DISCUSSION PAPER SERIES

DP18089

MARKET VERSUS OPTIMUM DIVERSITY IN OPEN ECONOMIES: THEORY AND QUANTITATIVE EVIDENCE

Peter Egger and Ruobing Huang

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

CEPR

ISSN 0265-8003

MARKET VERSUS OPTIMUM DIVERSITY IN OPEN ECONOMIES: THEORY AND QUANTITATIVE EVIDENCE

Peter Egger and Ruobing Huang

Discussion Paper DP18089
Published 18 April 2023
Submitted 13 April 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

MARKET VERSUS OPTIMUM DIVERSITY IN OPEN ECONOMIES: THEORY AND QUANTITATIVE EVIDENCE

Abstract

The inefficiency of the market under monopolistic competition and heterogeneous firms in closed economies is well established. However, insights regarding asymmetric countries which are open to trade are less developed. This paper offers a systematic analysis of this question using general demand structures and productivity distributions. First, we derive sufficient conditions that constrain the space of exogenous parameters, ensuring the existence and uniqueness of general equilibrium. Next, we demonstrate that distortions from a global social planner's perspective in open economies can be described as shift effects on production schedules, resulting from differences in the fundamentals of asymmetric countries, and rotation effects on production schedules, determined by the demand structure. We apply these insights to quantify the distortions in terms of cutoffs, quantities, and welfare for China and the Rest of the World. Our results indicate that both economies have weak selections, with relatively more overproducing firms selling to China and more underproducing firms selling abroad. Additionally, we find that in some cases, China attains higher welfare levels under the market than the global social optimum.

JEL Classification: F12, L11, D61

Keywords: Misallocation, Heterogeneous firms, Trade

Peter Egger - pegger@ethz.ch
Eth Zurich and CEPR

Ruobing Huang - ruhuang@ethz.ch
Eth Zurich

Market versus Optimum Diversity in Open Economies: Theory and Quantitative Evidence*

Peter H. Egger
ETH Zürich, CEPR, CESifo

Ruobing Huang †
ETH Zürich

April 12, 2023

Abstract

The inefficiency of the market under monopolistic competition and heterogeneous firms in closed economies is well established. However, insights regarding asymmetric countries which are open to trade are less developed. This paper offers a systematic analysis of this question using general demand structures and productivity distributions. First, we derive sufficient conditions that constrain the space of exogenous parameters, ensuring the existence and uniqueness of general equilibrium. Next, we demonstrate that distortions from a global social planner's perspective in open economies can be described as shift effects on production schedules, resulting from differences in the fundamentals of asymmetric countries, and rotation effects on production schedules, determined by the demand structure. We apply these insights to quantify the distortions in terms of cutoffs, quantities, and welfare for China and the Rest of the World. Our results indicate that both economies have weak selections, with relatively more overproducing firms selling to China and more underproducing firms selling abroad. Additionally, we find that in some cases, China attains higher welfare levels under the market than the global social optimum.

Keywords: Misallocation; Heterogeneous firms; Trade
JEL Codes: F12, L11, D61

*The authors gratefully acknowledge Carsten Eckel, Katharina Erhardt, Isabelle Méjean, Matthieu Parenti, John Morrow, Gianmarco Ottaviano, Jens Südekum, and participants at the European Trade Study Group 2022, the DICE Research Seminar (at University of Düsseldorf), the Research Workshop on International Trade and Multilateralism, and the 24th Göttinger Workshop for helpful comments. Huang gratefully acknowledges financial support from the 2022 MTEC foundation at ETH Zürich.

†Corresponding author. E-mail addresses: pegger@ethz.ch (P.H. Egger), ruhuang@ethz.ch (R.Huang).

1 Introduction

That markets provide effectively for an efficient outcome through an invisible hand − eventually even ensuring an optimal allocation − is a leading assumption of many economic models and economists. Such optimality is increasingly called into question. Recent academic work pointed to the fact that an optimal allocation is far from being guaranteed: in the closed economy it will only materialize under very restrictive model assumptions. That is, firms generally overproduce or underproduce compared to the optimal quantity from a social planner's perspective.

While the mechanism behind the potential misallocation of the market is well understood in a closed economy, it is still unclear with international trade among asymmetric countries. One key issue of the closed economy is that utility-maximizing consumers, profit-maximizing firms, and the social planner all have the same "address." The latter is fundamentally different for open economies: firms maximize profits given the resources they have at their domestic production site, but they face a consumer base located in different countries. In other words, what firms decide in the open economy affects not only their consumer home base but also the foreign one. Similarly, a global social planner who cares about the aggregate global welfare faces a customer base that does not have a single national "address." In substance, changing firms' decisions about output and pricing in one market induces externalities on the customers in other markets.

This paper addresses two questions in this context: (i) what are the determinants of market distortions regarding quantities, selections, and entries in open economies, and (ii) how do they quantitatively matter? The present paper provides a systematic analysis of the potential lack of optimality − also referred to as misallocation − of decentralized market outcome with general variable-elasticity-of-substitution (VES) preferences and general productivity distributions under monopolistic competition among the firms. We focus on a setup with two asymmetric countries as this is sufficient to gain key insights.

Before delving into the analysis of market outcome distortions, it is necessary to establish the existence and uniqueness of a market equilibrium. Proving such properties with country asymmetry regarding trade costs, fixed market entry costs, sunk costs, and market size under general demand and productivity distribution structures is challenging. To overcome this issue, we introduce two implicit conditions that jointly restrict the space of all exogenous parameters, ensuring the existence and uniqueness of the market general equilibrium. These conditions are applicable to a broad range of demand structures and productivity distributions under mild assumptions.

We present several insights into the misallocation in open economies. The distortions can be described as shift and rotation effects on the quantity schedules as a function of firm productivity for each country pair. The shift effects arise from the difference in destination-specific marginal utility of income, which depends on the fundamentals of countries. We demonstrate that, at the realization of a decentralized market equilibrium, the heterogeneous consumer-based marginal utility of income across countries generally leads to a heterogeneous firm-based marginal real revenue of resources across countries, and the respective demand shifters can be ranked clearly. This ranking incentivizes the global social planner to shift the quantity schedules of each country for the domestic and the foreign consumer market in opposite ways and to adjust more significantly the production for exports than the domestic market. The existence of such shift effects is independent of the demand structure, indicating that a market misallocation may occur even with a constant-elasticity-of-substitution (CES) demand.

The rotation effects originate from the difference in markup strategies between firms and the social planner. As demonstrated in Dhingra and Morrow (2019), the social planner rotates the quantity schedule as a function of productivity either clockwise or counterclockwise. We show that the direction of rotation is purely determined by the demand structure and is independent of the fundamentals of countries.

What is key is the combined impact of the shift and rotation effects on misallocation regarding cutoffs and production schedules. The interaction of these effects leads to elusive conclusions, as the shift effects can offset and potentially outweigh the rotation effects, at least for a set of the firms from a given origin serving a given destination.

To quantify the overall and relative importance of these effects, we use a parameterized version of the model and analyze data for China and the Rest of the World (RoW) from four consecutive years between 2004 and 2007. Our analysis identifies distortions in cutoff productivities, quantities (including the percentage of over- and under-producing firms), and country-level welfare. We find that selection effects are too weak for all producer-consumer country pairings, meaning that the counter-clockwise rotation effects on the quantity schedules as a function of firm productivity dominate the shift effects. However, the shift effects do induce a destination-specific heterogeneity of the distortions regarding cutoff productivities, production schedules, and welfare. For example, in 2006, ignoring the shift effects generates absolute biases for cutoff-productivity distortions ranging from -5.48% to 4.38%, and for country-level welfare distortions ranging from -0.53% to 7.85%. Interestingly, we find that China's customers fared better with the market outcome relative to the global optimum in the first half of the sample period.

With this research agenda, the present paper builds on the following earlier work analysing allocational efficiency in closed versus open economies under monopolistic competition and heterogeneous firms as in Melitz (2003). One strand of work focuses on the closed economy. One of the first systematic studies on this question is the one by Zhelobodko et al. (2012), who demonstrate how the relative love for variety in a Dixit and Stiglitz (1977) economy affects market outcome with heterogeneous elasticities of substitution between varieties (and, hence, with heterogeneous markups). Dhingra and Morrow (2019) show how variable elasticities of substitution (VES) among varieties − and demand − determine the misallocation of resources − in the sense of suboptimal outcome in the monopolistic market relative to a utilitarian social planner − in a single-sector economy where consumers feature additive preferences.[1] They highlight two margins of potential inefficiency: selection of the distribution of firms in terms of the cutoff productivity, only above which firms will to produce, and the allocation quantities produced across firms (in the sense of over- or under-production). In Dhingra and Morrow (2019), a demand structure based on constant-elasticity-of-substitution (CES) preferences is the only special case that leads to an equivalence between the market equilibrium and the social optimum. Behrens et al. (2020) extend the framework of Dhingra and Morrow (2019) to a multi-sector context with inter-sectoral labor mobility. In that case, the market generates inefficient selection and firm-level output, as in Dhingra and Morrow (2019), and on top of it, it generates inefficient masses of firm entrants.[2] They further quantify the sector-level misallocation with the Constant Absolute Risk Aversion (CARA) preferences and show an aggregate welfare loss of about 6%-10% of GDP for France and the United Kingdom.[3] Mrázová et al. (2021) assume Constant Revenue Elasticity of Marginal Revenue (CREMR) preferences and demand. They quantitatively compare the distributions of output in the market equilibrium with the one in a constrained social optimum in a single-sector closed economy, where the social planner can only reallocate output but not affect the cutoff productivity and firm entry.

A second strand of work focuses on the question of misallocation in open economies. Related work relies on one specific demand structure and/or relatively strong assumptions regarding the parameterization of economies such as country symmetry, costless trade, zero fixed market-entry costs, and a Pareto distribution about firm productivity. Most of the

[1]More recent research illustrates how the results of Dhingra and Morrow (2019) extend to a more general demand structure (see, e.g., Nocco et al., 2014, Bertoletti and Etro, 2021; Bagwell and Lee, 2022; Macedoni and Weinberger, 2022).

[2]Bagwell and Lee (2021) consider a two-sector model with non-additive preferences as in Melitz and Ottaviano (2008) to assess the efficiency of market entry relative to a second-best setting in which the planner can control only the entry of firms.

[3]See also Behrens and Murata (2012) and Behrens et al. (2014) for applications of CARA preferences.

related work in this vein utilizes the non-additive preference structure proposed by Melitz and Ottaviano (2008) with an outside sector, zero fixed cost, and a Pareto distribution about firm productivity for analytical tractability. Within this specific framework, Nocco et al. (2019) study market distortions (on cutoff productivities, quantities, and masses of entrants) and consider the question of a globally optimal multilateral trade policy under asymmetric countries. Bagwell and Lee (2020) study unilateral, efficient and Nash trade policies (specifically, about tariffs) between two symmetric countries with Pareto-distributed firms. Departing from the framework of Melitz and Ottaviano (2008), Arkolakis et al. (2019) compare the country-level gains from trade liberalization of gravity trade models with variable markups under a general demand structure, a Pareto firm distribution, and zero fixed costs. They derive a sufficient statistic of the welfare gains and show that gains under variable markups are no greater than those under constant markups. They further show how their results are related to the degree of misallocation at the country level. Assuming costless trade and symmetric countries, Baqaee and Farhi (2021) consider how an increase of market size (which they interpret as globalization) affects welfare and allocational efficiency. They demonstrate how allocational efficiency changes due to reallocations and emphasize a *Darwinian effect* resulting from increased firm entry, which leads to a reallocation of output among firms.[4]

Overall, a systematic analysis of both firm-level and country-level misallocations with general additive preferences and productivity distributions in asymmetric large open economies is hitherto not available. Moreover, by assuming a closed economy, symmetric countries, or outside sectors, earlier work eliminated one of the key mechanisms behind the misallocation problem for asymmetric open economies, namely the shift effect on quantity schedules. The latter arises from the heterogeneous consumer-based marginal utility of income across asymmetric countries.

The remainder of the paper is organized as follows. Section 2 outlines the general theoretical framework, and Sections 3 and 4 provide CES and VES examples, respectively. The section 5 carries out the quantification and analysis of VES example. The last section concludes.

2 Model: General Open Economy

In this section, we outline a model of heterogeneous firms, each of which produces a unique variety of a product, where the elasticity of substitution between those varieties is variable

[4] Based on the same assumptions, Dhingra and Morrow (2019) decompose the gains from market integration under heterogeneous firms compared to homogeneous firms.

as in Dhingra and Morrow (2019) and Behrens et al. (2020), and markups are variable. A key difference to earlier work on the topic here is that firms supply their output to the world market which is segmented into countries, so that whether and to which degree the allocation of market outcome relative to a hypothetical social planner's choice may be inefficient is specific to a market.[5] The analysis of the fundamental determinants of this problem is at the heart of this paper's interest.

Throughout the analysis, consumers optimize their consumption subject to a budget constraint. But whether this will be considered only indirectly by profit-maximizing firms in a decentralized market equilibrium or directly by a social planner depends on the objective pursued. As in earlier work on the topic, a comparison of the respective economic outcomes such as prices, quantities, or the number of varieties supplied will be at the center stage.

Towards outlining the model, it will be instructive to consider three rather than two types of objectives: the decentralized market objective; the social planner's objective; and a centralized market objective. The latter is an intermediate case to obtain otherwise potentially impossible comparability between the objectives of profit-maximizing firms and a utility-maximizing planner.

In preparation of the model outline and the respective objectives, it will be useful to introduce some general notation. Specifically, we will use indices $\{i, j\}$ to refer to countries. Whenever we use pairs of indices $\{ij\}$, the first index refers to the location of the output producer, and the second one to the consumer location. We use u, p, and q to refer to utility, price, and quantity, respectively, and we index firms by their unique productivity φ which in country i are distributed with c.d.f. $G_i(\varphi)$. Finally, we refer to the masses of potential producers by M and to the profits of producers by π.

The consumers in each country have homothetic preferences. Hence, we can portray the problem from the viewpoint of a representative consumer. Every consumer finances their expenses from an income w (depending on the country of residence), and we can refer to the masses of consumers as well as employees in a country by L. we will treat L as immobile between countries. Regarding utility, the key assumptions underlying the analysis are as follows.

Assumption 1 (Utility). *$u'(\cdot) > 0$ and $\lim_{q\to+\infty} u'(q) = 0$; $u''(\cdot) < 0$; and $u(\cdot)$ is triple continuously differentiable.*

Assumption 1 requires the utility function to be concave, ensuring consumers to love variety. For later use, we define the elasticity of utility and of marginal utility, both with

[5] As said, Dhingra and Morrow (2019) and Behrens et al. (2020) consider a closed economy or, equivalent to that, a perfectly integrated world market.

respect to the quantity consumed q, as follows:

$$\varepsilon_u(q) \equiv \frac{u'(q)q}{u(q)}, \quad r_u(q) \equiv -\frac{u''(q)q}{u'(q)}.$$

Specifically, $\frac{1}{1-r_u(q)}$ measures the private markup charged by firms in the market equilibrium, while $\frac{1}{\varepsilon_u(q)}$ measures the social markup assigned by the planner in the social optimum. More technical details will be shown in later sections. With CES preferences, $1 - r_u(q) = \varepsilon_u(q)$, $\forall q \geq 0$, while the equality does not hold for VES preferences. To establish the existence and uniqueness of equilibria, we further make the following assumption about the social markup:

Assumption 2. $\lim_{q \to 0} \varepsilon_u(q) > 0$.

Assumption 2 eliminates the case, where a social planner requires firms to sell zero quantity with an infinite markup. With L'Hôpital's rule, Assumption 2 also rules out that $\lim_{q \to 0} r_u(q) = 1$, where firms choose to sell zero quantity with an infinite markup in the market equilibria.

Note that some of these assumptions differ from the ones stated in some earlier work, such as Zhelobodko et al. (2012) and Dhingra and Morrow (2019). In particular, as in Behrens et al. (2020), we relax the Inada condition, $\lim_{q \to 0} u'(q) = +\infty$, and the assumption of bounded elasticities, $0 < r_u(q), \varepsilon_u(q) < 1$. Hence, Assumption 1 and 2 are less restrictive in terms of the utility functions covered than those adopted in some of the earlier research.

2.1 Decentralized Market Equilibrium

In what follows, we address the problem of each type of agents separately.

Consumers. Representative consumers in any country receive utility from consuming goods from home (domestically) and abroad (through imports). They have a love of variety and consume everything they can afford.

A supplier in market i with productivity φ faces demand from the representative consumer in j of $q_{ij}(\varphi)$, and she charges a price of $p_{ij}(\varphi)$. Of the M_i potential sellers in i, only $M_i \int_{\varphi_{ij}^*}^{+\infty} dG_i(\varphi) \leqslant M_i$ actually produce for/sell in j, where φ_{ij}^* is the cutoff productivity of producers in i selling in j. The latter is determined in the profit maximization of firms.

The maximization of utility by the representative customer in j can be cast in terms of the Lagrangian

$$\mathcal{L} = \underbrace{\sum_i \left[M_i \int_{\varphi^*_{ij}}^{+\infty} u(q_{ij}(\varphi)) dG_i(\varphi) \right]}_{\text{utility}} + \delta_j^{dmkt} \underbrace{\left\{ w_j - \sum_i \left[M_i \int_{\varphi^*_{ij}}^{+\infty} p_{ij}(\varphi) q_{ij}(\varphi) dG_i(\varphi) \right] \right\}}_{\text{budget constraint}},$$

where $p_{ij}(\varphi)q_{ij}(\varphi)$ is the consumer's expenditure on a single variety.

The Lagrange multiplier δ_j^{dmkt} is specific to customer market j, and it is invariant to the suppliers' location. It is the marginal utility of income and can be viewed as a demand shifter which is reflective of the competition intensity in j. Individual firms are atomistic and can neither influence δ_j^{dmkt} nor any other aggregate. However, δ_j^{dmkt} is endogenous to aggregate changes.

