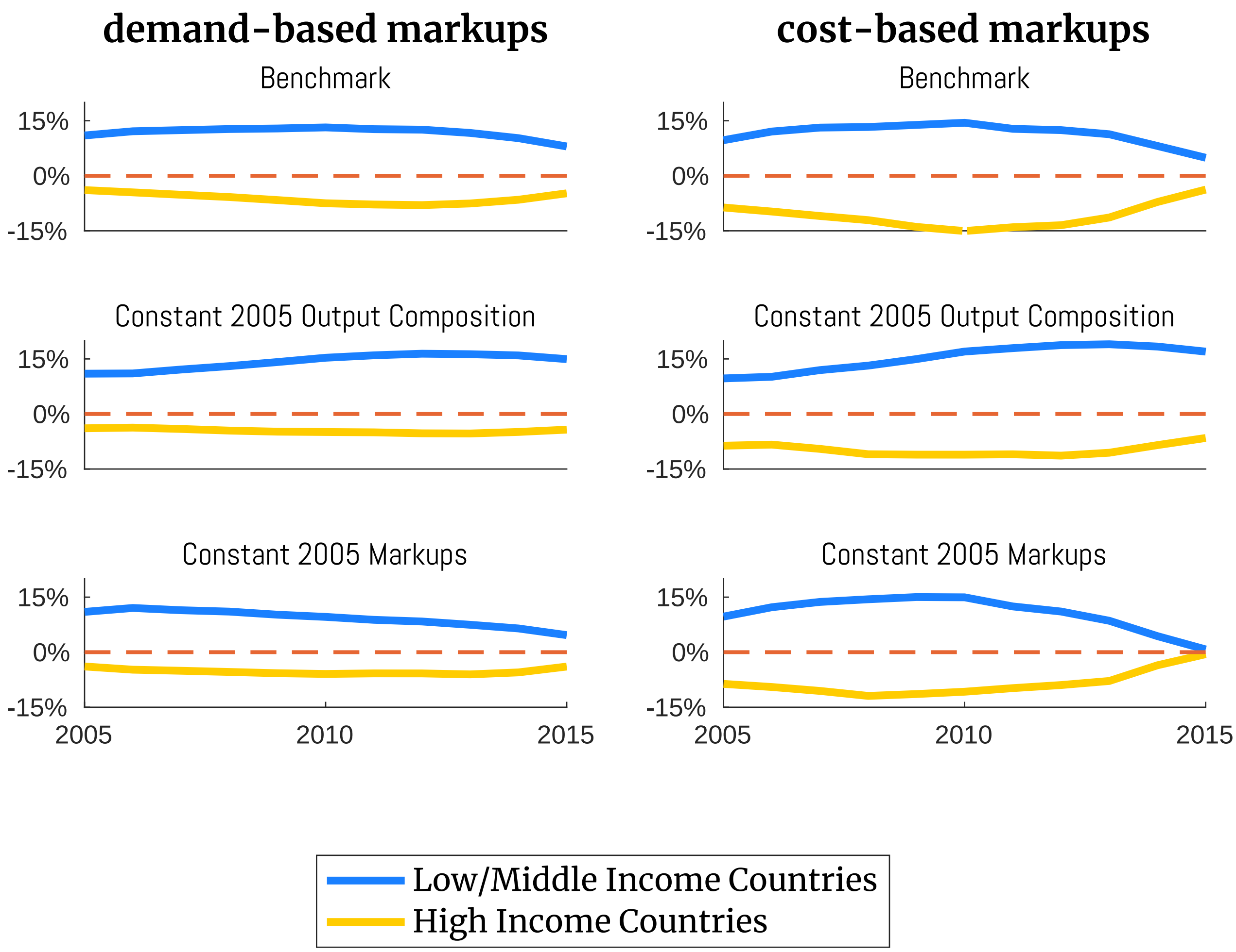
Figure A2: Trade-induced change in the deadweight loss of markups: fuel-exporting countries




*Note*: This figure displays the inferred % change in the deadweight loss of markups, $\mathscr{D}_i$, in response to trade openness. For example, a change of 5% corresponds to a 5% increase in the deadweight loss of markups 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.

international rent-shifting, $\Delta\mathscr{D}_i$, 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 A3 isolates the contribution of markup changes to the evolution of $\Delta\mathscr{D}_i$. It plots $\Delta\mathscr{D}_i$ for each year holding trade shares constant at their 2005 level. The evolution of$\Delta\mathscr{D}_i$ in the middle panel, as a result, merely reflects the change in markups over time. The bottom panel in Figure A3 isolates the contribution of changes in trade shares. It plots $\Delta\mathscr{D}_i$ for each year, holding markups constant at their 2005 level. As such, the evolution of $\Delta\mathscr{D}_i$ in the bottom panel merely reflects the impact of changing production specialization over time.

Figure A3: The drivers of rent-shifting patterns over time




*Note*: The above graph reports the trade-induced change in the deadweight loss of markups, $\Delta\mathscr{D}$. 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 $\Delta\mathscr{D}$ holding trade shares constant at their 2005 level. The bottom panel computes $\Delta\mathscr{D}$ holding markups constant at their 2005 level. In each case $\Delta\mathscr{D}$ is computed using the formula under Proposition 2. Data on industry-level expenditure, production and input-output shares are from the ICIO.

Comparing the three cases in Figure A3 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.

## N Quantitative Strategy for Calculating Implicit Tariffs

This appendix outlines our method to measure the implicit tariffs, $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, that replicate rent-shifting effects using observable data and our estimated markup values. We start by describing the equilibrium relationships that govern model outcomes. It's crucial to note that while our previous welfare calculations did not require trade elasticity data, we need this information to estimate the implicit tariffs. For a given vector of tariffs, $\{t_{in,k}\}$, the industry-level expenditure shares in our semi-parametric model are:

$$\lambda_{in,k} = \frac{\chi_{in,k} w_i^{-\theta_k} \left(1 + t_{in,k}\right)^{-\theta_k}}{\sum_{j=1}^{N} \chi_{jn,k} w_j^{-\theta_k} \left(1 + t_{jn,k}\right)^{-\theta_k}},$$

where $\chi_{in,k}$ is a constant that includes iceberg trade costs and policy-invariant technology parameters; $\theta_k$ is the trade elasticity in industry $k$, which equals the shape of the Pareto firm productivity distribution in our semi-parametric model, assuming tariffs are applied before markups and act as a cost shifter. Markups do not appear in the equation above because the markup distribution is the same across countries. The labor market clearing condition states that each country's wage bill equals the input cost, calculated as the value of sales minus markups and tariffs:

$$w_i L_i = \sum_k \sum_n \left[ \frac{1}{\mathbb{E}_{\rho_k}\left[\mu\right]\left(1 + t_{in,k}\right)} \lambda_{in,k} e_{n,k} E_n \right],$$

where $e_{n,k}$ is the constant expenditure share on industry $k$ in country $n$, based on the Cobb-Douglas assumption. A country's total expenditure, $E_n$, is equal to its earned income, which includes wage income, profits, and tariff revenues. The national-level budget constraint representing this condition can be written as:

$$E_i = \widetilde{\mathbb{E}}_{y_i}\left[\mu\right] w_i L_i + \sum_k \sum_n \left[ \frac{t_{in,k}}{1 + t_{in,k}} \lambda_{in,k} e_{n,k} E_n \right],$$

where $\widetilde{\mathbb{E}}_{y_i}\left[\mu\right]$ represents country $i$'s average profit margin and $\widetilde{\mathbb{E}}_{y_i}\left[\mu\right] w_i L_i$ is wage income adjusted for profit rebates. The second term on the right-hand side represents country $i$'s tariff revenue.

We want to investigate the equilibrium responses when transitioning from the factual markups and tariffs, denoted as $(\mathbf{t}, \boldsymbol{\mu})$, to a counterfactual scenario, $(\mathbf{t} + \tilde{\mathbf{t}}, \mathbf{1})$, where markups are replaced by an implicit tariff $\tilde{\mathbf{t}}$. To express the changes between the factual and counterfactual equilibria, we employ the hat-algebra notation, whereby $\hat{x}$ denotes the change in a generic variable $x$. For simplicity, we assume that applied tariffs are close to zero in the status quo, i.e., $\mathbf{t} \approx \mathbf{0}$, which aligns with real-world data.

In the counterfactual equilibrium, the labor market clearing condition ensures that wage payments in country $i$ equal sales net of only tariffs, since markups are zero. This condition can be expressed as:

$$\hat{w}_i w_i L_i = \sum_k \sum_n \left[ \frac{1}{\left(1 + \tilde{t}_n\right)} \frac{\lambda_{in,k} \left(1 + \tilde{t}_n\right)^{-\theta_k} \hat{w}_i^{-\theta_k}}{\lambda_{nn,k} \hat{w}_n^{-\theta_k} + \sum_{j \neq n} \lambda_{jn,k} \left(1 + \tilde{t}_n\right)^{-\theta_k} \hat{w}_j^{-\theta_k}} e_{n,k} \hat{E}_n E_n \right]. \tag{28}$$

The above formulation uses the fact that for any variable like the wage rate, the counterfactual value can be specified as $w_i' = \hat{w}_i w_i$. The constant elasticity import demand structure implies $\hat{\lambda}_{in,k} = \left(1 + \tilde{t}_n\right)^{-\theta_k} \hat{w}_i^{-\theta_k} \hat{P}_{n,k}^{-\theta_k}$, with the change in the consumer price index given by $\hat{P}_{n,k}^{-\theta_k} = \Big(\lambda_{nn,k} \hat{w}_n^{-\theta_k} + \sum_{j \neq n} \lambda_{jn,k} (1 + \tilde{t}_n$
The national budget constraint requires that total income equal wage payments plus tariff revenues in the

counterfactual equilibrium, as there are no markups or rents by assumption ($\Pi_i' = 0$). This condition can be formulated as:

$$\hat{E}_i E_i = \frac{\hat{w}_i w_i L_i}{1 - \sum_k \sum_{n \neq i} \left[ \frac{\left(1 - \frac{1}{1+\tilde{t}_i}\right) \lambda_{ni,k} (1+\tilde{t}_i)^{-\theta_k} \hat{w}_n^{-\theta_k}}{\lambda_{ii,k} \hat{w}_i^{-\theta_k} + \sum_{j \neq i} \lambda_{ji,k} (1+\tilde{t}_i)^{-\theta_k} \hat{w}_j^{-\theta_k}} e_{i,k} \right]}. \tag{29}$$

The welfare-neutrality condition ensures that replacing factual markups $\boldsymbol{\mu}$ with implicit tariffs $\tilde{\mathbf{t}}$, maintains national welfare in every country. This condition is expressed as:

$$\hat{W}_i = \frac{\hat{Y}_i}{\hat{P}_i} = 1, \qquad \text{where} \qquad \hat{P}_i = \prod_k \left[ \lambda_{ii,k} \hat{w}_i^{-\theta_k} + \sum_{n \neq i} \lambda_{ni,k} (1 + \tilde{t}_i)^{-\theta_k} \hat{w}_n^{-\theta_k} \right]^{-\frac{e_{i,k}}{\theta_k}}. \tag{30}$$

In summary, Equations (28)-(30) form a system of $3N$ independent equations with $3N$ unknowns: $\{\hat{w}_i, \hat{Y}_i, \tilde{t}_i\}$. By solving this system, we can recover the vector of hidden tariffs, $\tilde{\mathbf{t}} = \{\tilde{t}_1, ..., \tilde{t}_N\}$, which replicates the international rent-shifting effects associated with markups $\boldsymbol{\mu}$. This task, moreover, requires information on only observables, markups, and trade elasticities, $\mathbf{X} = \{\lambda_{ni,k}, e_{i,k}, Y_i, w_i L_i, \theta_k, \mathbb{E}_{\rho_k}[\mu]\}$.