The first-order conditions to the above maximization problem yield

$$\frac{\partial \mathcal{L}}{\partial q_{ij}(\varphi)} \Rightarrow u'(q_{ij}(\varphi)) = \delta_j^{dmkt} p_{ij}(\varphi), \forall i, j. \quad (1)$$

Operating Firms. Operating profits are defined as revenues in excess of (variable) operating costs. A firm faces factor prices of w_i per worker, and it delivers output to a customer's door in j at (iceberg) transport costs of $\tau_{ij} \geq 1$, where j could be the same as i or not. The production plus delivery costs per unit of shipment for a firm with productivity φ are $\frac{\tau_{ij} w_i}{\varphi}$. For setting up production, firms need to finance origin-destination-specific fixed costs. For a firm based in i those are $f_{ij} w_i$ for market j. Profits are obtained when aggregating real supply as $q_{ij}(\varphi) L_j$ and subtracting the fixed costs from operating profits to obtain

$$\pi_{ij}(\varphi) = (p_{ij}(\varphi) - \frac{\tau_{ij} w_i}{\varphi}) q_{ij}(\varphi) L_j - f_{ij} w_i.$$

With equ. (1), we obtain the pricing strategy for firms under profit maximization:

$$p_{ij}(\varphi) = \frac{u'(q_{ij}(\varphi))}{\delta_j^{dmkt}} = \frac{\tau_{ij} w_i}{[1 - r_u(q_{ij}(\varphi))]\, \varphi}. \quad (2)$$

$\frac{1}{1-r_u(q_{ij}(\varphi))}$ is the private markup charged by a firm producing output $q_{ij}(\varphi)$.

Equ. (2) equates marginal costs and marginal real revenues under profit maximization. From each firm's view, δ_j^{dmkt}, a common measure of competition intensity in j, is given. We can obtain the following properties for the quantity function. With a higher intensity of competition, δ_j^{cmkt}, a firm with productivity φ earns a lower marginal revenue $\frac{(1-r_u(q_{ij}))u'(q_{ij})}{\delta_j^{dmkt}}$. Besides, the marginal cost of firms, $\frac{\tau_{ij} w_i}{\varphi}$, declines with φ and increases with

$\tau_{ij} w_i$. Overall, we obtain an implicit solution for $q_{ij}(\delta_j^{dmkt} w_i, \frac{\tau_{ij}}{\varphi})$ and henceforth use the notation of $q_{ij}(\delta_j^{dmkt} w_i, \varphi)$ for it without ambiguity.

Firm Entry and Equilibrium. As in Melitz (2003) and Chaney (2008), firms draw φ prior to deciding on whether to produce. The participation costs in the lottery in country i for all potential producers, $F_i w_i$, are sunk. Only the sufficiently productive firms able to cover $f_{ij} w_i$ will choose to operate.

The zero-cutoff-profit condition (ZCPC) determines the minimum required productivity level φ_{ij}^*, at which an operating firm in i breaks even regarding its sales to j:

$$\pi_{ij}(\delta_j^{dmkt} w_i, \varphi_{ij}^*) = \left[\frac{1}{1 - r_u(q_{ij}(\delta_j^{dmkt} w_i, \varphi_{ij}^*))} - 1 \right] \frac{\tau_{ij} w_i}{\varphi_{ij}^*} q_{ij}(\delta_j^{dmkt} w_i, \varphi_{ij}^*) L_j - f_{ij} w_i = 0. \quad (3)$$

We further define the wage-adjusted profit as $\tilde{\pi}_{ij}(\delta_j^{dmkt} w_i, \varphi) \equiv \pi_{ij}(\delta_j^{dmkt} w_i, \varphi)/w_i$ to rewrite the ZCPC as $\tilde{\pi}_{ij}(\delta_j^{dmkt} w_i, \varphi) = 0$. At any lower productivity, the profits would be negative, and at higher productivity levels, they are positive. The aggregate profits of operating firms finance the lottery participation costs of all firms, the operating and the non-operating ones, so that all productivity-lottery participants have zero expected profits (ZEP):

$$\begin{aligned} &\sum_j \Pi_{ij}(\delta_j w_i) \\ &= \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left\{ \left[\frac{1}{1 - r_u(q_{ij}(\delta_j^{dmkt} w_i, \varphi))} - 1 \right] \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\delta_j^{dmkt} w_i, \varphi) L_j - f_{ij} w_i \right\} dG_i(\varphi) = F_i w_i. \end{aligned} \quad (4)$$

Defining $\tilde{\Pi}_{ij}(\delta_j^{dmkt} w_i) \equiv \Pi_{ij}(\delta_j^{dmkt} w_i)/w_i$, we can rewrite the ZEPC as $\sum_j \tilde{\Pi}_{ij}(\delta_j^{dmkt} w_i) = F_i$.

M_i is the mass of potential entrants participating in the productivity lottery. It is determined by the resource constraint (the labor-market-clearing condition):

$$M_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\delta_j^{dmkt} w_i, \varphi) L_j + f_{ij} w_i dG_i(\varphi) \right] + F_i w_i \right\} = L_i w_i. \quad (5)$$

Finally, with a numeraire, the relative wages between countries are established by the trade-balance condition (TBC): $\forall i \neq j$,

$$\begin{aligned}
&M_i \int_{\varphi_{ij}^*}^{+\infty} \frac{1}{1-r_u(q_{ij}(\delta_j^{dmkt} w_i, \varphi))} \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\delta_j^{dmkt} w_i, \varphi) L_j dG_i(\varphi) \\
=&M_j \int_{\varphi_{ji}^*}^{+\infty} \frac{1}{1-r_u(q_{ji}(\delta_i^{dmkt} w_j, \varphi))} \frac{\tau_{ji} w_j}{\varphi} q_{ji}(\delta_i^{dmkt} w_j, \varphi) L_i dG_j(\varphi).
\end{aligned} \tag{6}$$

Existence and Uniqueness of the Decentralized Market Equilibrium. We introduce the following lemma to characterize the scenario, where firms in the country i are restricted to sell only to the country j:

Lemma 1 (**Counterfactual partial equilibrium**). *$\forall i, j = H, F$, a counterfactual partial equilibrium is the solution $\left\{\delta_{ij} w_i, \varphi_{ij}^*, q_{ij}(\varphi)\right\}$ to the following conditions:*

$$\begin{cases}
\left[1 - r_u(q_{ij}(\delta_{ij} w_i, \varphi))\right] u'(q_{ij}(\delta_{ij} w_i, \varphi)) = \dfrac{\delta_{ij} \tau_{ij} w_i}{\varphi} \\
\tilde{\pi}_{ij}(\delta_{ij} w_i, \varphi_{ij}^*) = f_{ij} \\
\tilde{\Pi}_{ij}(\delta_{ij} w_i) = F_i
\end{cases}$$

Under Assumptions 1 and 2, the solution is uniquely determined. Specifically, the solution of δ_{ij} for a given w_i is defined as $\delta_{ij}(w_i)$, which is negatively related to τ_{ij}, f_{ij}, F_i, and w_i.

Given w_i, δ_{ij} is an inverse measure of the aggregate cost for firms in i to sell to j, as it decreases with all types of production costs. We fix the home wage, w_H, as the numeraire and treat the foreign wage, w_F, as endogenous. $\delta_{HH}(w_H)$ and $\delta_{HF}(w_H)$ are then exogenous, and we refer to them as δ_{HH} and δ_{HF}. Additionally, we assume:

Assumption 3. *Define:*

$$\overline{w_F} \equiv \{w_F \mid \delta_{FF}(w_F) = \delta_{HF}\}, \ \underline{w_F} \equiv \{w_F \mid \delta_{FH}(w_F) = \delta_{HH}\},$$

and require $\overline{w_F} > \underline{w_F} > 0$.

Assumption 3 establishes an implicit constraint on the space for all exogenous parameters jointly and ensures the existence of an exogenous range for the endogenous wage rate w_F where the conditions $\delta_{FF}(w_F) > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}(w_F)$ are met. Within this range, the aggregate cost of consuming only imported goods is higher than that of consuming only domestic goods. Hence, this assumption rules out that firms solely produce for the foreign market and consumers only purchase foreign goods in equilibrium. As we will show later, this assumption guarantees the existence of a partial equilibrium with the two sides of the TBC in the positive domain.

We now move to the analysis of the ZEPC, $\sum_j \tilde{\Pi}_{ij}(\delta_j w_i) = F_i$, within the wage domain that is jointly determined by all exogenous parameters. In the proof of Lemma 1, we make the following definition to ensure all firms charge positive markups:

$$\overline{q} \equiv \min\{q \geqslant 0 \text{ s.t. } r_u(q) = 1\}, \ \overline{B} \equiv r_u(\overline{q})u'(\overline{q})\overline{q}.$$

Then, we obtain the following properties of ZEPCs in both countries.

Lemma 2. *The ZEPC for home country, $\tilde{\Pi}_{HH}(\delta_H) + \tilde{\Pi}_{HF}(\delta_F) = F_H$, behaves:*

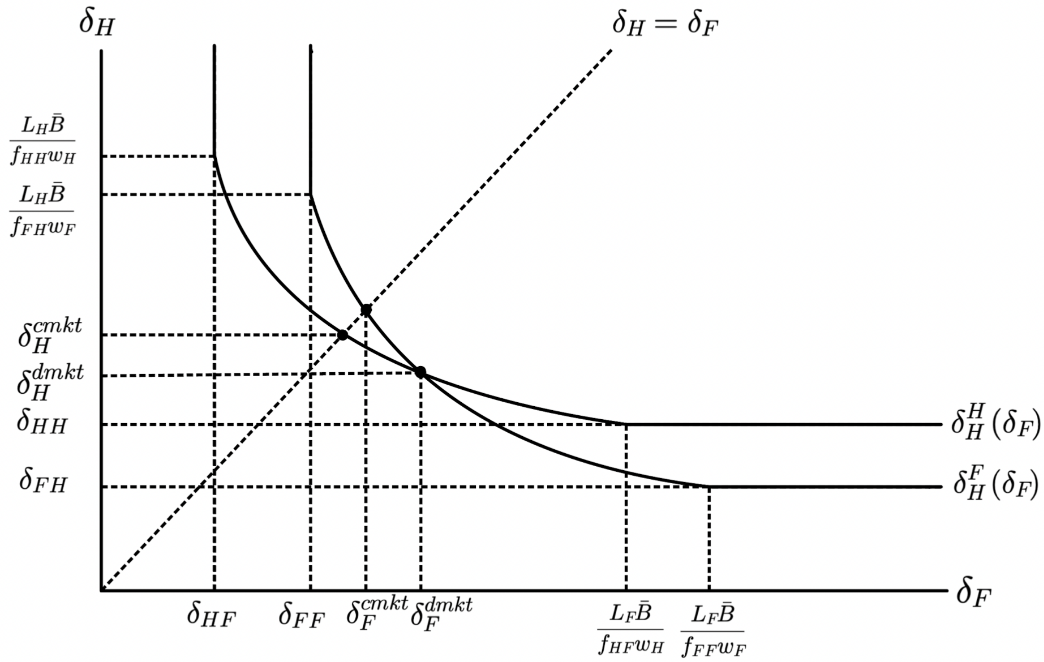
- *when δ_F decreases to δ_{HF}, δ_H increases to $\frac{L_H\overline{B}}{f_{HH}w_H}$; when δ_H decreases to δ_{HH}, δ_F increases to $\frac{L_F\overline{B}}{f_{HF}w_H}$.*
- *$\forall \delta_F \in (\delta_{HF}, \frac{L_F\overline{B}}{f_{HF}w_H})$, there exists a unique implicit function $\delta_H^H(\delta_F)$ s.t. $\tilde{\Pi}_{HH}(\delta_H^H(\delta_F)) + \tilde{\Pi}_{HF}(\delta_F) = F_H$, and $\mathrm{d}\delta_H^H(\delta_F)/\mathrm{d}\delta_F < 0$.*
- *$\forall \delta_F \in [\frac{L_F\overline{B}}{f_{HF}w_H}, +\infty)$, $\delta_H^H(\delta_F) = \delta_{HH}$.*

$\forall w_F \in (\underline{w_F}, \overline{w_F})$, the ZEPC for foreign country, $\tilde{\Pi}_{FF}(\delta_F|w_F) + \tilde{\Pi}_{FH}(\delta_H|w_F) = F_F$ behaves:

- *when δ_F decreases to $\delta_{FF}(w_F)$, δ_H increases to $\frac{L_H\overline{B}}{f_{FH}w_F}$; when δ_H decreases to $\delta_{FH}(w_F)$, δ_F increases to $\frac{L_F\overline{B}}{f_{FF}w_F}$.*
- *$\forall \delta_F \in (\delta_{FF}(w_F), \frac{L_F\overline{B}}{f_{FF}w_F})$, there exists a conditional implicit function $\delta_H^F(\delta_F|w_F)$ s.t. $\tilde{\Pi}_{FF}(\delta_F|w_F) + \tilde{\Pi}_{FH}(\delta_H^F(\delta_F|w_F)|w_F) = F_F$, and $\mathrm{d}\delta_H^F(\delta_F|w_F)/\mathrm{d}\delta_F < 0$.*
- *$\forall \delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$, $\delta_H^F(\delta_F|w_F) = \delta_{FH}(w_F)$.*

Lemma 2 and Figure 1 illustrates that, for a given wage rate, δ_H decreases with δ_F on the curve (δ_F, δ_H) that satisfies the ZEPC of home or foreign. Additionally, the value of δ_j approaches the corresponding value δ_{ij} under the counterfactual partial equilibrium, in which origin i sells exclusively to destination j. It should be noted that the curve satisfying ZEPC in either country might not be globally differentiable, as a specific demand structure could entail an upper bound on firm production. Nevertheless, from the ZEPCs of both countries, we can still express δ_H as an implicit function of δ_F when the point (δ_F, δ_H) is located within the differentiable interval of the curve. Next, we introduce another assumption that restricts the domain of parameters and guarantees the uniqueness of the market equilibrium.

Figure 1: δ_H and δ_F for a given wage.



Assumption 4. $\forall w_F \in (\underline{w_F}, \overline{w_F})$, *if* $\delta_{FF}(w_F) < \frac{L_F\overline{B}}{f_{HF}w_H}$, $\forall \delta_F \in \left(\delta_{FF}(w_F), \min\left\{\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F}\right\}\right)$*:*

$$\frac{\mathrm{d}\delta_H^F(\delta_F|w_F)}{\mathrm{d}\delta_F} = -\frac{\mathrm{d}\tilde{\Pi}_{FF}(\delta_F|w_F)/\mathrm{d}\delta_F}{\mathrm{d}\tilde{\Pi}_{FH}(\delta_H|w_F)/\mathrm{d}\delta_H} < \frac{\mathrm{d}\delta_H^H(\delta_F)}{\mathrm{d}\delta_F} = -\frac{\mathrm{d}\tilde{\Pi}_{HF}(\delta_F)/\mathrm{d}\delta_F}{\mathrm{d}\tilde{\Pi}_{HH}(\delta_H)/\mathrm{d}\delta_H} < 0.$$

With the implicit function theorem, Assumption 4 asserts that the destination-specific competition intensity has a greater impact on the average profit of domestic sellers there than that of exporting sellers. It limits the monotonicity of the system, ensuring a unique partial equilibrium for a given wage.

Proposition 1. *Let us specify* $\{i, j\} \in \{H, F\}$*, where* $\{H, F\}$ *denote home and foreign in a two-country world. Under Assumptions 1-4, the (decentralized) market equilibrium* $\left\{\delta_j^{dmkt}, w_i, \varphi_{ij}^*, q_{ij}(\varphi), M_i, \forall i, j = H, F\right\}$ *is uniquely determined.*

The proof of Proposition 1 illustrates the additional requirements for determining equilibrium existence and uniqueness in open economies and explains their functionality. Assumption 3 requires the existence of a feasible range for the endogenous wage to ensure the partial equilibrium for a specified wage and that both sides of the TBC for the two countries are in the positive domain. Assumption 4 imposes further constraints on the complete set of parameters to establish the singularity of partial equilibrium for a specified wage within the range and ultimately attain the uniqueness of the equilibrium wage. With a general demand structure and productivity distribution, these assumptions cannot be expressed explicitly. However, the intuition behind will become clear in later examples. It is worthwhile to emphasize that up to here, we do not make any assumptions about specific parameters (i.e., τ_{ij}, f_{ij}, F_i, and L_j). Instead, we derive general conditions that ensure the existence and uniqueness of the general equilibrium. Our framework can be applied to any preferences satisfying Assumptions 1 and 2 and any productivity distributions.

It will be useful to use $\{\delta_j^{dmkt}, \forall j = H, F\}$ to refer to the solutions for the (destination-specific) Lagrange multipliers under the decentralized market solution. We now illustrate how a firm's production $q_{ij}(\delta_j^{dmkt}, \frac{\tau_{ij}w_i}{\varphi})$ and profit $\pi_{ij}(\delta_j^{dmkt}, \frac{\tau_{ij}w_i}{\varphi})$ behave differently in open economies compared to a closed economy.

First, as described by Matsuyama and Ushchev (2022), δ_j^{dmkt} is a measure of competitive pressure, which captures all the equilibrium interactions across firms and acts as a magnifier of firm heterogeneity. Firms take the measure as given, and their production decrease with the intensity of competition. However, solving the destination-specific marginal utility of income δ_j^{dmkt} requires us to combine all equilibrium conditions for all origins since consumers in country j can spread their income over all available varieties from producers in all origins. Such characteristics will cause extra misalignment between the market equilibrium and the social optimum in the open economy compared to the closed economy, where consumers only spend their income domestically. In addition, while firms in origin i take the local wage rate as given, the endogenous wage rate adjusts, given a numeraire, to ensure balanced trade. This is not the case in a closed economy, where the wage rate does not have any impact on the equilibrium outcome.

We will show that, although the general equilibrium of open economies differs from that of a closed economy in two aspects, competition intensities and endogenous wages, only the competition intensities cause extra misalignment between the market equilibrium and the social optimum.

2.2 Social Optimum

We consider a benevolent social planner who maximizes the global aggregate households' utility with access to choose quantities, cutoffs, and the masses of entrants. She is agnostic about the distribution of welfare across countries and treats households from different destinations equally. To be specific, as in Nocco et al. (2019), the social planner's problem is defined as follows.