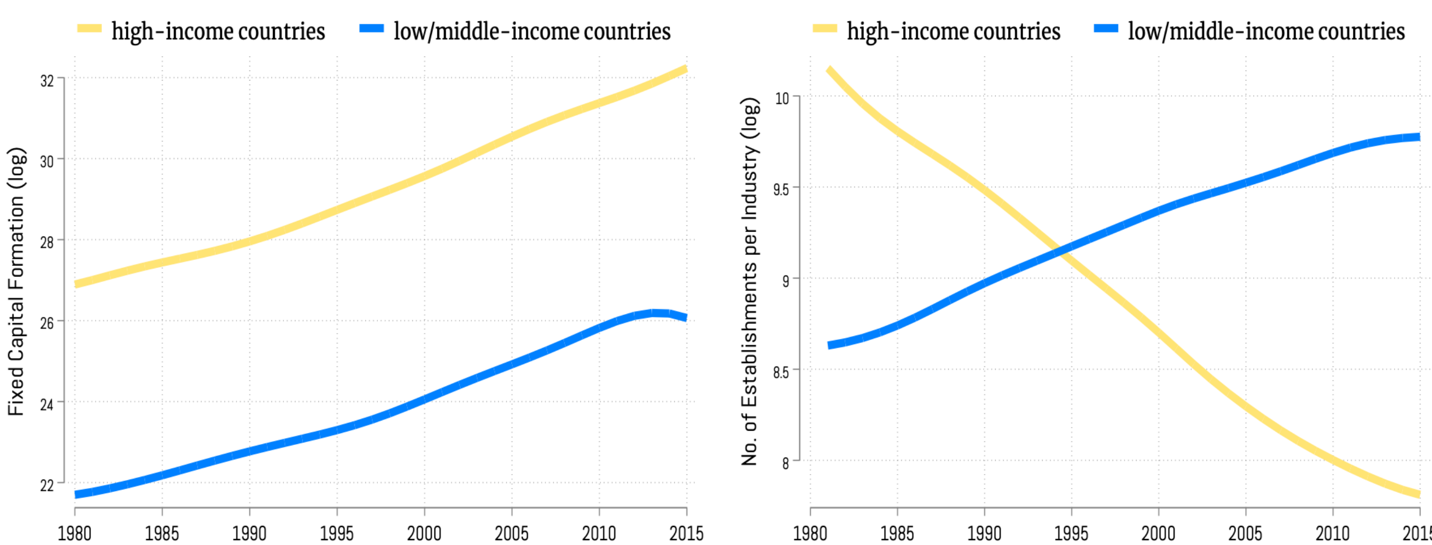
# O Additional Tables and Graph

Table A4: 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 | Chinese Taipe | | | |

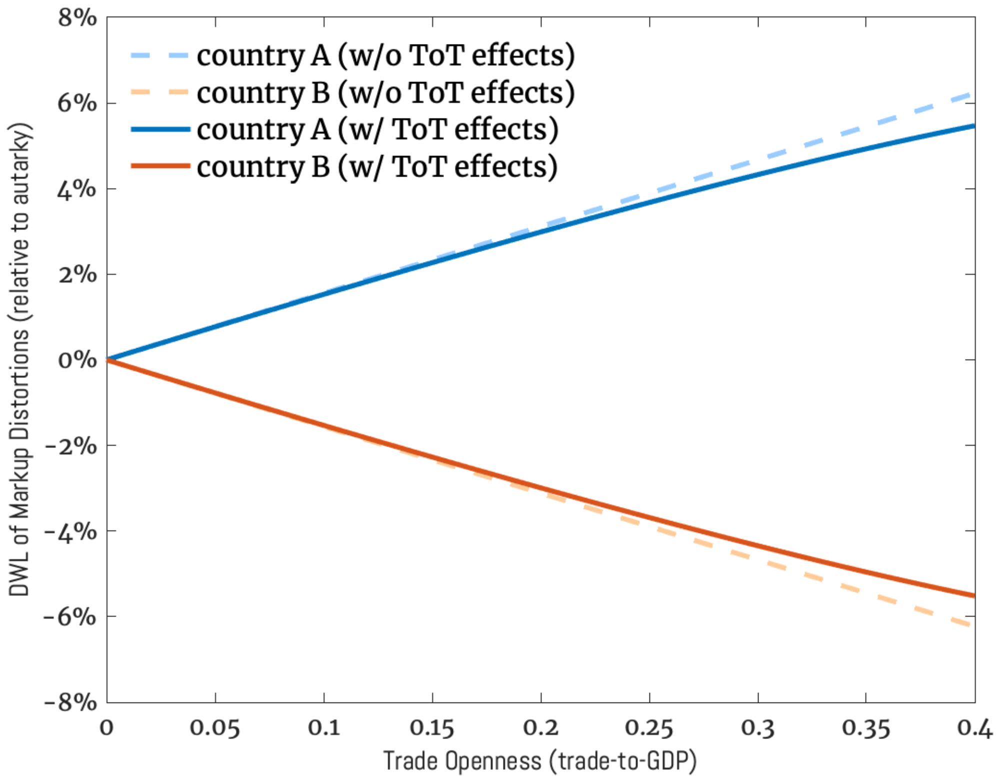
*Note*: The classification by income group is based on the UNITED NATIONS COUNTRY CLASSIFICATION.

Figure A4: Trends in fixed capital formation and firm entry by income group



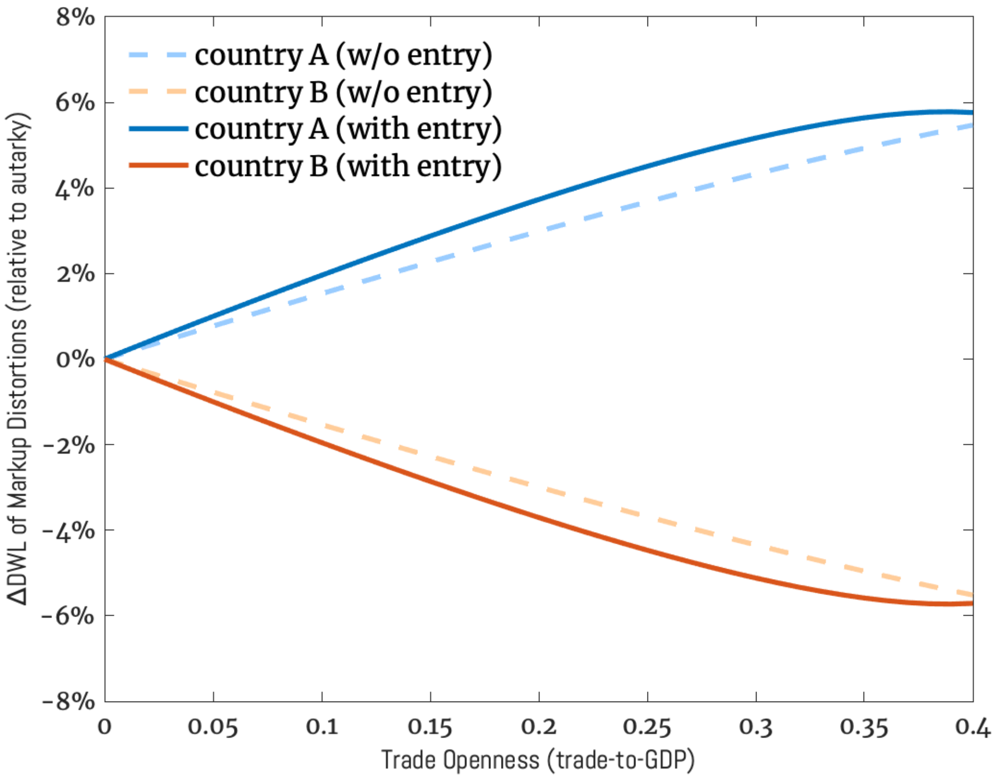

*Note*: the data is from UNIDO-INDSTAT covering 196 countries and 23 ISIC rev.3. The top panel demonstrates that the rate of fixed capital formation (which includes R&D investment) has not diverged across high and low-income countries in the same way aggregate profits margins have. The bottom panel demonstrates that diverging profit margins between low and high-income countries do not coincide with an increased rate firm entry in high-income countries. High-income countries are defined as those in top third of the GDP per capita distribution. Low and middle income countries are classified as those in the bottom two-thirds of the distribution.