Proposition 2 (**Social optimum**). *A (global) social optimum is the solution of the fol-*

lowing Lagrangian:

$$\mathcal{L} = \sum_i \sum_j \left\{ M_i L_j \int_{\varphi_{ij}^*}^{+\infty} u(q_{ij}(\varphi)) dG_i(\varphi) \right\}$$
$$+ \sum_i \left\{ \lambda_i^{opt} \left\{ L_i w_i - M_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} (\frac{q_{ij}(\varphi)\tau_{ij}L_j}{\varphi} + f_{ij}) w_i dG_i(\varphi) \right] + F_i w_i \right\} \right\} \right\}$$

The solution $\left\{\lambda_i^{opt}, \varphi_{ij}^*, q_{ij}(\varphi), M_i, \forall i, j = H, F\right\}$ *is uniquely determined.*

The origin-specific wage rates w_i are irrelevant to the social optimum and can be cancelled in the resource constraints. However, we retain them to obtain the optimal conditions that are comparable to the market equilibrium. The first-order conditions with respect to quantities are

$$\frac{\partial \mathcal{L}}{\partial q_{ij}(\varphi)} \Rightarrow u'(q_{ij}(\varphi)) = \frac{\lambda_i^{opt} \tau_{ij} w_i}{\varphi}, \tag{7}$$

which equate the marginal utility for consumers and marginal cost for firms. λ_i^{opt} is the marginal utility of resource and acts as an origin-specific demand shifter. From equ. (7), we obtain the implicit solution for quantities $q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi})$.

The first-order conditions regarding cutoff productivities and masses of entrants are

$$\frac{\partial \mathcal{L}}{\partial \varphi_{ij}^*} \Rightarrow (\frac{1}{\varepsilon_u(q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi_{ij}^*}))} - 1) \frac{\tau_{ij} w_i}{\varphi_{ij}^*} q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi_{ij}^*}) L_j = f_{ij} w_i, \tag{8}$$

and

$$\frac{\partial \mathcal{L}}{\partial M_i} \Rightarrow \sum_j \left\{ \int_{\varphi_{ij}^*}^{+\infty} \left[(\frac{1}{\varepsilon_u(q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi}))} - 1) \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi}) L_j - f_{ij} w_i \right] dG_i(\varphi) \right\} = F_i w_i. \tag{9}$$

$\frac{1}{\varepsilon_u(q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij} w_i}{\varphi}))}$ is the social markup a social planner assigns a firm with productivity φ to charge. Equs. (8) and (9) are the zero-cutoff-social-profit condition (ZCSPC) and the zero-expected-social-profit condition (ZESPC). We can further prove the existence and uniqueness of the social optimum.

The comparison of the decentralized market equilibrium and the social optimum is more complex with open economies for the following reasons. Firstly, λ_i^{opt} serves as a demand shifter and an amplifier of firm productivity from the social planner's perspective. Firms' production and social profits decrease with λ_i^{opt}, similar to firms' production and profits decrease with δ_j^{dmkt} in the market equilibrium. However, the impact of λ_i^{opt} is origin-specific rather than destination-specific, and importantly, it can be determined from the ZESPC of i, independent of that of j. On the other hand, the demand shifters in the decentralized equilibrium, δ_j^{dmkt}, are destination-specific, resulting in a misalignment between origins and destinations.

Second, while wage rates are endogenous with a numeraire in the decentralized market equilibrium, they are exogenous in the social optimum. As the social planner simply assigns production schedules to firms and products to consumers, wages act as a scaling factor in this scenario. As a result, we can require the social planner to reallocate resources under the market-equilibrium wages.

Lastly, the pricing strategies are different in the two equilibria. A firm with productivity φ charges a private markup of $\frac{1}{1-r_u(q_{ij}(\delta_j^{dmkt},\frac{\tau_{ij}w_i}{\varphi}))}$, while it is assigned a social markup of $\frac{1}{\varepsilon_u(q_{ij}(\lambda_i^{opt},\frac{\tau_{ij}w_i}{\varphi}))}$ in the social optimum. We can summarize the comparison as the misalignment between markup strategies, and one can see that the origin-destination misalignment also affects the markup misalignment.

Therefore, comparing the two equilibria is indirect and difficult unless we can isolate and analyze the two misalignment channels separately. One solution is to consider a planner who assigns the private markup strategies to the firms but with origin-specific demand shifters. In the following analysis, we construct a centralized market equilibrium, in which a planner has a real-revenue-maximizing objective, as an intermediate case to compare the decentralized market equilibrium with the social optimum.

2.3 Centralized Market Equilibrium

To construct the intermediate case, we evaluate equ. (2) at the equilibrium and rewrite it as $u'(q_{ij}(\varphi))\,[1-r_u(q_{ij}(\varphi))] = \frac{\delta_j^{dmkt}\tau_{ij}w_i}{\varphi}$, which equates the marginal real revenue to the marginal cost in real terms from the firms' perspective. Integrating the marginal real revenue, we obtain $\int u'(q_{ij}(\varphi))\,[1-r_u(q_{ij}(\varphi))]\,\mathrm{d}\,q_{ij}(\varphi) = u'(q_{ij}(\varphi))q_{ij}(\varphi)$, which is the real revenue for a firm selling quantity $q_{ij}(\varphi)$.

In a closed economy, Dhingra and Morrow (2019) compare the market equilibrium and social optimum by constructing a centralized market equilibrium that maximizes aggregate real revenue. Their comparisons are based on the equivalence of the decentralized and the centralized market equilibrium. In this subsection, we construct the global centralized market equilibrium similarly but apply their method in a different way, since the decentralized market equilibrium is distinct from the centralized one with large open economies. Specifically, the centralized market equilibrium applies when a global planner maximizes aggregate real revenues by choosing quantities, cutoff productivities, and masses of entrants.

Proposition 3 (Centralized market equilibrium). *A (global) centralized market equi-*

librium is the solution of the following Lagrangian:

$$\mathcal{L} = \sum_i \sum_j \left\{ M_i L_j \int_{\varphi_{ij}^*}^{+\infty} u'(q_{ij}(\varphi)) q_{ij}(\varphi) dG_i(\varphi) \right\}$$
$$+ \sum_i \left\{ \delta_i^{cmkt} \left\{ L_i w_i - M_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} \left(\frac{q_{ij}(\varphi)\tau_{ij} L_j}{\varphi} + f_{ij} \right) w_i dG_i(\varphi) \right] + F_i w_i \right\} \right\} \right\}$$

The solution $\left\{ \delta_i^{cmkt}, \varphi_{ij}^*, q_{ij}(\varphi), M_i, \forall i, j = H, F \right\}$ *is uniquely determined.*

Recall that under the decentralized market equilibrium, firms maximized profits, while recognizing that customers would adjust their consumption to the prevailing market prices in a utility-maximizing fashion. In the realization of such equilibrium, from the perspective of representative consumers, the marginal utility is equal to the marginal cost. Hence, consumers in a given destination are indifferent between consuming import and domestic goods for any specific variety. Therefore, the decentralized market equilibrium is characterized by destination-specific multipliers, which matter for producers in all origins. In the centralized market equilibrium, the planner maximizes aggregate real revenues for firms subject to their resource constraints. In realization, a real-revenue maximizing firm with productivity φ is indifferent between domestic and export production. Thus, the centralized market equilibrium is characterized by origin-specific multipliers. We will use the same acronym δ to refer to the Lagrange multiplier in the derivations, but note that this multiplier is indexed by the sellers' location.

The optimal conditions for the centralized market equilibrium with respect to quantities, cutoff productivities, and masses of entrants are the same as in the decentralized market equilibrium, except for the demand shifter δ_i^{cmkt}, which is origin-specific − see equs. (2), (3), (4), (5). Hence, the pricing strategies of firms and the shapes of the production schedules are the same between the two equilibria.

Mathematically, unlike solving the decentralized market equilibrium, solving the centralized market equilibrium does not require simultaneously considering the ZEPCs for all countries. The reason is that the demand shifter δ_i^{cmkt} is indexed by the producers' location instead of the consumers', and the ZEPCs are origin-specific. With the existence and uniqueness of the centralized market equilibrium, it is an ideal intermediate case for comparing the decentralized market equilibrium and the social optimum for the following reasons. First, in the decentralized market equilibrium, firms adopt a private markup strategy and face a destination-specific demand shifter, whereas the centralized market planner assigns firms the same private markup strategy but with an origin-specific shifter. Second, the maximization problems of the centralized market and the social planner are similar, except for their objective functions. Therefore, their demand shifters are both origin-specific, but they assign different markup strategies to firms: the social planner assigns a social markup while the centralized market planner assigns a private markup.

2.4 Comparison of Equilibria

In this subsection, we compare the two market equilibria and the social optimum in the open-economy case with two countries, called Home (H) and Foreign (F), with each other. To this end, we use $\{\delta_H^{dmkt}, \delta_F^{dmkt}\}$, $\{\delta_H^{cmkt}, \delta_F^{cmkt}\}$ to denote the solutions of the endogenous Lagrange multipliers for the two countries in the market equilibria and $\{\lambda_H^{opt}, \lambda_F^{opt}\}$ in the social optimum. We use $q_{ij}(\delta_j^{dmkt}, \varphi)$ and $q_{ij}^{dmkt}(\varphi)$ interchangeably for clearer illustration without ambiguity.

We devote separate parts to the pairwise comparisons of the decentralized with the centralized market equilibrium, of the centralized market with the social optimum, and of the decentralized market with the social optimum.

2.4.1 Decentralized vs. Centralized Market: Quantity-locus Shift

Note first that when countries are identical or an outside sector exists, $\delta_H^{dmkt} = \delta_F^{dmkt} = \delta_H^{cmkt} = \delta_F^{cmkt}$. Hence, competition intensities are identical across countries then, and associated open-economy models are identical to a closed-economy model with extra trade costs and exporting fixed costs.

When countries are asymmetric and an outside sector is absent, demand shifters in the decentralized and the centralized equilibria are generally different since the decentralized shifter is destination-specific and the centralized shifter is origin-specific. Therefore, it is necessary to examine the relative value of demand shifters to compare the decentralized market equilibrium to the centralized one, as the only difference between the two are the shifters. Put differently, once we comprehend the relationship between decentralized and centralized market shifters, the mapping of all decentralized outcomes can be obtained. The subsequent proposition establishes the corresponding relationships.

Proposition 4. *$\forall i, j = H, F$ and $i \neq j$*

$$\delta_i^{dmkt} > \delta_j^{dmkt} \Rightarrow \delta_i^{dmkt} \geqslant \delta_i^{cmkt} > \delta_j^{cmkt} \geqslant \delta_j^{dmkt}.$$

Then, $\forall \ell = H, F$,

- *quantity:* $q^{dmkt}_{\ell i}(\varphi) \leqslant q^{cmkt}_{\ell i}(\varphi)$, $q^{dmkt}_{\ell j}(\varphi) \geqslant q^{cmkt}_{\ell j}(\varphi)$.

- *cutoff productivity:* $(\varphi^*_{\ell i})^{dmkt} \geqslant (\varphi^*_{\ell i})^{cmkt}$, $(\varphi^*_{\ell j})^{dmkt} \leqslant (\varphi^*_{\ell j})^{cmkt}$.

Especially, when both countries export or $\overline{q} = +\infty$*, all inequalities strictly hold.*

Proposition 4 suggests how a specific ranking in the competitiveness across markets in the decentralized market equilibrium induces a specific ranking of the competition parameters across countries and decentralized versus centralized market equilibria as well as the associated firm-level quantity and cutoff-productivity levels in equilibrium. The clear-cut ranking follows from the fact that the first-order conditions are identical between the decentralized and the centralized market equilibria except for the Lagrange multipliers. Therefore, a ranking of the Lagrange multipliers dictates a ranking of the associated quantity levels and, in turn, of the cutoff productivities.

For intuition, consider that a real-revenue-maximizing planner faces the realization of the decentralized market equilibrium $\{\delta^{dmkt}_H, \delta^{dmkt}_F\}$ with $\delta^{dmkt}_H < \delta^{dmkt}_F$. Consumers in both countries are indifferent to consuming domestic or imported goods due to equal marginal utility and marginal costs for each variety. However, the planner prioritizes maximizing aggregate real revenue. This global maximization problem can be separated into two local (national) maximization problems. At the realization of ZEPCs in country H, $\tilde{\Pi}_{HH}(\delta^{dmkt}_H) + \tilde{\Pi}_{HF}(\delta^{dmkt}_F) = F_H$, the planner will increase aggregate real revenue in origin H by adjusting the shifters to equalize the marginal real revenue of resources for firms in i for both domestic and foreign sales. We denote the resulting marginal real revenue of resources in the ZEPC of country H as δ^{cmkt}_H and explore its properties.

Boundedness. To adjust the resource allocation between domestic and export production while maintaining the ZEPC, δ^{cmkt}_H must be between δ^{dmkt}_H and δ^{dmkt}_F, as proven in Lemma 2. The planner will change δ^{dmkt}_H and δ^{dmkt}_F towards the common equilibrium value. Hence, the planner will change the domestic and export output in opposite directions.

Consistency. The boundedness property allows us to view δ^{cmkt}_H as a weighted average of δ^{dmkt}_H and δ^{dmkt}_F. Assumption 4 ensures the dominance of the domestic market in both countries, so the sensitivity of average profits with respect to the shifter in the domestic market should be greater than in the export market. Therefore, the weight of δ^{dmkt}_H should be greater than that of δ^{dmkt}_F in order to obtain the equilibrium value of δ^{cmkt}_H. In general, this means δ^{cmkt}_i should be closer to δ^{dmkt}_i than to δ^{dmkt}_j for all $i \neq j$. The boundedness property implies that the ranking of demand shifters remains consistent in both equilibria: if $\delta^{dmkt}_H < \delta^{dmkt}_F$, then $\delta^{cmkt}_H < \delta^{cmkt}_F$. As for the outcomes, the centralized planner's adjustment will be greater for exports than for the domestic market.

Overall, we show the ranking of demand shifters in the decentralized and centralized market equilibria. Again, with the assumptions of a closed economy, symmetric countries or outside sectors, the outcomes of decentralized and centralized equilibria are the same because of the equivalence between the marginal utility of income and the marginal real revenue of resources for all countries. Without these assumptions, country-level fundamental differences generally result in differing marginal utilities of income, and the planner will adjust domestic and export production to equalize the marginal real revenue of resources. As a consequence, trade generally induces an extra channel when comparing the decentralized market equilibrium with the social optimum.

2.4.2 Centralized Market vs. Social Optimum: Quantity-locus Rotation

Recall that the centralized market equilibrium is constructed in a way such that it is comparable to the social optimum. Both are optimization problems from the planner's perspective, subjective to the same resource constraints but with different objectives. Therefore, the comparison between the centralized market and the socially optimal equilibrium can be conducted in the spirit of Dhingra and Morrow (2019). To start with, we make the following assumptions about markups.

Assumption 5 (Markups). *$(1 - r_u(q))'(\varepsilon(q))' > 0$; and, when $\lim_{q\to+\infty} \varepsilon_u(q) = 0$, $\lim_{q\to+\infty} 1 - r_u(q) \leqslant 0$.*

As in Dhingra and Morrow (2019), the first part of Assumption 5 suggests that we focus on aligned preferences, where the incentives of the market and the social planner are consistent. The second part relaxes their assumptions of interior markups. It guarantees that private markups and social markups converge for extreme quantities and that the social planner can assign at least the same quantity to firms as in the market equilibria. Specifically, when $\lim_{q\to+\infty} \varepsilon_u(q) > 0$, we obtain $\lim_{q\to+\infty} 1 - r_u(q) = \lim_{q\to+\infty} \varepsilon_u(q)$ with L'Hôpital's rule, where the range of markups and quantities in the market equilibria and the social optimum are the same. When $\lim_{q\to+\infty} \varepsilon_u(q) = 0$, we require the possible range of market quantities to be smaller than the one in the social optimum to obtain the same range for their markups. That is, there exists a $\overline{q}$, such that $\lim_{q\to+\infty} \varepsilon_u(q) = \lim_{q\to\overline{q}} 1 - r_u(q) = 0$. In what follows, we list some customary preferences (Mrázová and Neary, 2017; Dhingra and Morrow, 2019; Mayer et al., 2021) satisfying these assumptions in Table 1.

Building on the assumptions about markup properties, we introduce the following proposition, which characterizes how the misalignment between private markups and social markups can lead to misallocation in open economies.

Table 1: Properties for Common Utility Forms

	Bipower	HARA	Expo-power
$u(q)$	$\frac{aq^{1-\eta}}{1-\eta}+\frac{\beta q^{1-\theta}}{1-\theta}$	$\frac{[q/(1-\rho)+\alpha]^{\rho}-\alpha^{\rho}}{\rho/(1-\rho)}$	$\frac{1-\exp(-aq^{1-\rho})}{a}$
assumptions	$0<1-\eta<1-\theta<1$	$\alpha>0, 0<\rho<1$	$a>0, 0<p<1$
$u'(q)$	>0	>0	>0
$\lim_{q\to+\infty} u'(q)$	0	0	0
$u''(q)$	<0	<0	<0
$\lim_{q\to 0} \varepsilon_u(q)$	$1-\eta$	1	$1-\rho$
$[1-r_u(q)]'\,[\varepsilon_u(q)]'$	>0	>0	>0
$\lim_{q\to+\infty} \varepsilon_u(q)$	$1-\theta$	ρ	0
$\lim_{q\to+\infty} 1-r_u(q)$	$1-\theta$	ρ	$-\infty$

Proposition 5 (Quantity distortions). *$\forall i,j=H,F$, $q_{ij}^{cmkt}(\varphi)$ and $q_{ij}^{opt}(\varphi)$ have a unique intersection $\tilde{\varphi}_{ij}$:*[6]

- *If $(1-r_u(q))'<0$ and $\varepsilon_u'(q)<0$, $q_{ij}^{cmkt}(\varphi)<q_{ij}^{opt}(\varphi)$ for $\varphi>\tilde{\varphi}_{ij}$ and $q_{ij}^{cmkt}(\varphi)>q_{ij}^{opt}(\varphi)$ for $\varphi<\tilde{\varphi}_{ij}$.*
- *If $(1-r_u(q))'>0$ and $\varepsilon_u'(q)>0$, $q_{ij}^{cmkt}(\varphi)>q_{ij}^{opt}(\varphi)$ for $\varphi>\tilde{\varphi}_{ij}$ and $q_{ij}^{cmkt}(\varphi)<q_{ij}^{opt}(\varphi)$ for $\varphi<\tilde{\varphi}_{ij}$.*

In both cases, the domestic intersection is lower than the exporting intersection.