Figure A5: Zero-sum rent-shifting effects under factoral terms of trade adjustments



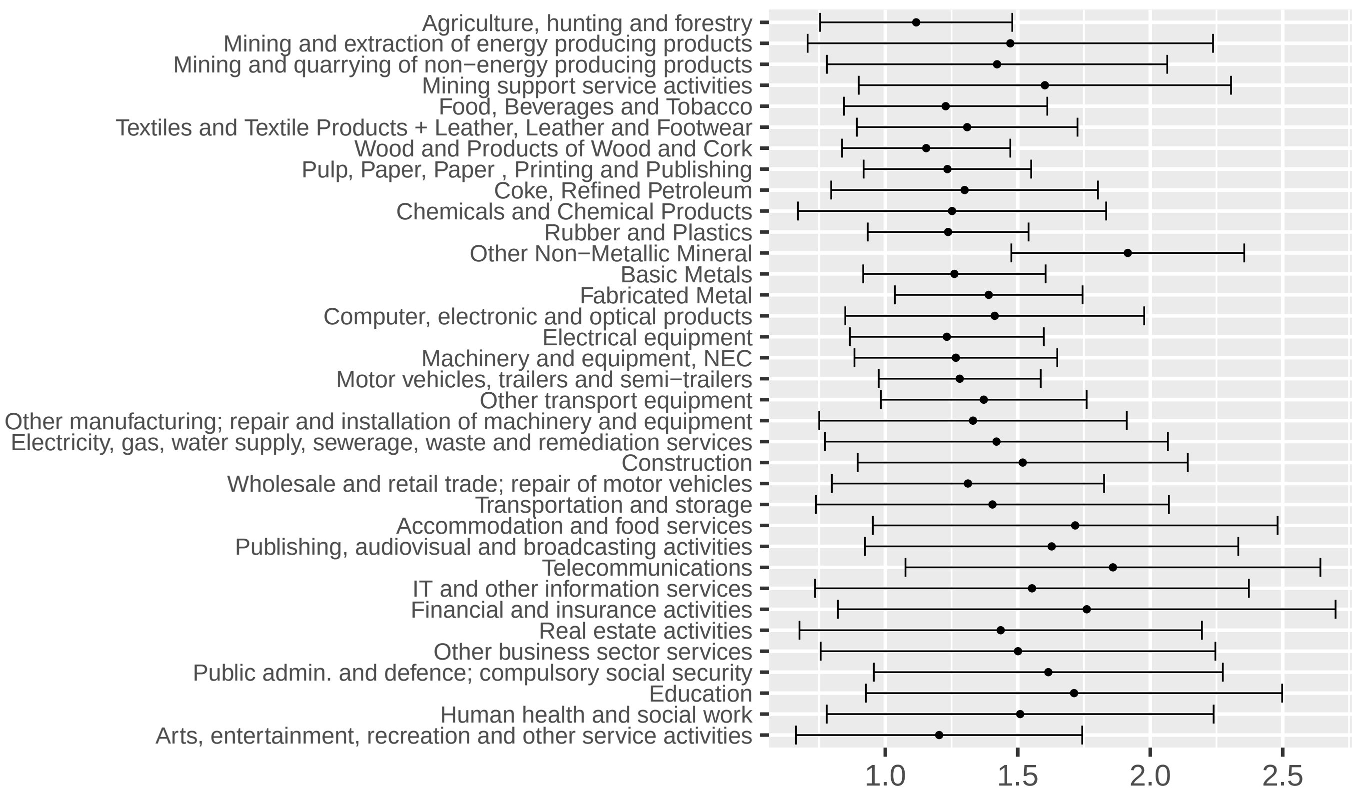

*Note*: This figure demonstrates that the factoral terms of trade (ToT) effects associated with markup distortions are relatively small compared to the direct welfare loss caused by these distortions. Specifically, the figure illustrates that the deadweight loss resulting from markup distortions is nearly identical whether we consider the associated wage effects (represented by the solid line) or measure the deadweight loss without accounting for wage effects (represented by the dashed line). The figure is generated through numerical simulations of a two-sector Krugman model with two countries, $A$ and $B$. Sector 2 has a higher markup ($\mu_2 = 1.5$) compared to Sector 1 ($\mu_1 = 1.1$). Country $A$ has a revealed comparative advantage in the high-markup sector, with $\lambda_{AB,2} = 0.75$, $\lambda_{AB,1} = 0.25$ , while Country $B$ has a revealed comparative advantage in the low-markup sector, with $\lambda_{BA,2} = 0.25$, $\lambda_{BA,1} = 0.75$. Both countries have equal population sizes ($L_A = L_B$) and the same Cobb-Douglas expenditure shares across sectors ($e_A = e_B = 1/2$). The trade elasticities in each sector are consistent with the underlying markup, as per the Krugman model: $\theta_1 = 2$ and $\theta_2 = 10$. In this numerical example, Country $A$'s pattern of revealed comparative advantage aligns with that of high-income countries, while Country $B$'s revealed comparative advantage is similar to that of low-income countries, as manifested in real-world data.

Figure A6: The internationally zero-sum welfare consequences of markup distortions under free entry



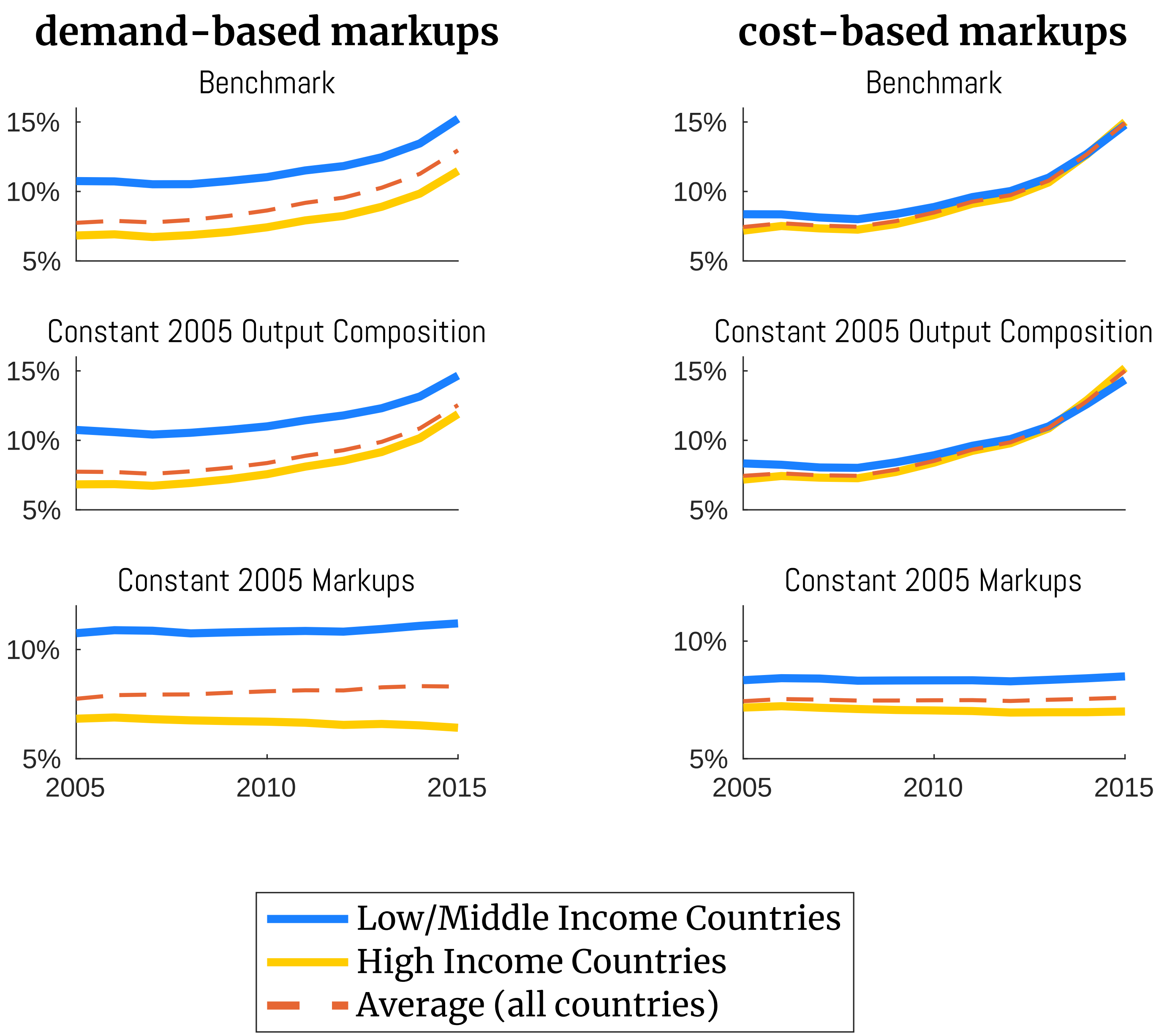

*Note*: This figure demonstrates that the zero-sum welfare effects of trade in markup-distorted economies persist even under free entry. The lines represent the change in the deadweight loss (DWL) of markups as a function of increased trade openness, which is measured by a higher trade-to-GDP ratio. This solid line corresponds to a scenario with free entry. In contrast, the dashed line traces the change in the deadweight loss under restricted entry conditions, which is our baseline specification. The figure is generated through numerical simulations of a two-sector Krugman model with two countries, $A$ and $B$. Sector 2 has a higher markup ($\mu_2 = 1.5$) compared to Sector 1 ($\mu_1 = 1.1$). Country $A$ has a revealed comparative advantage in the high-markup sector, with $\lambda_{AB,2} = 0.75$, $\lambda_{AB,1} = 0.25$ , while Country $B$ has a revealed comparative advantage in the low-markup sector, with $\lambda_{BA,2} = 0.25$, $\lambda_{BA,1} = 0.75$. Both countries have equal population sizes ($L_A = L_B$) and the same Cobb-Douglas expenditure shares across sectors ($e_A = e_B = 1/2$). The trade elasticities in each sector are consistent with the underlying markup, as per the Krugman model: $\theta_1 = 2$ and $\theta_2 = 10$. In this numerical example, Country $A$'s pattern of revealed comparative advantage aligns with that of high-income countries, while Country $B$'s revealed comparative advantage is similar to that of low-income countries, as manifested in real-world data.

Figure A7: Distributions of firm-level markups across industries




*Note*: The figure shows the within-industry distribution of firm-level cost-based 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.

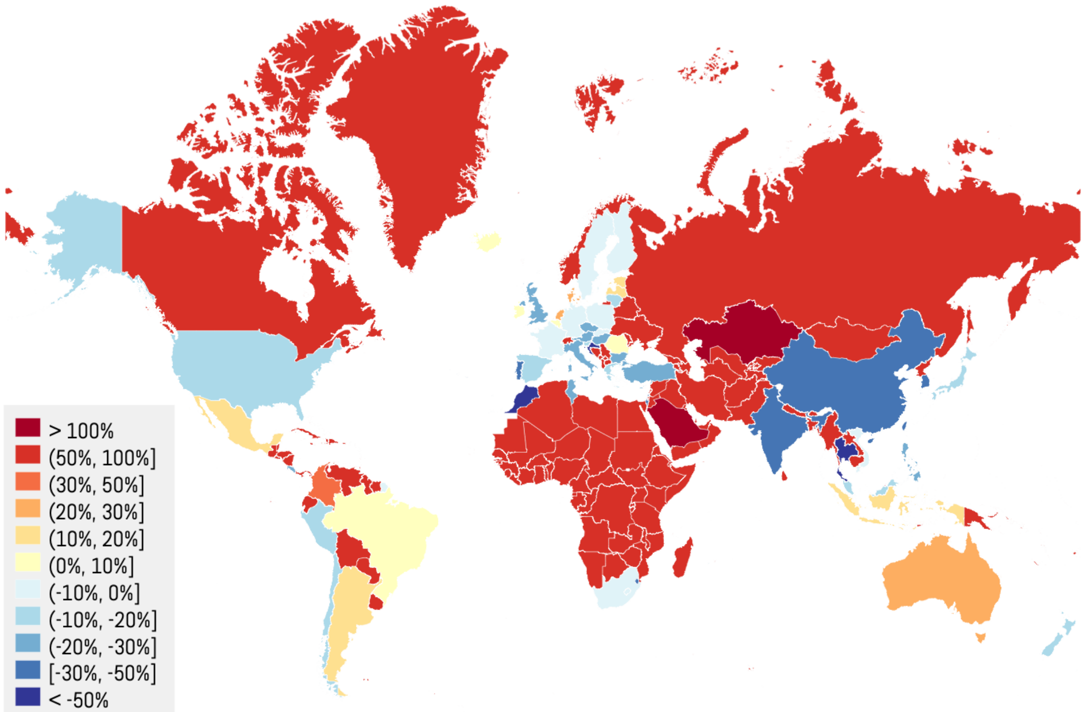
Figure A8: The deadweight loss markups adjusted for IO linkages, and its drivers over time