Proposition 5 relies on a portrait of the quantity schedules q_{ij}^{h} with $h\in\{cmkt, opt\}$ as a function of productivity φ, each. As in Dhingra and Morrow (2019), it turns out that these functions cross uniquely, and the progression and location of the loci depend on the prevailing demand structure, which further determines the monotonicity of private and social markups, without considering the truncation accruing from positive productivity cutoffs. It should be noticed that the direction of rotation is independent of the fundamentals of the countries and, therefore, the quantity loci for all country pairings are rotated clockwise or counterclockwise, depending on consumer preferences.

We do not repeat the explanation of such rotation effects, since it is clearly stated in Dhingra and Morrow (2019). The intuition is that, disregarding the truncation at the cutoff productivity, the centralized real-revenue-maximizing planner declares overproduction for a group of firms and underproduction for the rest in the decentralized market from a social planner's perspective; whether a firm overproduces or underproduces depends on the preference of consumers and its heterogeneous productivity. However, one should note that the extent of rotation differs across the destinations of sales, because the variable markups

[6] An intersection would not occur, if $\tilde{\varphi}_{ij}$ is lower than the respective cutoff productivity levels.

and exporting trade costs jointly determine the location of the intersection between two quantity loci. In the case of free trade, the intersections for domestic and exported sales are identical.

Recall that Proposition 5 illustrates how the optimal production loci can be derived by rotating the centralized market curves, disregarding the cutoff productivity induced from non-zero fixed costs. In order to discuss the cutoff effects, we postulate the following definition:

$$\underline{j}^i \equiv \left\{ j \mid \frac{L_j}{f_{ij}} = \min\left\{ \frac{L_j}{f_{ij}}, \forall j = H, F \right\} \right\}, \quad \overline{j}^i \equiv \left\{ j \mid \frac{L_j}{f_{ij}} = \max\left\{ \frac{L_j}{f_{ij}}, \forall j = H, F \right\} \right\},$$

where $\underline{j}^i$ and $\overline{j}^i$ denote the destinations with relatively higher and lower fixed costs per capita. We use $\underline{j}$ and $\overline{j}$ for simplification. This allows us to obtain the ranking of inverse fixed costs per consumer, which are the key variable affecting cutoff distortions.

Proposition 6 (**Cutoff distortions**)**.** $\forall i = H, F$,

- *If* $\varepsilon'_u(q) > 0$, $(\varphi^*_{i\overline{j}})^{cmkt} > (\varphi^*_{i\overline{j}})^{opt}$ *and* $(\varphi^*_{i\underline{j}})^{cmkt} \gtreqless (\varphi^*_{i\underline{j}})^{opt}$.
- *If* $\varepsilon'_u(q) < 0$, $(\varphi^*_{i\overline{j}})^{cmkt} < (\varphi^*_{i\overline{j}})^{opt}$ *and* $(\varphi^*_{i\underline{j}})^{cmkt} \gtreqless (\varphi^*_{i\underline{j}})^{opt}$.

Specially, when all fixed costs are zero:

- *If* $\varepsilon'_u(q) > 0$, $(\varphi^*_{i\overline{j}})^{cmkt} > (\varphi^*_{i\overline{j}})^{opt}$ *and* $(\varphi^*_{i\underline{j}})^{cmkt} > (\varphi^*_{i\underline{j}})^{opt}$.
- *If* $\varepsilon'_u(q) < 0$, $(\varphi^*_{i\overline{j}})^{cmkt} < (\varphi^*_{i\overline{j}})^{opt}$ *and* $(\varphi^*_{i\underline{j}})^{cmkt} < (\varphi^*_{i\underline{j}})^{opt}$.

Proposition 6 establishes that only the cutoff productivity levels of economies with the lower fixed costs per consumer are clearly ranked in each market as in the closed economy in Dhingra and Morrow (2019). However, the ranking of cutoff productivity levels for the country with the higher fixed costs per consumer is elusive, if all fixed market-access costs are positive. At zero fixed market-access costs, the cutoff productivity levels for market entry are clearly ranked between the centralized market equilibrium and the social optimum. This can be intuitively explained as follows.

When fixed market-access costs are positive and different and social markups increase with quantity, the lack of appropriability of a marginal variety in the market with the lower fixed cost per consumer dominates the business-stealing effect, encouraging the production of the marginal variety and decreasing the cutoffs in the centralized market equilibrium. However, such entry causes business stealing and reallocation across markets. With the extra stealing effect, the lack of appropriability of a marginal variety in the market with higher fixed costs per consumer will not necessarily dominate the business-stealing effect, resulting in the elusive cutoff ranking between the two equilibria. This statement is true even when the fixed costs of domestic and export production are the same, because the market sizes of asymmetric destinations are different. When all fixed costs are zero, there is no priority for producing the marginal variety for a specific market, and the contagion effect between markets disappears. The entry of the marginal variety in both markets can appear simultaneously, indicating the clear and consistent ranking between the two equilibria.

Overall, the comparison between the quantity loci in the centralized market equilibrium and the social optimum can be summarized as a rotation effect, which depends on the demand structure and on domestic vs. export production. However, the centralized market equilibrium is constructed as an intermediate case to comparing the decentralized market equilibrium and social optimum, and the centralized and decentralized market equilibria are generally different in open economies. Therefore, the arguments regarding the closed economy in Dhingra and Morrow (2019) do not simply extend to asymmetric large open economies. In the decentralized market equilibrium, the effect of economy-level heterogeneity measured by the demand shifters is stronger than in the centralized market equilibrium. That is, as we show in Proposition 4, the dispersion between δ_i^{dmkt} and δ_j^{dmkt} is greater than that between δ_i^{cmkt} and δ_j^{cmkt}. Thus, the value of $\delta_j^{dmkt}/\lambda_i^{opt}$ might not be bounded by the interval of private markups $1 - r_u(q)$, and an intersection of the quantity schedules for the decentralized equilibrium and the social optimum is not guaranteed.[7] Below, we will combine the shift and the rotation effects to discuss the distortions in a world with asymmetric countries.

2.4.3 Decentralized Market vs. Social Optimum

In the open-economy case, the comparison of outcomes is elusive between the decentralized market equilibrium and the social optimum. The quantity distortions of the decentralized equilibrium can be decomposed into two parts: shift effects caused by different destination-specific competition intensities and rotation effects depending on demand-side elasticities and fixed costs per capita. The shift effects depend on the fundamentals of each country and move the quantity curves of two destinations oppositely. The rotation effects rely on the monotonicity of markups and rotate all quantity schedules in the same direction with generally different strengths. How these two effects jointly matter for cutoffs and quantities are different, as we will illustrate.

[7] We show in the proof of Proposition 5 that, $\forall i = H, F$, the value of $\delta_i^{cmkt}/\lambda_i^{cmkt}$ is bounded by the interval, guaranteeing the intersection of the production schedules for the centralized market equilibrium and the social optimum.

The analysis of cutoff distortions is a one-dimensional problem: the variable-markup effects are strengthened by the competition-intensity effects in one country but counteracted by them in the other country. If the variable-markup effects dominate the competition-intensity effects on productivity cutoffs, the country with aligned variable-markup and competition-intensity effects is further away from the social optimum, while the other country is closer to it. If the competition-intensity effects dominate the variable-markup effects, the selection effects are too strong in one country and too weak in the other, indicating an elusive conclusion about which country is closer to the social optimum.

Regarding the quantity loci, the effect of variable markups causes all origin-destination-specific loci to rotate in the same direction, resulting in underproduction and overproduction. This guarantees intersections between the centralized-market and the social-optimum quantity loci. However, the competition-intensity effects shift the loci differently across destinations, leading to relatively more overproducing firms selling to one destination and relatively more underproducing firms selling to the other destination. When these effects are combined, the existence of intersections between the decentralized-market quantity loci and the social-optimum loci is not guaranteed, and all firms in one origin-destination pair may overproduce or underproduce. Moreover, there are no explicit general results regarding entry distortions in a decentralized market equilibrium.[8] We discuss some specific cases for illustration, here.

Free trade. With free trade among two asymmetric countries, the rotation effects on domestic and export quantities are the same, when disregarding the cutoff productivities. However, cutoff distortions in the rotation effects persist due to the differences in fixed costs and market sizes. The shift effects caused by varying competition intensities also persist due to country-level asymmetries.

Free trade and zero fixed market-access costs. In this case, the rotation effects on domestic and export quantities and cutoffs are identical, as zero fixed costs eliminate the impact of varying market sizes on cutoff distortions. Nonetheless, competition intensities differ across destinations, meaning that quantity-locus shift effects persist.

These two examples illustrate that shift effects generally persist when the two countries are asymmetric. A significant factor contributing to the heterogeneity of competition intensities are differences in sunk costs. This is not accounted for in a closed economy.

Overall, the difference between the open- and closed-economy cases is fundamental.

[8]The ratio $\frac{M_i^{cmkt}}{M_i^{opt}}$ can be decomposed into two terms as in Behrens et al. (2020), one measuring the effective fixed costs and the other measuring the gap between private and social markups. However, the decomposition of $\frac{M_i^{dmkt}}{M_i^{opt}}$ cannot be obtained in a similar way.

E.g., in Dhingra and Morrow (2019) or Behrens et al. (2020) distortions induced by destination-versus-origin-specific competition intensities are absent. This is not the case with open economies. In the following section, we will highlight these differences by examining a general example of constant elasticity of substitution (CES) preferences.

Our findings also extend to work on the misallocation in open economies with specific demand structures and strong assumptions. E.g., Nocco et al. (2019) consider a Melitz-Ottaviano model with an outside sector. They find that the planner is more selective than the market regarding both domestic production and exports. Moreover, the proportions of inefficiently under-supplied and over-supplied varieties are solely determined by the shape parameter of the Pareto distribution and independent of country characteristics. Baqaee and Farhi (2021) examine the case of symmetric countries with free trade and, thus, find that market inefficiency is the same across all countries. However, these studies do not allow for differences in competition intensities across destinations. As a result, they demonstrate rotation effects but not shift effects. As discussed earlier, heterogeneity in competition intensities across destinations renders the misallocation in open economies generally ambiguous. Additionally, our findings concerning shift effects suggest that a destination-specific (domestic vs. imported) rather than origin-specific (domestic vs. exported) framework should be used when comparing market and optimum outcomes.

3 Example: CES Preferences

3.1 General Discussion

In the analysis above, we only made relatively mild assumptions about preferences, whereby the case of constant elasticity of substitution (CES) preferences, the workhorse framework in modern international trade theory, was covered. CES preferences guarantee that the decentralized market allocation is efficient in a single-sector closed economy (see Dhingra and Morrow, 2019). In a multi-sector closed economy, Behrens et al. (2020) show that with CES preferences, the cutoff and output distortions are absent but the masses of entrants are distorted because of the inter-sectoral mobility of labor. The latter creates a potential misalignment between the allocation of resources from a consumer's versus a producer's perspective. In the following lemma, we prove that in a single-sector large open economy, CES preferences do not guarantee efficient cutoff levels and outputs.

Proposition 7. *Under CES utility with $u(q) = q^\rho$ for $0 < \rho < 1$, the decentralized market* *equilibrium may be inefficient. Specifically, if* $\forall i, j = F, H$ *and* $i \neq j$*:*

$$\delta_i^{dmkt} > \delta_j^{dmkt} \Rightarrow \delta_i^{dmkt} > \delta_i^{cmkt} > \delta_j^{cmkt} > \delta_j^{dmkt}.$$

Then, for $\forall \ell = H, F$*:*

- *quantity:* $q_{\ell i}^{dmkt}(\varphi) < q_{\ell i}^{opt}(\varphi)$, $q_{\ell j}^{dmkt}(\varphi) > q_{\ell j}^{opt}(\varphi)$.
- *cutoff productivity:* $(\varphi_{\ell i}^*)^{dmkt} > (\varphi_{\ell i}^*)^{opt}$, $(\varphi_{\ell j}^*)^{dmkt} < (\varphi_{\ell j}^*)^{opt}$.

Recall that the equilibrium conditions of the decentralized and centralized market frameworks are the same, except for the demand shifters, which are measures of the competition intensity. Also, with a CES demand the outcomes of the centralized market equilibrium and the social optimum are equivalent. Therefore, the destination-specific competition intensity, δ_j^{dmkt}, systematically shifts all outcomes to country j away from the optimal outcomes. In other words, under CES with asymmetric countries, the market quantity schedules never intersect with the optimal quantity schedules. Instead, depending on the cross-country ranking of competition intensities, all firms in a country overproduce for one market and underproduce for the other. Moreover, the selection effects are too strong for one market and too weak for the other.

The reason why our conclusion for open economies differs from the closed-economy ones in Dhingra and Morrow (2019) and Behrens et al. (2020) is as follows. With a CES demand, all firms charge a constant markup, which is the same in both the market equilibrium and the social optimum. In a closed economy, the equivalence between the decentralized and the centralized market equilibrium is ensured. However, in asymmetric open economies firms face different competition intensities at home and abroad, regardless of the constant markup constraint. At the realization of the decentralized equilibrium, the marginal real revenues of resources differ between domestic and export production. To map the decentralized equilibrium into a centralized one, a real-revenue-maximizing planner would adjust the production until the origin-specific marginal real revenues of resources are equal across asymmetric destinations, leading to misallocations in market-pairings-specific quantities and cutoff productivities.

3.2 A Special Case with Pareto-distributed Productivity

We further provide an example with CES demand and Pareto productivity to explain Proposition 7 explicitly. In this setting, the utility function can be described as follows:

$\forall \rho \in (0,1)$,

$$u(q) = q^{\rho}, u'(q) = \rho q^{\rho-1}, u''(q) = -\rho(1-\rho)q^{\rho-2}, 1 - r_u(q) = \varepsilon_u(q) = \rho,$$

where private and social markups are both constant at $\frac{1}{\rho}$.

Assumption 3 requires $\overline{w_F} > \underline{w_F} > 0$ such that $\forall w_F \in (\underline{w_F}, \overline{w_F})$, $\delta_{HH} > \delta_{FH}(w_F)$ and $\delta_{FF}(w_F) > \delta_{HF}$ when we choose w_H as the numeraire. We can express these two conditions explicitly with the CES demand structure:

$$\begin{aligned}&\delta_{HH}(w_F) - \delta_{FH}\\ =&D(L_H)^{\frac{(1-\rho)\gamma}{\rho}}\left[(\frac{1}{F_H w_H})(\frac{1}{f_{HH} w_H})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{HH} w_H})^{\gamma} - (\frac{1}{F_F w_F})(\frac{1}{f_{FH} w_F})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{FH} w_F})^{\gamma}\right] > 0\end{aligned}$$

and

$$\begin{aligned}&\delta_{FF}(w_F) - \delta_{HF}\\ =&D(L_F)^{\frac{(1-\rho)\gamma}{\rho}}\left[(\frac{1}{F_F w_F})(\frac{1}{f_{FF} w_F})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{FF} w_F})^{\gamma} - (\frac{1}{F_H w_H})(\frac{1}{f_{HF} w_H})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{HF} w_H})^{\gamma}\right] > 0,\end{aligned}$$

where $D = \frac{\rho^{\frac{\gamma(1+\rho)+\rho}{\rho}}(1-\rho)^{\frac{(1-\rho)\gamma}{\rho}}}{(1-\rho)\gamma-\rho}$. We further obtain the sufficient constraint to guarantee Assumption 3:

$$(\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HF}}{\tau_{FF}})^{\gamma} > (\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HH}}{\tau_{FH}})^{\gamma}. \tag{10}$$

As for Assumption 4, since $\overline{B} = +\infty$ under CES preferences, $\delta_{FF}(w_F) < \frac{L_F \overline{B}}{f_{HF} w_H}$ holds. Then we can express the corresponding explicit functions as follows: $\forall \delta_F \in (\delta_{FF}(w_F), +\infty)$,

$$\frac{d\delta_H^F(\delta_F \mid w_F)}{d\delta_F} = -\frac{\partial \Pi_{FF}/\partial \delta_F}{\partial \Pi_{FH}/\partial \delta_H} = -(\frac{L_F}{L_H})^{\frac{(1-\rho)\gamma}{\rho}}(\frac{f_{FH}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{FH}}{\tau_{FF}})^{\gamma}(\frac{\delta_F}{\delta_H})^{-(\frac{\gamma}{\rho}+1)}$$

and

$$\frac{d\delta_H^H(\delta_F)}{d\delta_F} = -\frac{\partial \Pi_{HF}/\partial \delta_F}{\partial \Pi_{HH}/\partial \delta_H} = -(\frac{L_F}{L_H})^{\frac{(1-\rho)\gamma}{\rho}}(\frac{f_{HH}}{f_{HF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HH}}{\tau_{HF}})^{\gamma}(\frac{\delta_F}{\delta_H})^{-(\frac{\gamma}{\rho}+1)}.$$

Assumption 4 requires $\frac{d\delta_H^F(\delta_F|w_F)}{d\delta_F} < \frac{d\delta_H^H(\delta_F)}{d\delta_F}$, which can also be simplified to condition (10).