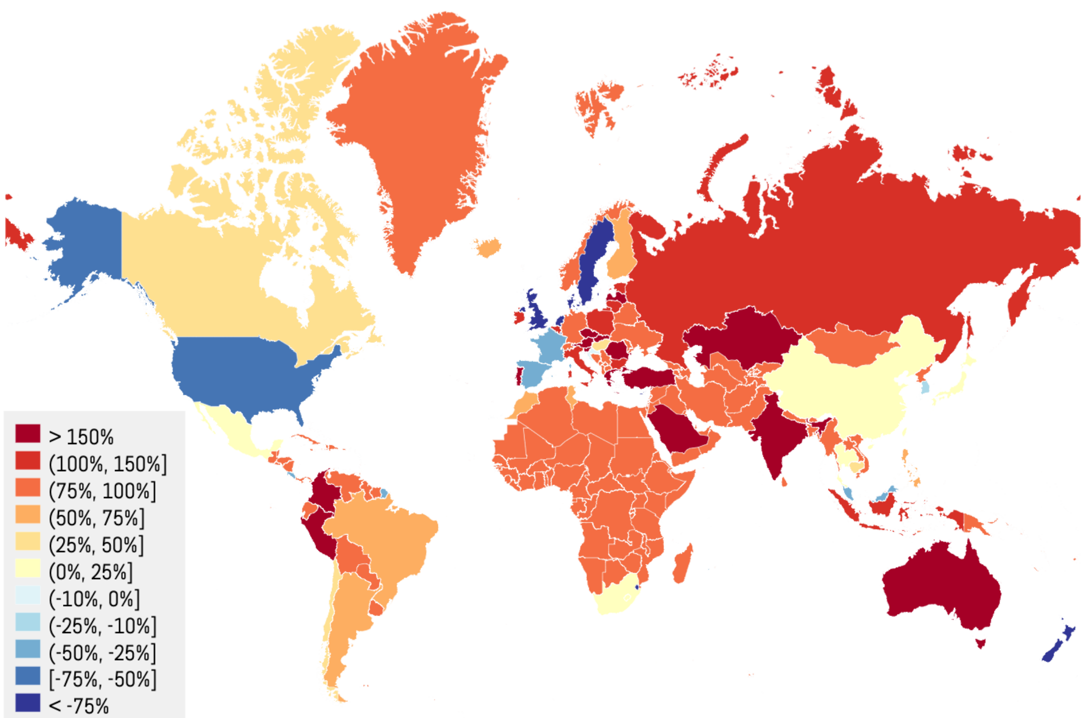

*Note*: The above graph reports the deadweight loss (DWL) of markups for averaged across the various countries. For example, a 5% DWL implies that markups lower real consumption by 5% relative to the efficient level. The figures in the top panel are computed using Lemma 3 (which accounts for input-output 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.

Figure A9: Trade-Induced change in deadweight loss of markups (exhibit A)