Therefore, we can assume that $\forall i,j = H,F$ and $i \neq j$, $f_{ij} > f_{jj}$ and $\tau_{ij} > \tau_{jj}$ to guarantee the existence and uniqueness of a decentralized market equilibrium. We report the explicit solutions in Table 2. The endogenous wage in j satisfies:

$$\frac{w_j}{w_i} = \frac{L_i}{L_j} \frac{\left[(\frac{F_j w_j}{F_i w_i})(\frac{f_{jj} w_j}{f_{ij} w_i})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^{\gamma} - (\frac{f_{jj} f_{ii}}{f_{ji} f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^{\gamma} \right]}{\left[(\frac{F_i w_i}{F_j w_j})(\frac{f_{ii} w_i}{f_{ji} w_j})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{ii} w_i}{\tau_{ji} w_j})^{\gamma} - (\frac{f_{jj} f_{ii}}{f_{ji} f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^{\gamma} \right]}$$

Table 2: Explicit Solution with CES preferences and demand

	dmkt	*cmkt, opt*
φ_{ii}^*	$\left\{ \frac{\rho f_{ii}\left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma}-1\right]}{[(1-\rho)\gamma-\rho]F_i\left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma}-\frac{F_j w_j}{F_i w_i}(\frac{f_{ji}w_j}{f_{ii}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^{\gamma}\right]} \right\}^{\frac{1}{\gamma}}$	$\left\{ \frac{\rho f_{ii}}{[(1-\rho)\gamma-\rho]F_i} \left[1 + (\frac{\tau_{ii}}{\tau_{ij}})^{\gamma} (\frac{f_{ii}}{f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{L_j}{L_i})^{\frac{(1-\rho)\gamma}{\rho}} \right] \right\}^{\frac{1}{\gamma}}$
φ_{ij}^*	$\left\{ \frac{\rho f_{ij}\left[1-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]}{[(1-\rho)\gamma-\rho]F_i\left[\frac{F_j w_j}{F_i w_i}(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]} \right\}^{\frac{1}{\gamma}}$	$\left\{ \frac{\rho f_{ij}}{[(1-\rho)\gamma-\rho]F_i} \left[1 + (\frac{\tau_{ij}}{\tau_{ii}})^{\gamma} (\frac{f_{ij}}{f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{L_i}{L_j})^{\frac{(1-\rho)\gamma}{\rho}} \right] \right\}^{\frac{1}{\gamma}}$
q_{ii}	$\frac{\rho}{1-\rho}\frac{f_{ii}}{L_i\tau_{ii}}(\frac{1}{\varphi_{ii}^*})^{\frac{\rho}{1-\rho}}\varphi^{\frac{1}{1-\rho}}$	
q_{ij}	$\frac{\rho}{1-\rho}\frac{f_{ij}}{L_j\tau_{ij}}(\frac{1}{\varphi_{ij}^*})^{\frac{\rho}{1-\rho}}\varphi^{\frac{1}{1-\rho}}$	
M_i	$\frac{L_i\rho}{F_i\gamma}$	

We use φ_{ij}^* under each of the three equilibria for illustration. Under the decentralized market equilibrium, $(\varphi_{ij}^*)^{dmkt}$ consists of the following components. First, $\frac{\rho}{(1-\rho)\gamma-\rho}$ matters and is pinned down by the CES-demand and Pareto-technology parameters. Second, $\frac{f_{ij}}{F_i}$, an origin-destination-specific shifter, enters. Third, $1-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}$, a measure of global trade frictions consisting of all trade costs and fixed costs across all markets, enters. Finally, $\frac{F_j w_j}{F_i w_i}(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}$, measuring the relative advantage of domestic sales to imports in country j adjusted for global trade frictions, matters.

Given the equivalence between the centralized market equilibrium and the social optimum in this case, $(\varphi_{ij}^*)^{opt}$ depends on the same origin-destination-specific shifter and the same technology and preference parameters as $(\varphi_{ij}^*)^{dmkt}$. What differs is that the social planner only cares about the relative advantage from the perspective of firms as resource users. Hence, the second term in the expressions for $(\varphi_{ij}^*)^{dmkt}$ versus $(\varphi_{ij}^*)^{opt}$ differs.

The following proposition permits ranking the competition intensities and demonstrates the associated impact on the efficiency of market allocations.

Lemma 3. *$\forall i, j = H, F$ and $i \neq j$, if*

$$\begin{aligned} &(L_i)^{\frac{(1-\rho)\gamma}{\rho}} \left[(\frac{1}{F_i w_i})(\frac{1}{f_{ii} w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{ii} w_i})^{\gamma} - (\frac{1}{F_j w_j})(\frac{1}{f_{ji} w_j})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{ji} w_j})^{\gamma} \right] \\ >&(L_j)^{\frac{(1-\rho)\gamma}{\rho}} \left[(\frac{1}{F_j w_j})(\frac{1}{f_{jj} w_j})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{jj} w_j})^{\gamma} - (\frac{1}{F_i w_i})(\frac{1}{f_{ij} w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{ij} w_i})^{\gamma} \right], \end{aligned}$$

then $\delta_i^{dmkt} > \delta_j^{dmkt}$ and $\forall \ell = H, F$:

- *quantity:* $q^{dmkt}_{\ell i}(\varphi) < q^{opt}_{\ell i}(\varphi)$, $q^{dmkt}_{\ell j}(\varphi) > q^{opt}_{\ell j}(\varphi)$.
- *cutoff productivity:* $(\varphi^*_{\ell i})^{dmkt} > (\varphi^*_{\ell i})^{opt}$, $(\varphi^*_{\ell j})^{dmkt} < (\varphi^*_{\ell j})^{opt}$.

Lemma 3 is an application of Proposition 7. As discussed in Section 3.1, the equivalence between the centralized market equilibrium and the social optimum with a CES demand eliminates the rotation effect, but the shift effect persists when competition intensities vary across destinations. By providing explicit measures of competition intensities, we can quantify the shift effect. In the presence of only the shift effect, distortions in sales to different destinations behave in opposite ways. Specifically, the global social planner with a CES demand will systematically reduce all sales to one destination and increase all sales to the other.

4 Example: CARA Preferences

In the previous section, we have established a set of feasible general results and acknowledged that a number of established results for the closed economy do not carry over simply to the open economy. However, it is possible to obtain sharper comparison results in the case of variable-elasticity-of-substitution (VES) preferences, since they incorporate both shift and rotation effects. Specifically, we rely on the Constant Absolute Risk Aversion (CARA) preferences with parameter $a > 0$, an absence of fixed market-access costs ($f_{ij} = 0 \ \forall i, j = H, F$),[9] and Pareto-distributed firm productivities with a cumulative density function of $G(\varphi) = 1 - (\frac{1}{\varphi})^{\gamma}$ with $\gamma > 1$.

Utility can then be described as:

$$u(q) = 1 - e^{-aq}, u'(q) = ae^{-aq}, u''(q) = -a^2 e^{-aq}, r_u(q) = -\frac{u''(q)q}{u'(q)} = aq.$$

Note that CARA preferences fall into the domain of $\varepsilon'_u(q) < 0$ and $r'_u(q) > 0$, whereby markups are increasing with both productivity and quantity across firms.

[9]E.g., Melitz and Ottaviano (2008) use a setting of open economies without fixed market-access costs. Behrens et al. (2020) work without fixed costs in the closed economy with Pareto firms. Fixed market-access costs impede the tractability of the model substantially. However, as the marginal utility is bounded, high-cost firms will eventually not be able to survive even without fixed costs, and variable trade costs are sufficient to induce a selection of high-productivity firms into the supply to foreign markets.

4.1 Solutions under Different Equilibra

We relegate most of the analytical details to the Online Appendix. One can verify that CARA preferences satisfy Assumptions 1, 2, and 5. We now verify Assumptions 3 and 4 to derive an explicit constraint on the parameter space to ensure the existence and uniqueness of the decentralized market equilibrium.

Assumption 3 requires $\overline{w_F} > \underline{w_F} > 0$ such that $\forall w_F \in (\underline{w_F}, \overline{w_F})$, $\delta_{HH} > \delta_{FH}(w_F)$ and $\delta_{FF}(w_F) > \delta_{HF}$, when choosing w_H as the numeraire. We can express these two conditions explicitly with the CARA demand structure as:

$$\delta_{HH} - \delta_{FH}(w_F) = a^{\frac{\gamma}{\gamma+1}} (L_H \gamma \kappa_1)^{\frac{1}{\gamma+1}} \left[(\frac{1}{F_H w_H})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{HH} w_H})^{\frac{\gamma}{\gamma+1}} - (\frac{1}{F_F w_F})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{FH} w_F})^{\frac{\gamma}{\gamma+1}} \right] > 0$$

and

$$\delta_{FF}(w_F) - \delta_{HF} = a^{\frac{\gamma}{\gamma+1}} (L_F \gamma \kappa_1)^{\frac{1}{\gamma+1}} \left[(\frac{1}{F_F w_F})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{FF} w_F})^{\frac{\gamma}{\gamma+1}} - (\frac{1}{F_H w_H})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{HF} w_H})^{\frac{\gamma}{\gamma+1}} \right] > 0,$$

and further obtain the sufficient constraint to guarantee Assumption 3:

$$\frac{\tau_{HF}}{\tau_{FF}} > \frac{\tau_{HH}}{\tau_{FH}}.$$

As for Assumption 4, we can express the corresponding explicit functions as follows: $\forall \delta_F \in (\delta_{FF}(w_F), +\infty)$, we have

$$\frac{d\delta_H^F(\delta_F \mid w_F)}{d\delta_F} = -\frac{\partial \Pi_{FF}/\partial \delta_F}{\partial \Pi_{FH}/\partial \delta_H} = -\frac{L_F}{L_H} (\frac{\tau_{FH}}{\tau_{FF}})^{\gamma} (\frac{\delta_F}{\delta_H})^{-(\gamma+2)}$$

and

$$\frac{d\delta_H^H(\delta_F)}{d\delta_F} = -\frac{\partial \Pi_{HF}/\partial \delta_F}{\partial \Pi_{HH}/\partial \delta_H} = -\frac{L_F}{L_H} (\frac{\tau_{HH}}{\tau_{HF}})^{\gamma} (\frac{\delta_F}{\delta_H})^{-(\gamma+2)}.$$

Assumption 4 requires $\frac{d\delta_H^F(\delta_F|w_F)}{d\delta_F} < \frac{d\delta_H^H(\delta_F)}{d\delta_F}$, which can be simplified to $\frac{\tau_{HF}}{\tau_{FF}} > \frac{\tau_{HH}}{\tau_{FH}}$.

Overall, assuming $\forall i \neq j$, $\tau_{ij} > \tau_{jj}$ can sufficiently restrict the parameter space such that Assumptions 3 and 4 hold and, thus, the decentralized market equilibrium is uniquely determined. We further report the explicit solutions in Table 3, where $\kappa_1 = \int_0^1 (\frac{1}{z} + z - 2)\frac{z+1}{z}(ze^{z-1})^{\gamma+1} dz$ and $\mathbf{W}$ is the Lambert function (Corless et al., 1996), which satisfies $z = \mathbf{W}(z)e^{\mathbf{W}(z)}$.

Table 3: Explicit Solution of CARA Demand

	dmkt	*cmkt*	*opt*
φ^*_{ii}	$\left\{\frac{\gamma\kappa_1 L_i\tau_{ii}\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma-1\right]}{aF_i\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma-(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^\gamma\right]}\right\}^{\frac{1}{\gamma+1}}$	$\left\{\frac{\gamma\kappa_1 L_i\tau_{ii}\left[1+\frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{aF_i}\right\}^{\frac{1}{\gamma+1}}$	$\left\{\frac{L_i\tau_{ii}\left[1+\frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{a(\gamma+1)^2F_i}\right\}^{\frac{1}{\gamma+1}}$
φ^*_{ij}	$\left\{\frac{\gamma\kappa_1 L_j\tau_{ij}\left[1-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma\right]}{aF_i\left[(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma\right]}\right\}^{\frac{1}{\gamma+1}}$	$\left\{\frac{\gamma\kappa_1 L_j\tau_{ij}\left[1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^\gamma\right]}{aF_i}\right\}^{\frac{1}{\gamma+1}}$	$\left\{\frac{L_j\tau_{ij}\left[1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^\gamma\right]}{a(\gamma+1)^2F_i}\right\}^{\frac{1}{\gamma+1}}$
q_{ii}	$\frac{1}{a}\left[1-\mathbf{W}(e\frac{\varphi^*_{ii}}{\varphi})\right]$		$\frac{1}{a}\ln(\frac{\varphi}{\varphi^*_{ii}})$
q_{ij}	$\frac{1}{a}\left[1-\mathbf{W}(e\frac{\varphi^*_{ij}}{\varphi})\right]$		$\frac{1}{a}\ln(\frac{\varphi}{\varphi^*_{ij}})$
M_i	$\frac{L_i}{F_i(\gamma+1)}$		

The endogenous relative wage ratio satisfies:

$$\frac{w_j}{w_i} = \frac{L_i}{L_j}\frac{(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}{(\frac{F_i w_i}{F_j w_j})(\frac{\tau_{ii}w_i}{\tau_{ji}w_j})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}.$$

Starting with the decentralized market equilibrium, we can see that $\forall i \neq j$, $\tau_{ij} > \tau_{jj}$ ensure that $(\varphi_{ii})^{dmkt}$, $(\varphi_{ij})^{dmkt}$, and $\frac{w_j}{w_i}$ are positive. Furthermore, $(\varphi^*_{ij})^{dmkt}$ is determined as follows. First, $\gamma\kappa_1$, which only depends on the Pareto shape parameter, as well as the CARA parameter a matter. Second, $\frac{L_j\tau_{ij}}{F_i}$, an origin-destination-specific shifter, matters. Third, $1-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma$, a measure of global trade frictions consisting of all trade costs across all markets enters. Finally, $(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma$, measuring the relative advantage of domestic sales to imports in country j adjusted for global trade frictions, enter. The last two terms describe how the fundamentals of both countries jointly determine the market equilibrium, where consumers in both countries are indifferent between consuming domestic and imported goods for any variety, and changes in any parameters will systematically affect cutoffs and sales for all countries.

In the centralized market equilibrium, the first two components are the same as those in the decentralized equilibrium. However, the effects of global trade frictions, the relative advantage between domestic sales and imports, and wages disappear. This is because optimization under a centralized market is an origin-based allocation problem, and the centralized market planner does not consider wages or the effects of imports. Instead, she maximizes aggregate real revenues for each origin, such that the marginal real revenues of resources are the same between outputs for the domestic and the export markets. The tradeoff between selling to the two destinations is measured by $1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^\gamma$.

Facing another origin-based allocation problem, the social optimal planner has a similar spirit to the centralized market planner but has a different objective. Therefore, $(\varphi_{ij}^{*})^{opt}$ consists of the same origin-based fundamentals $1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^{\gamma}$, origin-destination-specific shifter $\frac{L_j\tau_{ij}}{F_i}$, and the preference parameter $\frac{1}{a}$. What differs is that the social planner's cutoff productivities depend on $(\gamma+1)^2$ in the denominator.

Regarding quantity, it can be observed that the decentralized and centralized market equilibria share the same solution, except for the cutoffs. This indicates that the quantity function inherits all effects of exogenous fundamentals from cutoffs. However, in the social optimum, the social planner assigns markups to firms by a different strategy, resulting in a different quantity function compared to the two market equilibria. Additionally, with the assumption of zero fixed cost and the properties of the Pareto distribution, the masses of entrants are the same across all three equilibria, consistent with Behrens et al. (2020) and Bagwell and Lee (2022). In the following sections, we demonstrate how these differences between the solutions correspond to our general results.

4.2 Comparisons in Cutoffs

With CARA preferences and Pareto productivities, we are able to derive an explicit expression for the destination-specific demand shifters δ_j^{dmkt}. This allows us to analyze the shift effects between the decentralized and centralized market equilibria.

Lemma 4. *$\forall i,j=H,F$ and $i\neq j$. If*

$$\frac{L_i\left[(\frac{1}{F_iw_i})(\frac{1}{\tau_{ii}w_i})^{\gamma}-(\frac{1}{F_jw_j})(\frac{1}{\tau_{ji}w_j})^{\gamma}\right]}{L_j\left[(\frac{1}{F_jw_j})(\frac{1}{\tau_{jj}w_j})^{\gamma}-(\frac{1}{F_iw_i})(\frac{1}{\tau_{ij}w_i})^{\gamma}\right]}>1, \tag{11}$$

then $\delta_i^{dmkt}>\delta_j^{dmkt}$ and $\forall \ell=H,F$, $(\varphi_{\ell i}^{*})^{dmkt}>(\varphi_{\ell i}^{*})^{cmkt}$, $(\varphi_{\ell j}^{*})^{dmkt}<(\varphi_{\ell j}^{*})^{cmkt}$.

Lemma 4 is an application of Proposition 4 in terms of cutoffs. To be specific, the explicit measure of $\delta_i^{dmkt}/\delta_j^{dmkt}$ allows us to compare the competition intensities across different markets and further compare the cutoffs and production schedules across the decentralized and centralized equilibria. Note that in Lemma 4, all inequalities strictly hold due to our assumption of zero fixed costs and, thus, both countries export.

Lemma 5. *$\forall i,j=H,F$, $\left[\frac{(\varphi_{ij}^{*})^{cmkt}}{(\varphi_{ij}^{*})^{opt}}\right]^{\gamma+1}=(\gamma+1)A<1$, where $A=\int_0^1 z^{\gamma+1}\left(e^{z-1}\right)^{\gamma+1}dz$.*

Lemma 5 exemplifies Proposition 6 and summarizes the rotation effect discussed in Section 2.4.2. Due to the increasing private and social markups associated with CARA preferences, the selection effect under the centralized equilibrium is weaker than that in the social optimum.

Proposition 8. *$\forall i,j = H,F$ and $i \neq j$, if condition (11) holds, then $\forall \ell = H,F$, $(\varphi^*_{\ell j})^{dmkt} < (\varphi^*_{\ell j})^{opt}, (\varphi^*_{\ell i})^{dmkt} \gtrless (\varphi^*_{\ell i})^{opt}$.*

Proposition 8 can be obtained by combining Lemma 4 and 5. We can observe that the shift effects depend on the fundamentals of destinations, while the rotation effects depend on the demand structure of consumers and are independent of other fundamentals. In an open economy, the heterogeneity of countries creates additional market distortions. Specifically, for destinations with lower competition intensity, the shift and rotation effects are consistent, leading to weak market selection. In contrast, for destinations with higher competition intensity, the shift and rotation effects are opposite, resulting in an elusive effect on cutoff productivities. Since the mass of entrants is efficient in the market equilibrium, distortions in the mass of producing firms are determined by the cutoff distortions.

4.3 Comparisons in Quantities

Lemma 6. *$\forall i,j = H,F$, $q^{cmkt}_{ij}(\varphi)$ and $q^{opt}_{ij}(\varphi)$ have a unique crossing $\tilde{\varphi}^{cmkt}_{ij}$: $q^{cmkt}_{ij}(\varphi) < q^{opt}_{ij}(\varphi)$ for $\varphi > \tilde{\varphi}^{cmkt}_{ij}$ and $q^{cmkt}_{ij}(\varphi) > q^{opt}_{ij}(\varphi)$ for $\varphi < \tilde{\varphi}^{cmkt}_{ij}$. Besides, $\forall i \neq j$, $\tilde{\varphi}^{cmkt}_{ij} = \frac{\tau_{ij}}{\tau_{ii}} \tilde{\varphi}^{cmkt}_{ii}$.*

Lemma 6 is an example of Proposition 5, illustrating the rotation effect on the quantity schedules. Since CARA preferences exhibit increasing markups, the more productive firms with $\varphi > \tilde{\varphi}_{ij}$ underproduce, while the less productive firms with $\varphi < \tilde{\varphi}_{ij}$ overproduce, in the centralized market equilibrium compared to the social optimum. One can see that the direction of rotation is independent of origins and destinations, but the relative location of the intersections depends on the ratio of exporting and domestic trade costs.