**Baseline Model**

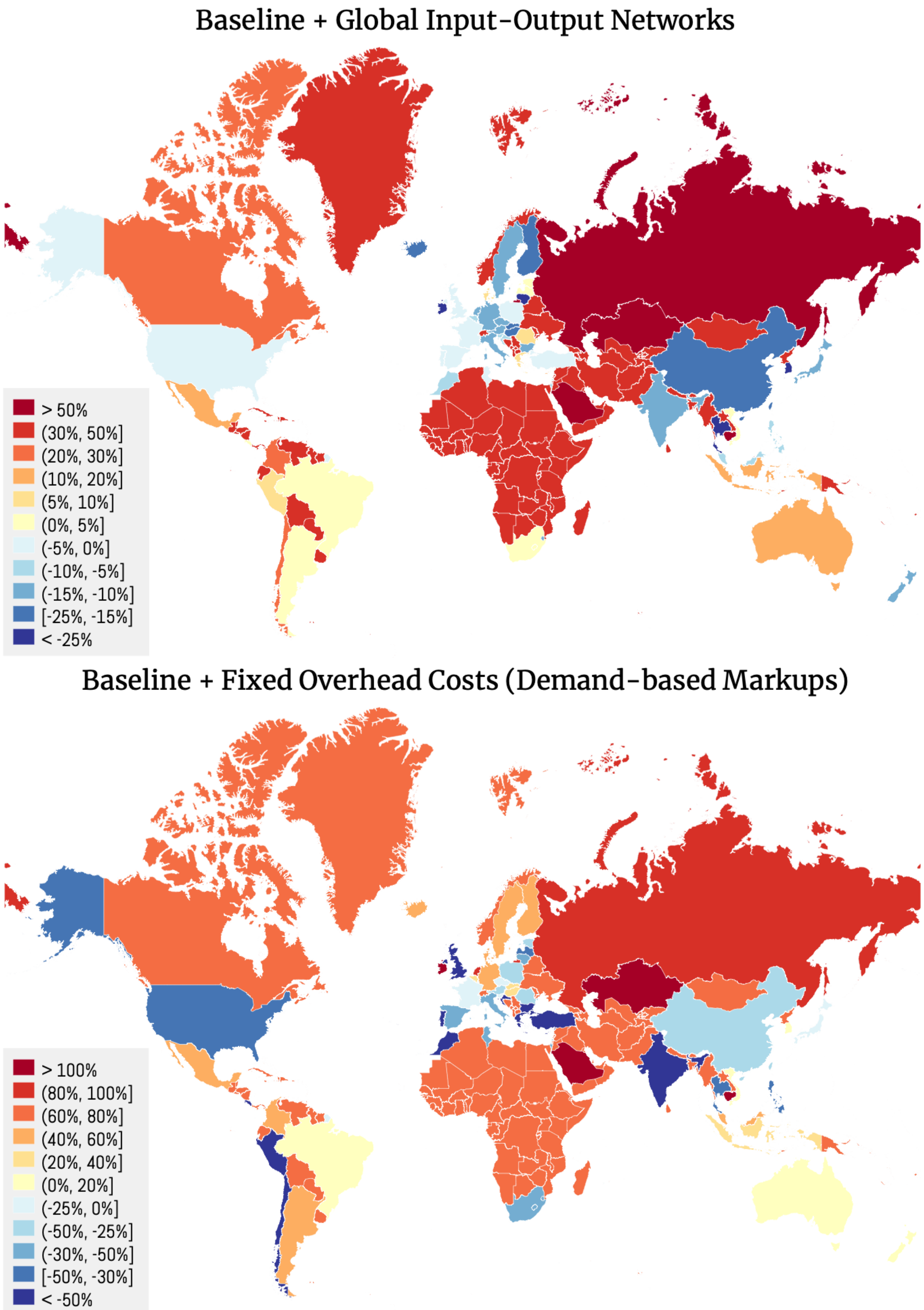



**Baseline + Multinatoinal Ownership**



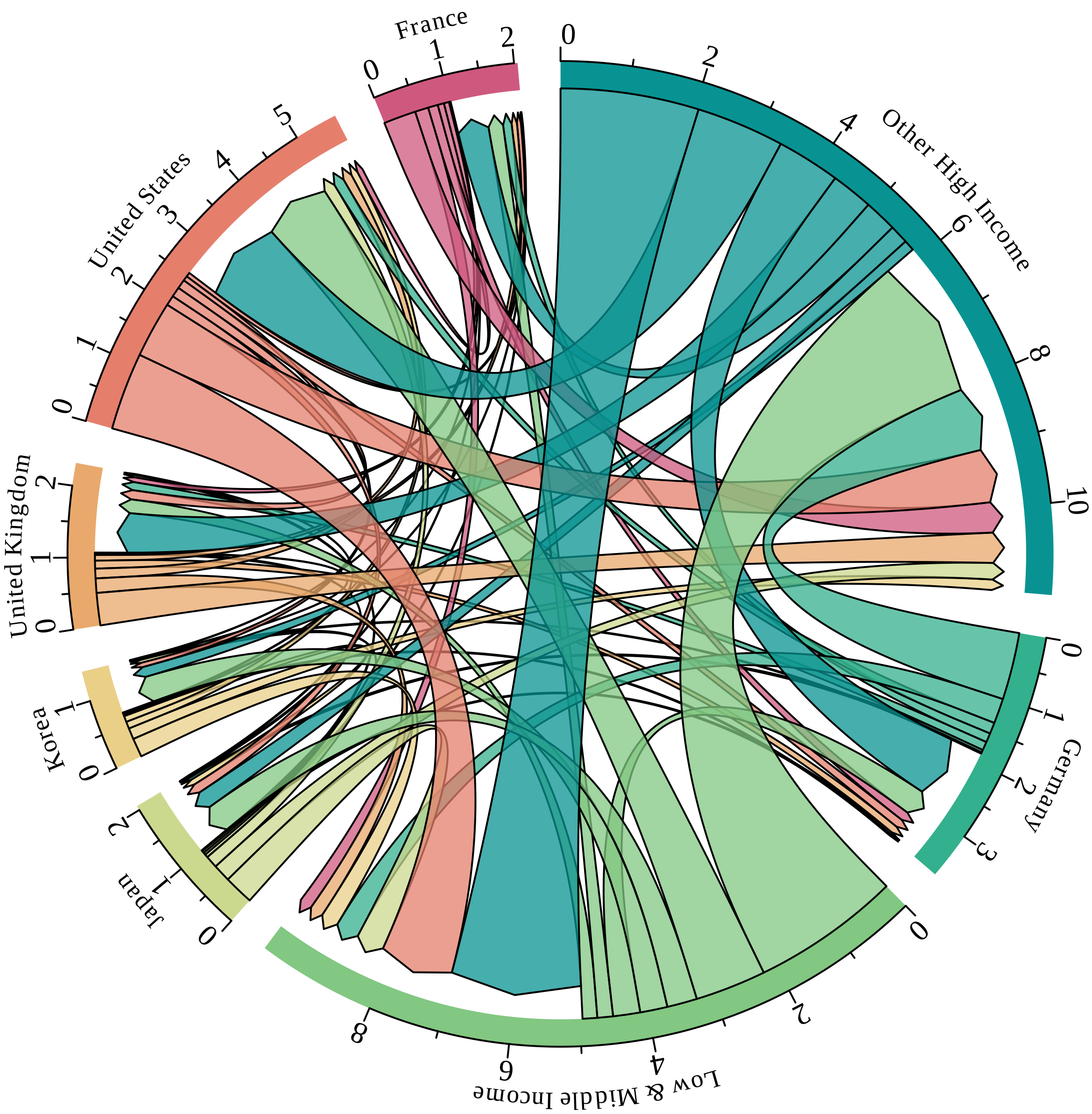

*Note*: The map illustrates the unequal exposure to rent-shifting externalities among high and low-income country groups. It displays the per-cent change in the deadweight loss of markups due to international rent-sifting ($\triangle\mathscr{D}_i$) on a country by country basis. The reported change in deadweight losses are calculated using Proposition 2 and represent the average effect implied by the demand-based and cost-based markup estimates. Data on national and industry-level output and expenditure shares are from ICIO in 2015.