Proposition 9. *$\forall i,j = H,F$ and $i \neq j$, if condition (11) holds, then $\delta^{dmkt}_i > \delta^{dmkt}_j$ and $\forall \ell = H,F$,*

- *$q^{dmkt}_{\ell j}(\varphi)$ and $q^{opt}_{\ell j}(\varphi)$ intersect uniquely at $\tilde{\varphi}^{dmkt}_{\ell j}$, $q^{dmkt}_{\ell j}(\varphi) < q^{opt}_{\ell j}(\varphi)$ for $\varphi > \tilde{\varphi}^{dmkt}_{\ell j}$ and $q^{dmkt}_{\ell j}(\varphi) > q^{opt}_{\ell j}(\varphi)$ for $\varphi < \tilde{\varphi}^{dmkt}_{\ell j}$.*
- *If $q^{dmkt}_{\ell i}(\varphi)$ and $q^{opt}_{\ell i}(\varphi)$ intersect at $\tilde{\varphi}^{dmkt}_{\ell i}$, the intersection is unique. Besides, $q^{dmkt}_{\ell i}(\varphi) < q^{opt}_{\ell i}(\varphi)$ for $\varphi > \tilde{\varphi}^{dmkt}_{\ell i}$ and $q^{dmkt}_{\ell i}(\varphi) > q^{opt}_{\ell i}(\varphi)$ for $\varphi < \tilde{\varphi}^{dmkt}_{\ell i}$.*
- *If $q^{dmkt}_{\ell i}(\varphi)$ and $q^{opt}_{\ell i}(\varphi)$ do not intersect, then $q^{opt}_{\ell i}(\varphi) > q^{dmkt}_{\ell i}(\varphi)$ for all φ.*

Proposition 9 shows the joint shift and rotation effects on the quantity schedules. Since CARA preferences feature increasing markups, the shift and rotation effects are consistent for sales to destinations with the lower competition intensity. Therefore, the quantity loci of the market equilibrium and social optimum intersect, so that high-productivity firms underproduce and low-productivity firms overproduce.

However, for sales to destinations with the higher competition intensity, the distortions in quantity are elusive since the shift and rotation effects are counteracting. The existence of an intersection of the quantity schedules depends on which effect dominates. If the rotation effect dominates the shift effect, an intersection exists, resulting in high-productivity firms underproducing and low-productivity firms overproducing. If the shift effect dominates the rotation effect, a quantity-locus intersection does not exist, and all firms underproduce.

Overall, most of the results are general and do not permit a uniform conclusion regarding the distortions existing in the market. More specifically, the patterns of misallocation do not only depend on the properties of the demand structure but also on the fundamentals of countries, which represents a critical distinction between open and closed economies. In the following section, we offer a specific parameterization and quantification to further assess this issue.

5 Quantification: China versus the Rest of the World

In this section, we conduct a quantification based on the CARA demand in Section 4, which exhibits variable markups and induces both shift and rotation effects, using data on China and the Rest of the World (RoW). Specifically, we use two sources of data for parameter estimation and model calibration: China's Annual Surveys of Industrial Firms (CASIF) and the United Nations Industrial Development Organization's (UNIDO) Industry Statistics. We use data for the years 2004-2007 for reasons of quality and coverage in CASIF.

Following Yu (2015) and Dai and Xu (2017), we exclude firm observations in CASIF if they meet any of the following criteria: (1) missing or negative values for any of the following variables: exports and total revenues (also dubbed output); (2) exports exceed 85% of total revenues;[10] (3) state-owned firms;[11] (4) firms with less than eight employees. Furthermore, we follow Behrens et al. (2020) to trim the top and bottom 2% of the domestic revenue distribution to mitigate the impact of extreme values. Following these cleaning procedures, our sample comprises between 197,174 and 264,566 firms, depending on the considered year.

Our subsequent analysis draws on the following statistics. We begin by using the

[10] This eliminates export-processing firms as well as all those firms, where reported exports exceed total revenues (an obvious data error).

[11] Defined as ones with more than 50% capital from the state.

fraction of exporters in China as well as firm-level domestic revenues in the manufacturing sector, both from CASIF. The fraction of exporters is obtained by dividing the number of firms with positive exports by the total number of manufacturing firms in CASIF. Domestic sales are calculated by subtracting export sales from total revenues in CASIF. In addition, we utilize data on total revenues (output), total exports of China, total exports of the RoW to China, employment, and wages per capita from UNIDO. We will elaborate on how we utilize these data in the following sections.

As a precursor to the subsequent analysis, let us state that we will throughout assume that domestic trade costs are unity, so that $\tau_{ii} = 1$ for $i \in \{C, W\}$, where C and W are short hands for China and the RoW, respectively. Any ratio of foreign to domestic trade costs then reduces to $\tau_{ij}/\tau_{ii} \equiv \tau_{ij}$.

5.1 Pareto Shape (γ) and Other Fundamental Parameters

In what follows, we will describe all steps towards obtaining the relevant parameters. We summarize their respective values in Table 4 for each year in 2004-2007 separately.

First, we follow Behrens et al. (2020) to compute the Pareto shape parameter γ from the standard deviation of log (here, domestic) sales of the firms in China's manufacturing sector. Using $x_{CC}^{dmkt}(\varphi) = p_{CC}^{dmkt}(\varphi) q_{CC}^{dmkt}(\varphi)$ to denote the domestic sales per customer in China, we can define the average of the log of total domestic sales as:

$$\begin{aligned}\text{Mean}(\ln(x_{CC}^{dmkt}(\varphi)L_C)) =& \frac{1}{1 - G\left((\varphi_{CC}^*)^{dmkt}\right)} \int_{(\varphi_{CC}^*)^{dmkt}}^{+\infty} \ln\left\{\left[\frac{1}{1 - r_u\left(q_{CC}^{dmkt}(\varphi)\right)}\right] \frac{w_C}{\varphi} q_{CC}^{dmkt}(\varphi) L_C\right\} dG(\varphi) \\ =& \ln(w_C) + \ln(L_C) - \ln(a) - \ln\left((\varphi_{CC}^*)^{dmkt}\right) + \tilde{M}, \end{aligned} \tag{12}$$

where a is the CARA parameter and $\tilde{M} = \gamma \int_0^1 \ln(\frac{1}{z} - 1) \frac{z+1}{z} (ze^{z-1})^\gamma dz - \frac{1}{\gamma}$, which depends on γ only. The standard deviation of firm-level domestic sales, $\ln(x_{CC}^{dmkt}(\varphi)L_C)$, is then:

$$\begin{aligned}&\text{Std.dev.}(\ln(x_{CC}^{dmkt}(\varphi)L_C)) \\ =&\sqrt{\frac{1}{1 - G\left((\varphi_{CC}^*)^{dmkt}\right)} \int_{(\varphi_{CC}^*)^{dmkt}}^{+\infty} \left\{\ln\left\{\left[\frac{1}{1 - r_u\left(q_{CC}^{dmkt}(\varphi)\right)}\right] \frac{w_C}{\varphi} q_{CC}^{dmkt}(\varphi) L_C\right\} - \text{Mean}(\ln(x_{CC}^{dmkt}(\varphi)L_C))\right\}^2 dG(\varphi)} \\ =&\sqrt{\gamma \int_0^1 \left[\ln\left(\frac{1-z}{z} ze^{z-1}\right) - \tilde{M}\right]^2 \frac{z+1}{z} (ze^{z-1})^\gamma dz}. \end{aligned} \tag{13}$$

which also depends on γ only. We obtain annual values of Std.dev.$(\ln(x_{CC}^{dmkt}(\varphi))L_C)$ and γ as reported at the top of Table 4. Note that γ is slightly increasing over the period of investigation. The latter indicates, consistent with the progression of Std.dev.$(\ln(x_{CC}^{dmkt}(\varphi))L_C)$ across the years, that the dispersion of firms increases with time.

From the CASIF data, we compute the ratio of the number of exporting firms (N_{CW}^{dmkt})

and the number of domestically selling firms in China (N_{CC}^{dmkt}) and report it in Table 4.[12] According to the table, about 18% of the firms in China's manufacturing are exporters in the earlier years of the sample, and this ratio drops in the last two years of the data by several percentage points. Note that the exporter ratio features an important relationship with the cutoff-productivity ratio of domestic-seller versus exporter:

$$\frac{N_{CW}^{dmkt}}{N_{CC}^{dmkt}} = \frac{M_C \left[1 - G((\varphi_{CW}^*)^{dmkt})\right]}{M_C \left[1 - G((\varphi_{CC}^*)^{dmkt})\right]} = \left[\frac{(\varphi_{CC}^*)^{dmkt}}{(\varphi_{CW}^*)^{dmkt}}\right]^{\gamma}. \tag{14}$$

With an estimate of γ at hand, we can compute the ratio of exporter-to-domestic-seller cutoff-productivity ratio in the decentralized market equilibrium, $\widehat{\frac{(\varphi_{CW}^*)^{dmkt}}{(\varphi_{CC}^*)^{dmkt}}}$, based on the exporter ratio. The corresponding figures in Table 4 suggest that, consistent with the exporter ratio being smaller than unity and declining, the exporter-to-domestic-seller cutoff ratio is larger than unity and increasing over time.

We use data on total employment in manufacturing (L_C and L_W) and wages per worker (w_C and w_W) from UNIDO.[13] We report on the ratios of employment and wages per employee for the RoW versus China across the years in Table 4. Consider the aggregate revenues from exporting (X_{CW}^{dmkt}) relative to domestic sales for China (X_{CC}^{dmkt}):

$$\frac{X_{CW}^{dmkt}}{X_{CC}^{dmkt}} = \frac{M_C \int_{(\varphi_{CW}^*)^{dmkt}}^{+\infty} x_{CW}^{dmkt}(\varphi) L_W dG(\varphi)}{M_C \int_{(\varphi_{CC}^*)^{dmkt}}^{+\infty} x_{CC}^{dmkt}(\varphi) L_C dG(\varphi)} = \frac{\tau_{CW} L_W}{L_C} \left[\frac{(\varphi_{CC}^*)^{dmkt}}{(\varphi_{CW}^*)^{dmkt}}\right]^{\gamma+1}. \tag{15}$$

With an estimate of $\widehat{\frac{(\varphi_{CW}^*)^{dmkt}}{(\varphi_{CC}^*)^{dmkt}}}$ and data on (L_W/L_C), we obtain an estimate of $\hat{\tau_{CW}}$ from the aggregate sales ratio. It is useful to consider another expression for the aggregate sales ratio, this time for China and the RoW:

$$\begin{cases} \dfrac{X_{CW}^{dmkt}}{X_{CC}^{dmkt}} = \dfrac{\left[(\frac{F_W w_W}{F_C w_C})(\frac{w_W}{\tau_{CW} w_C})^{\gamma} - (\frac{1}{\tau_{WC}\tau_{CW}})^{\gamma}\right]}{\left[1 - (\frac{F_W w_W}{F_C w_C})(\frac{w_W}{\tau_{CW} w_C})^{\gamma}\right]} \\ \dfrac{X_{WC}^{dmkt}}{X_{WW}^{dmkt}} = \dfrac{\left[(\frac{F_C w_C}{F_W w_W})(\frac{w_C}{\tau_{WC} w_W})^{\gamma} - (\frac{1}{\tau_{CW}\tau_{WC}})^{\gamma}\right]}{\left[1 - (\frac{F_C w_C}{F_W w_W})(\frac{w_C}{\tau_{WC} w_W})^{\gamma}\right]} \end{cases} \tag{16}$$

From the latter two equations we can solve for the unknown sunk entry-cost ratio, $\frac{F_W}{F_C}$, and

[12]Note that the number of exporting firms is smaller than that of domestic sellers, and the set of domestic sellers includes the one of exporting firms by assumption and design.

[13]In equilibrium, wages are endogenous. However, it is customary to identify the model parameters given data on wages.

the variable export costs from the RoW to China, τ_{WC}. The figures in Table 4 suggest that the variable trade costs for China's exporters to te RoW declined over the years at a declining rate. The trade costs of foreign exporters to China declined much more massively in comparison during the period of investigation. The sunk costs appear to have been considerably higher in China than in the RoW throughout the considered time span, with the ratio staying roughly constant.

Table 4: Parameters in 2004-2007

	Year			
Parameter	2004	2005	2006	2007
N_{CC}^{dmkt}	$197,174$	$200,989$	$228,385$	$264,566$
Std.dev($\ln(x_{CC}^{dmkt}(\varphi)L_C)$)	1.0096	1.0568	1.0820	1.1035
γ	1.3775	1.7622	2.0388	2.3375
$N_{CW}^{dmkt}/N_{CC}^{dmkt}$	0.1877	0.1963	0.1837	0.1554
$(\varphi_{CW}^*)^{dmkt}/(\varphi_{CC}^*)^{dmkt}$	3.3690	2.5195	2.2960	2.2175
w_W/w_C	11.1400	10.0272	8.2769	7.2935
L_W/L_C	2.0067	1.9351	1.9356	1.7502
τ_{CW}	2.8323	2.1525	2.1289	2.2080
τ_{WC}	11.2655	6.6691	4.5238	3.3511
F_W/F_C	0.0034	0.0017	0.0019	0.0019

Using the parameters in Table 4, we can proceed with quantifying various distortions in the next subsection.

5.2 Quantifying Distortions

We focus on market distortions for cutoff productivities, quantity, and welfare, here.

Productivity-cutoff distortions. We continue using superscripts $\{dmkt, cmkt, opt\}$ to refer to the decentralized and centralized market and the optimum outcomes, respectively.

Let us start with a comparison of the centralized market versus the socially optimal cutoff-productivity ratios:

$$\left[\frac{(\varphi_{CC}^*)^{cmkt}}{(\varphi_{CC}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{CW}^*)^{cmkt}}{(\varphi_{CW}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{WW}^*)^{cmkt}}{(\varphi_{WW}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{WC}^*)^{cmkt}}{(\varphi_{WC}^*)^{opt}}\right]^{\gamma+1} = A(\gamma+1), \tag{17}$$

Hence, all these cutoff ratios are constant, as $A = \int_0^1 z^{\gamma+1}(e^{z-1})^{\gamma+1}dz$, and they depend on the Pareto shape parameter γ alone. This underscores the fact that the rotation effects on cutoffs are independent of the country-specific fundamentals. Table 5 presents the reported ratios.

Let us next compare the decentralized with the centralized market in terms of cutoff-productivity levels $\forall i, j = C, W, i \neq j$:

$$\begin{cases} \left[\frac{(\varphi_{ii}^*)^{dmkt}}{(\varphi_{ii}^*)^{cmkt}}\right]^{\gamma+1} = \frac{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^{\gamma} - 1\right]}{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^{\gamma} - (\frac{F_j w_j}{F_i w_i})(\frac{\tau_{ji} w_j}{\tau_{ii} w_i})^{\gamma}\right]\left[1 + \frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^{\gamma}\right]} \\ \left[\frac{(\varphi_{ij}^*)^{dmkt}}{(\varphi_{ij}^*)^{cmkt}}\right]^{\gamma+1} = \frac{\left[1 - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]}{\left[(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^{\gamma} - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]\left[1 + \frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^{\gamma}\right]} \end{cases} \tag{18}$$

With the parameters listed in Table 4, we can solve for the cutoff ratios, which are summarized in the top block of Table 5.

For a more detailed analysis of these cutoff productivity ratios, consider Lemma 4. It is worth noting that, between 2004 and 2007, we observe that $(\varphi_{iC}^*)^{dmkt}/(\varphi_{iC}^*)^{cmkt} < 1$ and $(\varphi_{iW}^*)^{dmkt}/(\varphi_{iW}^*)^{cmkt} > 1$ for all $i = C, W$. Consequently, we have $\delta_W^{dmkt} > \delta_W^{cmkt} > \delta_C^{cmkt} > \delta_C^{dmkt}$ and, hence,

$$L_W\left[(\frac{1}{F_W w_W})(\frac{1}{w_W})^{\gamma} - (\frac{1}{F_C w_C})(\frac{1}{\tau_{CW} w_C})^{\gamma}\right] > L_C\left[(\frac{1}{F_C w_C})(\frac{1}{w_C})^{\gamma} - (\frac{1}{F_W w_W})(\frac{1}{\tau_{WC} w_W})^{\gamma}\right].$$

Additionally, consistent with Lemma 5, we have $\forall i, j = C, W$, $(\varphi_{ij})^{cmkt} < (\varphi_{ij})^{opt}$. It is worth noting that $(\varphi_{ij})^{cmkt}/(\varphi_{ij})^{opt}$ solely depends on γ and ranges from 0.7103 to 0.7925. Referred to as the "rotation effect," this component of the cutoff distortions is driven by the discrepancy between the pricing strategies of real-revenue-maximizing and utility-maximizing planners and it is the only distortion present in a closed economy.

The cutoff-productivity ratios for the decentralized market versus the social optimum are consistently lower than unity across all market pairings and years in the bottom block of Table 5. These ratios are proportionately shifted to the ones in the center block, as discussed above. The selection effects are all too weak, indicating a domination of the rotation effects over the shift effects. Specifically, for sales to the RoW, where the competition intensity is higher, the rotation and shift effects are counteracting, but the market still offers too wide a variety of products compared to the social optimum. Using the year 2006 as an example, the ratios $(\varphi_{ij})^{dmkt}/(\varphi_{ij})^{opt}$ range from 0.7175 to 0.8161, demonstrating a big heterogeneity of cutoff distortions and emphasizing the requirement of an open-economy model to quantify misallocation distortions.

Intersections with optimal quantity curves. Due to the dominance of rotation effects, the production schedules of the market and the social optimum intersect for all country pairings. We can numerically solve for the location of these intersections, which

Table 5: Cutoff distortions in 2004-2007

Cutoff ratio	Year 2004	2005	2006	2007
$(\varphi_{CC}^{*})^{dmkt}/(\varphi_{CC}^{*})^{cmkt}$	0.9525	0.9557	0.9798	0.9991
$(\varphi_{CW}^{*})^{dmkt}/(\varphi_{CW}^{*})^{cmkt}$	1.1329	1.1186	1.0567	1.0034
$(\varphi_{WW}^{*})^{dmkt}/(\varphi_{WW}^{*})^{cmkt}$	1.0037	1.0035	1.0020	1.0001
$(\varphi_{WC}^{*})^{dmkt}/(\varphi_{WC}^{*})^{cmkt}$	0.8438	0.8574	0.9290	0.9958
$(\varphi_{ij}^{*})^{cmkt}/(\varphi_{ij}^{*})^{opt}$	0.7103	0.7499	0.7723	0.7925
$(\varphi_{CC}^{*})^{dmkt}/(\varphi_{CC}^{*})^{opt}$	0.6765	0.7166	0.7567	0.7918
$(\varphi_{CW}^{*})^{dmkt}/(\varphi_{CW}^{*})^{opt}$	0.8047	0.8388	0.8161	0.7952
$(\varphi_{WW}^{*})^{dmkt}/(\varphi_{WW}^{*})^{opt}$	0.7129	0.7525	0.7739	0.7926
$(\varphi_{WC}^{*})^{dmkt}/(\varphi_{WC}^{*})^{opt}$	0.5993	0.6429	0.7175	0.7892

indicate the distribution of overproducing and underproducing firms. In Section 4.3, we demonstrate that $\forall \ell \in \{dmkt, cmkt\}$, the unique intersection $\tilde{\varphi}_{ij}^{\ell} \in ((\varphi_{ij}^{*})^{\ell}, +\infty)$ satisfies $q_{ij}^{opt}(\tilde{\varphi}_{ij}^{\ell}) = q_{ij}^{\ell}(\tilde{\varphi}_{ij}^{\ell})$ and

$$\frac{(\varphi_{ij}^{*})^{\ell}}{(\varphi_{ij}^{*})^{opt}} = \mathbf{W}(e\frac{(\varphi_{ij}^{*})^{\ell}}{\tilde{\varphi}_{ij}^{\ell}}), \tag{19}$$

where $\mathbf{W}$ is the Lambert function. Therefore, we can solve for $(\varphi_{ij}^{*})^{\ell}/\tilde{\varphi}_{ij}^{\ell}$ numerically, $\forall i, j = C, W$ and report the intersection ratio in Table 6.

Table 6: Intersections between the benchmark and optimal quantity curves in 2004-2007

Intersection ratio	Year 2004	2005	2006	2007
$(\varphi_{ij}^{*})^{cmkt}/\tilde{\varphi}_{ij}^{cmkt}$	0.5316	0.5839	0.6151	0.6440
$(\varphi_{CC}^{*})^{dmkt}/\tilde{\varphi}_{CC}^{dmkt}$	0.4895	0.5398	0.5933	0.6429
$(\varphi_{CW}^{*})^{dmkt}/\tilde{\varphi}_{CW}^{dmkt}$	0.6619	0.7139	0.6791	0.6479
$(\varphi_{WW}^{*})^{dmkt}/\tilde{\varphi}_{WW}^{dmkt}$	0.5350	0.5875	0.6172	0.6442
$(\varphi_{WC}^{*})^{dmkt}/\tilde{\varphi}_{WC}^{dmkt}$	0.4015	0.4498	0.5409	0.6392

We find that the intersection ratios are identical for all centralized market quantities (i.e., irrespective of i and j), as the rotation effects only depend on γ. Besides, for the destination with the lower competition intensity (here, China), the intersection ratios under the decentralized market are smaller than their counterparts under the centralized equilibrium. In 2006, $(\varphi_{CC}^{*})^{dmkt}/\tilde{\varphi}_{CC}^{dmkt}$ and $(\varphi_{WC}^{*})^{dmkt}/\tilde{\varphi}_{WC}^{dmkt}$ are 0.5933 and 0.5409, respectively, which are both smaller than $(\varphi_{ij}^{*})^{cmkt}/\tilde{\varphi}_{ij}^{cmkt} = 0.6151$. On the contrary, $(\varphi_{CW}^{*})^{dmkt}/\tilde{\varphi}_{CW}^{dmkt}$ and $(\varphi_{WW}^{*})^{dmkt}/\tilde{\varphi}_{WW}^{dmkt}$, with values of 0.6791 and 0.6172, respectively, are larger than $(\varphi_{ij}^{*})^{cmkt}/\tilde{\varphi}_{ij}^{cmkt}$.

Underproducing versus overproducing firms. We can now determine the percentage of firms that should exit and those that over- or under-produce. Specifically, we use the following two equations to calculate the relative producer masses in the optimum versus decentralized market and below the intersection of the optimum and decentralized market quantity curves for all $i, j = C, W$:

$$\frac{N_{ij}^{opt}}{N_{ij}^{dmkt}} = \left[\frac{(\varphi_{ij}^{*})^{dmkt}}{(\varphi_{ij}^{*})^{opt}}\right]^{\gamma}, \quad \frac{\tilde{N}_{ij}^{dmkt}}{N_{ij}^{dmkt}} = \left[\frac{(\varphi_{ij}^{*})^{dmkt}}{(\tilde{\varphi}_{ij})^{dmkt}}\right]^{\gamma}. \tag{20}$$

We present the results in Table 7 and visualize the quantity loci for all country-pair-specific equilibria, including the decentralized and centralized markets as well as the global social optimum, in Figure 2.

Table 7: Composition of inefficient firms in 2004-2007

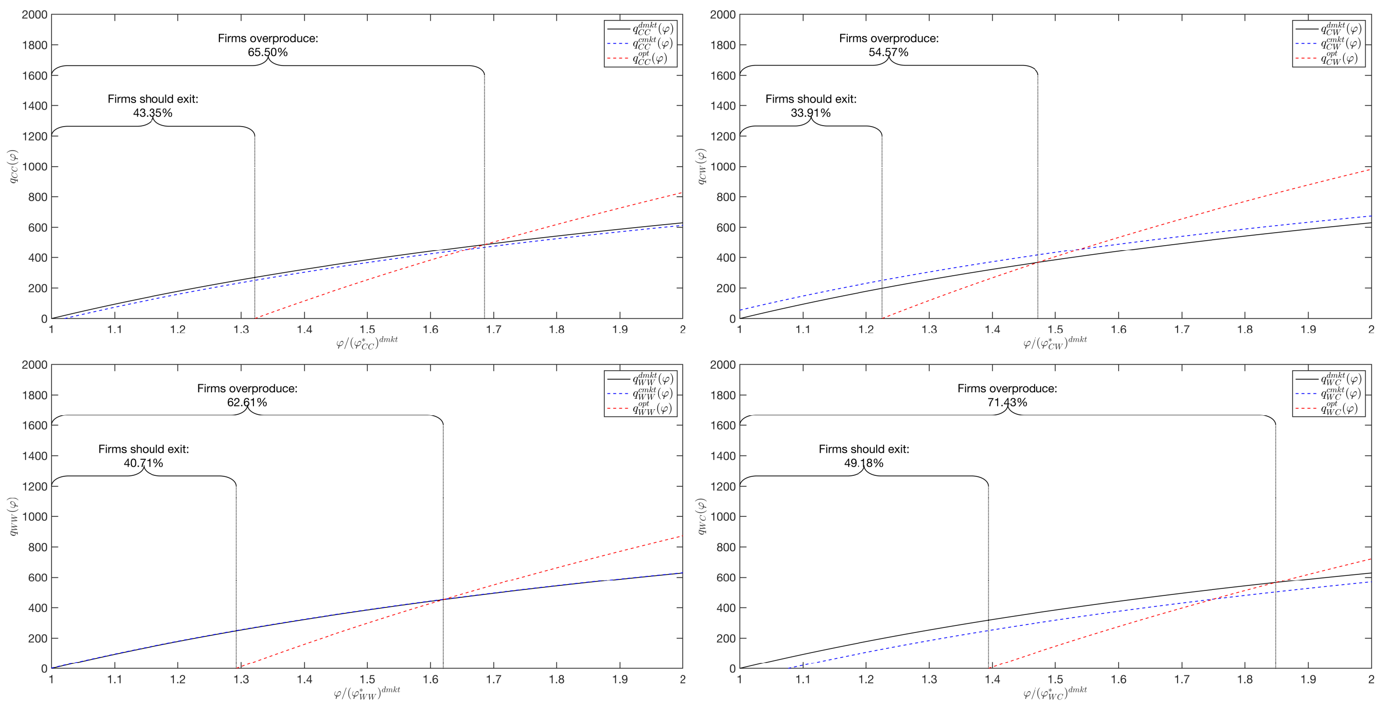
	Year			
Ratio	2004	2005	2006	2007
$N_{CC}^{opt}/N_{CC}^{dmkt}$	0.5837	0.5559	0.5665	0.5794
$\tilde{N}_{CC}^{dmkt}/N_{CC}^{dmkt}$	0.3738	0.3374	0.3450	0.3561
$N_{CW}^{opt}/N_{CW}^{dmkt}$	0.7413	0.7336	0.6609	0.5853
$\tilde{N}_{CW}^{dmkt}/N_{CW}^{dmkt}$	0.5664	0.5522	0.4543	0.3626
$N_{WW}^{opt}/N_{WW}^{dmkt}$	0.6274	0.6059	0.5929	0.5809
$\tilde{N}_{WW}^{dmkt}/N_{WW}^{dmkt}$	0.4225	0.3917	0.3739	0.3577
$N_{WC}^{opt}/N_{WC}^{dmkt}$	0.4940	0.4591	0.5082	0.5750
$\tilde{N}_{WC}^{dmkt}/N_{WC}^{dmkt}$	0.2845	0.2447	0.2857	0.3513

The results in Figure 2 suggest that for China as the customer market (the second index is C), the relatively smaller market with lower competition intensity, the producer ratios are smaller than the comparison ratios for the RoW as the customer market (the second index is W). To see this, compare the value for $N_{CC}^{opt}/N_{CC}^{dmkt}$ with that for $N_{CW}^{opt}/N_{CW}^{dmkt}$ and the value for $N_{WC}^{opt}/N_{WC}^{dmkt}$ with that for $N_{WW}^{opt}/N_{WW}^{dmkt}$ in the table. As a consequence, as shown in Table 7 using results from 2006, firms in the bottom 43.45% (100% - 56.65%) of the productivity distribution should exit for domestic sales in China, while for exports to the RoW, firms in the bottom 33.91% (100% - 66.09%) should exit to achieve a global optimum. Similarly, for firms domestically selling in the RoW and those exporting to China, the bottom 40.71% (100% - 59.29%) and 49.18% (100% - 50.82%) should exit, respectively.

Consistent with the latter results, there are relatively fewer underproducing firms for sales to the country with the lower competition intensity, China. In contrast, there are relatively fewer overproducing firms for sales to the country with the higher competition intensity, the RoW. E.g., the percentages of overproducing firms in domestic sales and exports to China are 65.50% and 71.43%, respectively, while those in domestic sales and exports to the RoW are 62.61% and 54.57%, respectively.

One should also note that these results exhibit the properties of boundedness and consistency mentioned in Proposition 4. Specifically, δ_i^{cmkt} should be closer to δ_i^{dmkt} than to δ_j^{dmkt} for all $i \neq j$. Therefore, for a given origin, the shift effects should be greater on exports than on domestic production, which can be clearly seen in Figure 2.

Figure 2: Quantity curves under different equilibria in 2006 ($a = 0.0005$)



Welfare. $\forall \ell \in \{dmkt, cmkt\}$, with $q_{ij}^{\ell}(\varphi) = \frac{1}{a}\left[1 - \mathbf{W}(e\frac{(\varphi_{ij}^*)^{\ell}}{\varphi})\right]$, we can obtain an expression for aggregate welfare in country j in the decentralized/centralized market equilibrium:

$$
\begin{aligned}
U_j^{\ell} &= L_j \sum_i M_i \int_{(\varphi_{ij}^*)^{\ell}}^{+\infty} u(q_{ij}^{\ell}(\varphi))dG(\varphi) \\
&= L_j \frac{1}{(\gamma+1)}(\frac{1}{\gamma+1} - A)\frac{L_C}{F_C}\left[\frac{1}{(\varphi_{Cj}^*)^{\ell}}\right]^{\gamma}\left[1 + \frac{L_W}{L_C}\frac{F_C}{F_W}(\frac{(\varphi_{Cj}^*)^{\ell}}{(\varphi_{Wj}^*)^{\ell}})^{\gamma}\right].
\end{aligned} \tag{21}
$$

Similarly, noting $q_{ij}^{opt}(\varphi) = \frac{1}{a}\ln(\frac{\varphi}{(\varphi_{ij}^*)^{opt}})$, we obtain the aggregate welfare in country j in the social optimum:

$$U_j^{opt} = L_j \frac{1}{(\gamma+1)^2} \frac{L_C}{F_C} \left[\frac{1}{(\varphi_{Cj}^*)^{opt}} \right]^\gamma \left[1 + \frac{L_W}{L_C} \frac{F_C}{F_W} \left(\frac{(\varphi_{Cj}^*)^{opt}}{(\varphi_{Wj}^*)^{opt}} \right)^\gamma \right], \tag{22}$$

The country-specific allocative efficiency of the decentralized market equilibrium can then be compactly measured by:

$$\frac{U_j^{dmkt}}{U_j^{opt}} = \underbrace{[1-(\gamma+1)A] \left[\frac{1}{A(\gamma+1)} \right]^{\frac{\gamma}{\gamma+1}}}_{\text{Planner's objective}} \underbrace{\left[\frac{(\varphi_{Cj}^*)^{cmkt}}{(\varphi_{Cj}^*)^{dmkt}} \right]^\gamma}_{\text{Competition-intensity wedge}} \underbrace{\frac{\left[1 + \frac{L_W}{L_C} \frac{F_C}{F_W} \left(\frac{(\varphi_{Cj}^*)^{dmkt}}{(\varphi_{W_j}^*)^{dmkt}} \right)^\gamma \right]}{\left[1 + \frac{L_W}{L_C} \frac{F_C}{F_W} \left(\frac{(\varphi_{Cj}^*)^{opt}}{(\varphi_{Wj}^*)^{opt}} \right)^\gamma \right]}}_{\text{Reallocation between origins}} \tag{23}$$

We decompose the destination-specific welfare distortion into three parts: the planner's objective, the competition-intensity wedge, and the reallocation between origins (resource bases). The first part arises due to the misalignment of pricing strategies between profit-maximizing firms and the utility-maximizing planner. This distortion is the only one that exists in a single-sector closed economy. The second term arises due to the destination-specific competition intensity, which is determined by the fundamentals of all countries selling to destination j. The third term measures the differences in fundamentals among all countries selling to j. The global allocational efficiency, $\frac{U^{dmkt}}{U^{opt}}$, the social planner's target, can be expressed as:

$$\begin{aligned} \frac{U^{dmkt}}{U^{opt}} &= \frac{\sum_j U_j^{dmkt}}{\sum_j U_j^{opt}} \\ &= \frac{U^{cmkt}}{U^{opt}} \left[\frac{(\varphi_{CC}^*)^{cmkt}}{(\varphi_{CC}^*)^{dmkt}} \right]^\gamma \frac{1 + \frac{L_W}{L_C} \frac{F_C}{F_W} \left[\frac{(\varphi_{CC}^*)^{dmkt}}{(\varphi_{WC}^*)^{dmkt}} \right]^\gamma + \frac{L_W}{L_C} \left[\frac{(\varphi_{CC}^*)^{dmkt}}{(\varphi_{CW}^*)^{dmkt}} \right]^\gamma + (\frac{L_W}{L_C})^2 \frac{F_C}{F_W} \left[\frac{(\varphi_{CC}^*)^{dmkt}}{(\varphi_{WW}^*)^{dmkt}} \right]^\gamma}{1 + \frac{L_W}{L_C} \frac{F_C}{F_W} \left[\frac{(\varphi_{CC}^*)^{opt}}{(\varphi_{WC}^*)^{opt}} \right]^\gamma + \frac{L_W}{L_C} \left[\frac{(\varphi_{CC}^*)^{opt}}{(\varphi_{CW}^*)^{opt}} \right]^\gamma + (\frac{L_W}{L_C})^2 \frac{F_C}{F_W} \left[\frac{(\varphi_{CC}^*)^{opt}}{(\varphi_{WW}^*)^{opt}} \right]^\gamma} \end{aligned} \tag{24}$$

A similar decomposition can be applied to measure the global allocation efficiency, but the last term measures reallocation between all countries. Additionally, it can be verified that $\frac{U^{cmkt}}{U^{opt}} = \frac{U_j^{cmkt}}{U_j^{opt}}$ for all countries j, indicating that the planner's objective is the only difference between the centralized market equilibrium and the social optimum. Using the parameters and cutoff ratios in Tables 4 and 5, we compute the country-level and global welfare distortions and present the results in Table 8.

Table 8: Welfare distortions in 2004-2007

	Year			
	2004	2005	2006	2007
U^{cmkt}/U^{opt}	0.8918	0.9109	0.9210	0.9296
U_C^{dmkt}/U_C^{opt}	1.0218	1.0632	0.9995	0.9341
U_W^{dmkt}/U_W^{opt}	0.8854	0.9028	0.9157	0.9292
U^{dmkt}/U^{opt}	0.8912	0.9103	0.9208	0.9296

We observe that the aggregate efficiency of the decentralized market equilibrium ranges from 89.12% to 92.96% of the one of the global social optimum for the world altogether, Hence, welfare losses due to misallocation range from 10.82% to 7.04% throughout the sample period with the chosen parametrization. Moreover, the ratios U^{cmkt}/U^{opt} and U^{dmkt}/U^{opt} are generally quite close to each other. This suggests that the global aggregate distortions are mainly driven by variable markups (misalignment with the planner's objective) rather than heterogeneous country-level competition intensities (wedge and reallocation across countries).

The results suggest that while the overall welfare effects of misallocation are similar when treating the world as integrated, there are significant differences in allocational inefficiencies across countries when benchmarked against a global social planner. Specifically, in 2004 and 2005, China gained 2.18% and 6.32% of a welfare surplus under the decentralized market relative to what a global social planner would have allocated, while the RoW lost 11.46% and 9.72%, respectively. In 2006 and 2007, the situation changes, and China lost about 0.05% and 6.59%, while the RoW lost 8.43% and 7.08%, respectively, relative to the allocation of a global social planner. Ignoring the effect of competition intensities would have led to non-negligible biases at the country level, especially for China. For instance, in 2006, dropping the competition intensity effect generates absolute errors of about 7.85 percentage points for China and -0.53 percentage points for the RoW.