Figure A10: Trade-Induced change in deadweight loss of markups (exhibit B)



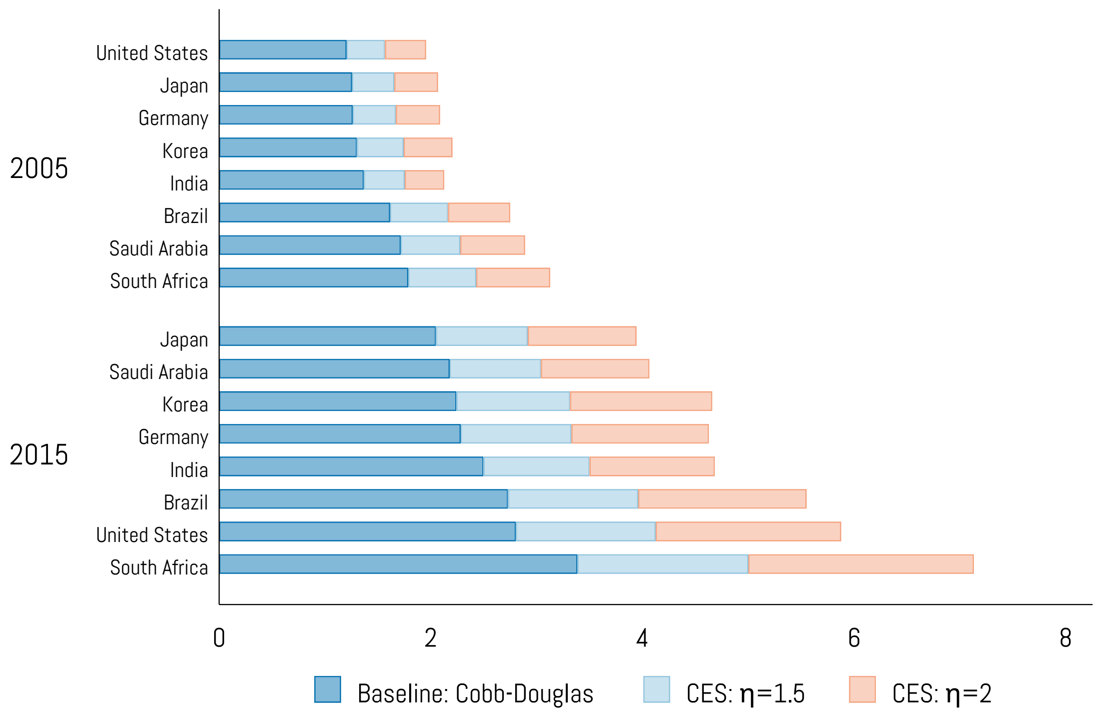

*Note*: The map illustrates the heterogeneous exposure to rent-shifting externalities within high and low-income country groups. It displays the per-cent change in the deadweight loss of markups due to international rent-sifting ($\triangle \mathscr{D}_i$) on a country by country basis. The reported change in deadweight losses are calculated using Proposition 2 and represent the average effect implied by the demand-based and cost-based markup estimates. The markups levels are taken from our demand-based estimation. Data on national and industry-level output and expenditure shares are from ICIO in 2015.

Figure A11: International patterns of rent-shifting



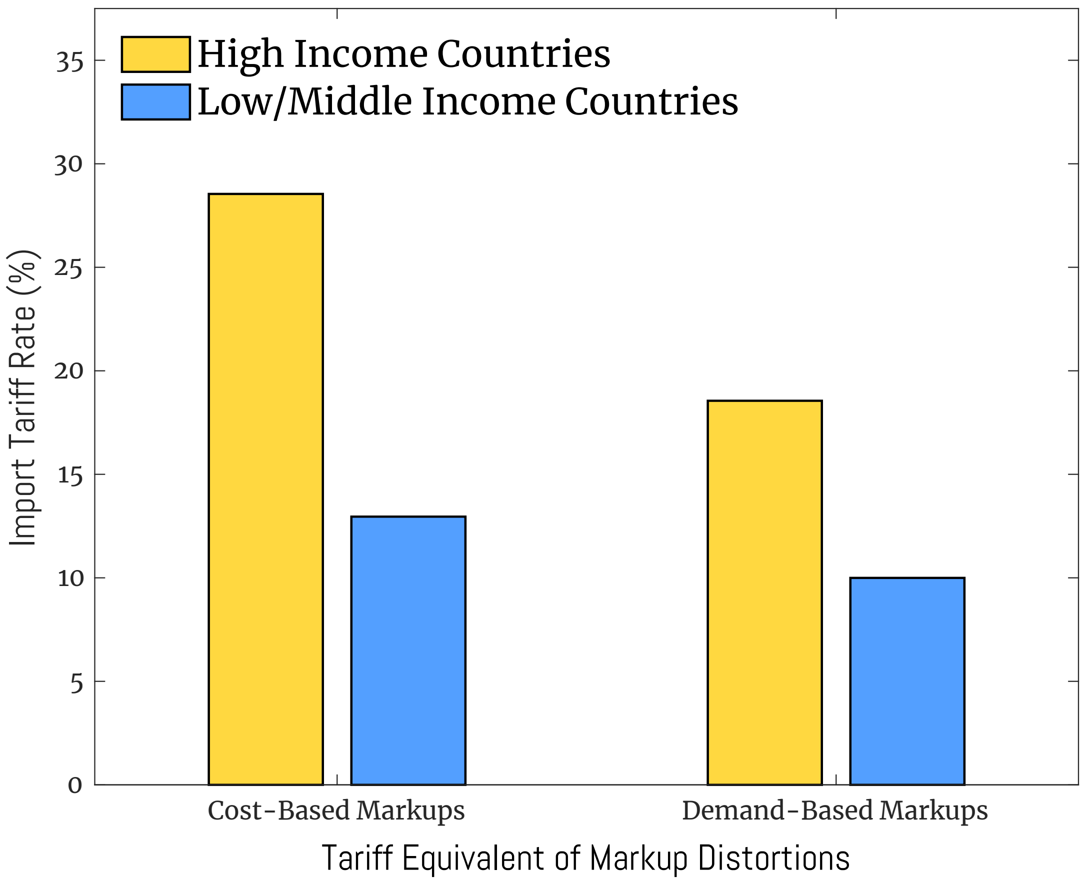

*Note*: An arrow from country *i* to *j* corresponds to the profits or rents collected by firms in country *j* from consumers in country *i* as a percentage of the profits collected by domestic firms. Profits or rents are calculated using sales data from Icio and markups estimated by applying the cost-based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

Figure A12: The deadweight loss of markups under different levels of cross-industry substitutability




*Note*: The above graph shows how a higher substitution elasticity between industries amplifies the deadweight loss (DWL) of markups. Regarding units: a 5% DWL implies that markups lower real consumption by 5% relative to the efficient level. Data on industry-level expenditure, trade, production, and input-output shares are from the ICIO. Data on markups are based on our demand-based markup estimates.

Figure A13: Strong duality between import tariffs and monopolistic markups




*Note*: This figure illustrate strong duality between import tariffs and markups as defined in Section 7. It reports the tariffs that exactly replicate the welfare effects of monopolistic markups by country group. The tariff rates for each country group are the GDP-weight average rates and calculated using a similar algorithm to that described in Appendix N. Our calculation uses expenditure and output data from the ICIO TABLES and our estimated markups (described in Section (8)) in year 2015. Industry-level trade elasticity values are from Caliendo and Parro (2015).