Overall, a decentralized market equilibrium results in global inefficiency in resource allocation. However, the extent of misallocation across countries is influenced by the varying levels of competition intensity, which can lead to certain countries benefiting from the market system over a global planner. Failure to account for the distribution of misallocation across countries can result in an under- or overestimation of its impact on different countries. This discrepancy is particularly significant for China as compared to the RoW in the present application.

6 Conclusions

The question of misallocation appears particularly interesting with open economies. The reason is that a social planner, who is concerned with the maximization of utility for customers around the globe, faces a mismatch between the cross-national distribution of customers and the one of resources for output production at costly cross-border shipment of output. This mismatch creates a source of misallocation even with Pareto-distributed monopolistic competitor firms which charge a fixed markup. A situation with variable markups adds to the problem in a way which is qualitatively analogous to the one in closed economies. Overall, what is interesting with asymmetric open countries is that the profit maximization and market segmentation of firms which utilize resources from a common domestic source generate externalities across customers with different addresses.

The present paper characterizes this problem and derives a set of novel results to the end of misallocation with monopolistic competitors that face a variable elasticity of substitution between output varieties and heterogeneous markups over their variable costs. To start with, it introduces sufficient constraints for the space of all exogenous parameters to ensure the existence and uniqueness of a general equilibrium. It further obtains conditions for the extensive-margin (market entry) and intensive-margin (markup and quantity) components of the misallocation problem, which are specific to each customer-and-producer market pair. In open economies, the distortions resulting from misallocation can be characterized as shift-and-rotation effects on the production schedule as a function of firm productivity. Shift effects arise from variations in the marginal utility of income in different destinations, which are in turn determined by the asymmetric fundamentals of different countries. Output schedules are shifted (in quantity-productivity space) in opposite directions for domestic sales versus exports, and shifts are larger for exports than for domestic sales. The rotation effects, which are qualitatively similar to those in closed-economy models, cause all production schedules to rotate either clockwise or counterclockwise, depending on the properties of the demand structure.

The combination of the two effects does not permit unambiguous conclusions regarding the direction of the misallocation bias of the market (too weak versus too strong selection; which and how many firms over- versus under-produce relative to a utilitarian global social optimum). This is because shift effects counteract the rotation effects in specific economies.

We provide a specific parameterization with China versus the Rest of the World as two open economies, where firms are Pareto distributed and customers feature Constant Absolute Risk Aversion (CARA) preferences. When looking at the data between 2004 and 2007 through this lens, we find the following insights. First, the selection effects are too weak in both the domestic and the foreign from either economy's perspective, indicating the domination of quantity-schedule shift effects by the rotation effects. Second, low-productivity (high-productivity) firms tend to overproduce (underproduce) in both markets, with this effect being more pronounced in sales to China (the Rest of the World). Finally, we highlight the fact that disregarding the shift effects resulting from openness could lead to significant biases in the predicted distortions of individual countries regarding cutoffs, quantities, and welfare.

References

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

Bagwell, K. and S. Lee (2022). Monopolistic competition and efficiency under firm heterogeneity and non-additive preferences. Manuscript.

Bagwell, K. and S. H. Lee (2020). Trade policy under monopolistic competition with firm selection. *Journal of International Economics 127*, 103379.

Bagwell, K. and S. H. Lee (2021). Entry and welfare in general equilibrium with heterogeneous firms and endogenous markups. Manuscript.

Baqaee, D. and E. Farhi (2021). Darwinian returns to scale. Discussion Paper No. DP15712, Centre for Economic Policy Research.

Behrens, K., G. Mion, Y. Murata, and J. Südekum (2014). Trade, wages, and productivity. *International Economic Review 55*(4), 1305–1348.

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

Behrens, K. and Y. Murata (2012). Trade, competition, and efficiency. *Journal of International Economics 87*(1), 1–17.

Bertoletti, P. and F. Etro (2021). Monopolistic competition with generalized additively separable preferences. *Oxford Economic Papers 73*(2), 927–952.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–21.

Corless, R. M., G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth (1996). On the lambert w function. *Advances in Computational mathematics 5*, 329–359.

Dai, M. and J. Xu (2017). Firm-specific exchange rate shocks and employment adjustment:

Evidence from china. *Journal of International Economics 108*, 54–66.

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

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review 67*(3), 297–308.

Macedoni, L. and A. Weinberger (2022). Quality heterogeneity and misallocation: The welfare benefits of raising your standards. *Journal of International Economics 134*, 103544.

Matsuyama, K. and P. Ushchev (2022). Selection and sorting of heterogeneous firms through competitive pressures. Discussion Paper No. DP17092, Centre for Economic Policy Research.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2021). Product mix and firm productivity responses to trade competition. *Review of Economics and Statistics 103*(5), 874–891.

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

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *Review of Economic Studies 75*(1), 295–316.

Mrázová, M. and J. P. Neary (2017). Not so demanding: Demand structure and firm behavior. *American Economic Review 107*(12), 3835–74.

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: Papers and Proceedings 104*(5), 304–309.

Nocco, A., G. I. Ottaviano, and M. Salto (2019). Geography, competition, and optimal multilateral trade policy. *Journal of International Economics 120*, 145–161.

Yu, M. (2015). Processing trade, tariff reductions and firm productivity: Evidence from chinese firms. *The Economic Journal 125*(585), 943–988.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012). Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica 80*(6), 2765–2784.

A Online Appendix

Proof of Lemma 1. We first show the properties of optimal conditions under general equilibrium and then apply them under the counterfactual partial equilibrium. For brevity we refer to δ_j^{dmkt} by δ_j without ambiguity. Recall we assume $r_u(0) < 1$. Firms charge non-negative markups, so $r_u(q) \in [0, 1)$, $\forall q \in [0, +\infty)$. However, if $r'_u(q) > 0$, $r_u(q)$ might become greater than 1 as q increases. We therefore define: $\overline{q} \equiv \min\{q \geqslant 0 \text{ s.t. } r_u(q) = 1\}$. If $r_u(q) < 1$ for all $q > 0$, the respective $\overline{q} = +\infty$.[14]

From the FOCs, $\forall i, j = H, F$, we have

$$[1 - r_u(q_{ij})]\, u'(q_{ij}) = \frac{\delta_j \tau_{ij} w_i}{\varphi}. \tag{25}$$

Taking the derivative of the LHS w.r.t. q_{ij} obtains $\frac{\partial\{[1-r_u(q_{ij})]\cdot u'(q_{ij})\}}{\partial q_{ij}} = u''(q_{ij})\cdot[2 - r_{u'}(q_{ij})] < 0$,[15] where $r_{u'}(q_{ij}) \equiv -\frac{q_{ij}\cdot u'''(q_{ij})}{u''(q_{ij})}$. Recall we assume $\lim_{q_{ij}\to\overline{q}} [1 - r_u(q_{ij})]\, u'(q_{ij}) = 0$ and $\lim_{q_{ij}\to+\infty} u'(q_{ij}) = 0$.

The LHS of (25) could be bounded or not. When $\lim_{q_{ij}\to 0} u'(q_{ij}) = +\infty$, we have $\lim_{q_{ij}\to 0} [1 - r_u(q_{ij})]\, u'(q_{ij}) = +\infty$ and $\lim_{q_{ij}\to\overline{q}} [1 - r_u(q_{ij})]\, u'(q_{ij}) = 0$. Since the LHS goes from $+\infty$ to 0 as q_{ij} increases, a unique quantity $q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})$ is decreasing in $\delta_j w_i$ and $\frac{\tau_{ij}}{\varphi}$ for any $\delta_j w_i$ and $\frac{\tau_{ij}}{\varphi}$. When $u'(0)$ is finite, then $[1 - r_u(q_{ij})]\, u'(q_{ij})$ is bounded and $[1 - r_u(q_{ij})]\, u'(q_{ij}) \in [0, (1 - r_u(0))u'(0)]$. If $\frac{\tau_{ij}}{\varphi} \in (0, \frac{[1-r_u(0)]u'(0)}{\delta_j w_i}]$, we obtain the unique quantity $q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})$ that declines in $\delta_j w_i$ and $\frac{\tau_{ij}}{\varphi}$. If $\frac{\tau_{ij}}{\varphi} > \frac{[1-r_u(0)]u'(0)}{\delta_j w_i} \geqslant \frac{[1-r_u(q_{ij})]u'(q_{ij})}{\delta_j w_i}$ for all possible quantities q_{ij}, a firm's productivity is too low to face positive demand, and $q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = 0$. Besides, in both cases the unique $q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})$ satisfies

$$\lim_{\frac{\tau_{ij}}{\varphi}\to+\infty} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i\to+\infty} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = 0, \tag{26}$$

$$\lim_{\frac{\tau_{ij}}{\varphi}\to 0} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i\to 0} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \overline{q}. \tag{27}$$

When (25) holds, we can write a firm's profit divided by the wage in the decentralized

[14] Here, $\overline{q}$ depends on the setting of preferences. As preferences are identical across countries, no subscript is required for $\overline{q}$.

[15] Zhelobodko et al. (2012) assume $[2 - r_{u'}(q_{ij})] > 0$ as a "second-order condition" under constant unit costs.

equilibrium as:

$$\begin{aligned}
\tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) &= \left[\frac{1}{1-r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))} - 1\right] \frac{\tau_{ij}}{\varphi} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) L_j - f_{ij} \\
&= \frac{L_j}{\delta_j w_i} r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})) u'(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})) q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) - f_{ij} \\
&= \frac{L_j}{\delta_j w_i} r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})) \varepsilon_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})) u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})) - f_{ij} \\
&= \frac{L_j}{\delta_j w_i} \left[-q_{ij}^2(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) u''(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))\right] - f_{ij}.
\end{aligned}$$

Since $\frac{\partial\left[-q_{ij}^2 u''(q_{ij})\right]}{\partial q_{ij}} = -q_{ij}u''(q_{ij})\left[2 - r_{u'}(q_{ij})\right] > 0$, we obtain $\frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})} > 0$. Since $\frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \delta_j w_i} < 0$ and $\frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \frac{\tau_{ij}}{\varphi}} < 0$, we have

$$\frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \delta_j w_i} < 0 \text{ and } \frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \frac{\tau_{ij}}{\varphi}} < 0. \tag{28}$$

By definition, $0 \leqslant r_u(q)\varepsilon_u(q) \leqslant 1$ and $u(0) = 0$. Then, $\lim_{q\to 0} r_u(q)\varepsilon_u(q)u(q) = 0$. With (26), we obtain

$$\lim_{\frac{\tau_{ij}}{\varphi}\to+\infty} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i\to+\infty} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = -f_{ij}. \tag{29}$$

Because $r_u(q_{ij})u'(q_{ij})q_{ij} = -q_{ij}^2 u''(q_{ij})$ is increasing in q_{ij} and $q_{ij} \in [0, \overline{q}]$, we define the upper bound as $\overline{B} \equiv r_u(\overline{q})u'(\overline{q})\overline{q}$. Dhingra and Morrow (2019) focus on the case, where the utility function satisfies the Inada conditions, which is consistent with $\overline{B} = +\infty$. However, we wish to allow for $\overline{B}$ being finite. With (27), we obtain

$$\lim_{\frac{\tau_{ij}}{\varphi}\to 0} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \frac{L_j\overline{B}}{\delta_j w_i} - f_{ij} \text{ and } \lim_{\delta_j w_i\to 0} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = +\infty. \tag{30}$$

When $\delta_j w_i \to 0$, all firms make strictly positive profits, and the cutoff productivity approaches 0.

Combine (28), (29), and (30), if $\delta_j w_i \in (0, \frac{L_j\overline{B}}{f_{ij}}]$, we can solve for a unique cutoff $\varphi_{ij}^*(\delta_j w_i)$ such that $\tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi_{ij}^*}) = 0$. We refer to $\tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi_{ij}^*}) = 0$ as the zero-cutoff-profit condition, ZCPC. Evaluating (28) at $\varphi = \varphi_{ij}^*$, we can apply the implicit function theorem $\frac{d(\frac{\tau_{ij}}{\varphi_{ij}^*})}{d(\delta_j w_i)} = -\frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi_{ij}^*})}{\partial(\delta_j w_i)} / \frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi_{ij}^*})}{\partial(\frac{\tau_{ij}}{\varphi_{ij}^*})} < 0$ and obtain $\frac{d\varphi_{ij}^*}{d(\delta_j w_i)} > 0$. When $\delta_j w_i > \frac{L_j\overline{B}}{f_{ij}}$,

$\tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) < \frac{L_j}{\delta_j w_i} r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))u'(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) - \frac{L_j \overline{B}}{\delta_j w_i} < 0$, in which case no firm will operate. We can summarize the properties of cutoffs as

$$\lim_{\delta_j w_i \to 0} \varphi_{ij}^*(\delta_j w_i) = 0, \text{ and } \lim_{\delta_j w_i \to \frac{L_j \overline{B}}{f_{ij}}} \varphi_{ij}^*(\delta_j w_i) = +\infty. \tag{31}$$

The average profit divided by w_i from origin i to destination j reads

$$\tilde{\Pi}_{ij}(\delta_j w_i) = \int_{\varphi_{ij}^*(\delta_j w_i)}^{+\infty} \left[\frac{1}{1 - r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))} - 1 \right] \frac{\tau_{ij}}{\varphi} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) L_j - f_{ij} dG_i(\varphi).$$

From the ZCPC, we know that $\frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d\varphi_{ij}^*(\delta_j w_i)} = 0$. Then $\tilde{\Pi}_{ij}(\delta_j w_i)$ behaves as

$$\begin{aligned}\frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d(\delta_j w_i)} =& \frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d\varphi_{ij}^*(\delta_j w_i)} \frac{d\varphi_{ij}^*(\delta_j w_i)}{d(\delta_j w_i)} + \frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{dq_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})} \frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial(\delta_j w_i)} \\ =& \int_{\varphi_{ij}^*(\delta_j w_i)}^{+\infty} \left\{ \frac{r_u'(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\left[1 - r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))\right]^2} + \frac{r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))}{1 - r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))} \right\} \frac{\tau_{ij}}{\varphi} L_j \frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial(\delta_j w_i)} dG_i(\varphi).\end{aligned}$$

Given that $\frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial(\delta_j w_i)} < 0$, the derivative can be written after simplification as

$$\int_{\varphi_{ij}^*(\delta_j w_i)}^{+\infty} \left\{ \frac{\left[2 - r_{u'}(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))\right] r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))}{\left[1 - r_u(q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}))\right]^2} \right\} \frac{\tau_{ij}}{\varphi} L_j \frac{\partial q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial(\delta_j w_i)} dG_i(\varphi) < 0.$$

Therefore, $\frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d(\delta_j w_i)} < 0$. Consider the limit, with (29), (30), and (31), we obtain

$$\lim_{\delta_j w_i \to 0} \tilde{\Pi}_{ij}(\delta_j w_i) = +\infty \text{ and } \lim_{\delta_j w_i \to \frac{L_j \overline{B}}{f_{ij}}} \tilde{\Pi}_{ij}(\delta_j w_i) = 0. \tag{32}$$

Now we prove the existence and uniqueness of the counterfactual partial equilibrium $\{\delta_{ij} w_i, \varphi_{ij}^*, q_{ij}(\varphi)\}$, which satisfies the following equilibrium conditions:

$$\begin{cases} \left[1 - r_u(q_{ij}(\delta_{ij} w_i, \frac{\tau_{ij}}{\varphi}))\right] u'(q_{ij}(\delta_{ij} w_i, \frac{\tau_{ij}}{\varphi})) = \frac{\delta_{ij} \tau_{ij} w_i}{\varphi} \\ \tilde{\pi}_{ij}(\delta_{ij} w_i, \frac{\tau_{ij}}{\varphi_{ij}^*}) = f_{ij} \\ \tilde{\Pi}_{ij}(\delta_{ij} w_i) = F_i. \end{cases}$$

Since $\frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d(\delta_j w_i)} < 0$ and (32), there exists a unique $\delta_{ij} w_i$ s.t. $\tilde{\Pi}_{ij}(\delta_{ij} w_i) = F_i$, which implies a unique solution for $\varphi^*_{ij}(\delta_{ij}w_i)$ and $q_{ij}(\delta_{ij}w_i, \frac{\tau_{ij}}{\varphi})$. Therefore, for a given w_i, we obtain a unique solution for $\delta_{ij}(w_i)$. We can show the relation between δ_{ij} and τ_{ij}, f_{ij}, and F_i for a given wage. Since $\frac{\partial\tilde{\Pi}_{ij}}{\partial\tau_{ij}} < 0$ and $\frac{\partial\tilde{\Pi}_{ij}}{\partial f_{ij}} < 0$, the equilibrium value of δ_{ij} must decrease s.t. $\tilde{\Pi}_{ij}(\delta_{ij}w_i) = F_i$. Besides, with an increase of F_i and $\frac{\partial\Pi_{ij}}{\partial\delta_{ij}} < 0$, δ_{ij} must decrease to satisfy the counterfactual ZEPC. Therefore, for a given wage, we show that δ_{ij} is negatively related to τ_{ij}, f_{ij}, and F_i. Also, it is clear that $\delta_{ij}(w_i)$ decreases with w_i. □

Proof of Lemma 2. Recall that w_H is exogenous and $w_F \in (\underline{w_F}, \overline{w_F})$ s.t. $\delta_{HF} < \delta_{FF}(w_F)$ and $\delta_{FH}(w_F) < \delta_{HH}$ hold. We consider the standard ZEPCs for a given wage vector and rewrite $\delta_{ij}(w_i)$, $q_{ij}(\delta_{ij}w_i, \frac{\tau_{ij}}{\varphi})$, $\pi_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})$, $\varphi^*_{ij}(\delta_j w_i)$ and $\Pi_{ij}(\delta_j w_i)$ as δ_{ij}, $q_{ij}(\delta_{ij}, \frac{\tau_{ij}w_i}{\varphi})$, $\pi_{ij}(\delta_j, \frac{\tau_{ij}w_i}{\varphi})$, $\varphi^*_{ij}(\delta_j)$ and $\Pi_{ij}(\delta_j)$, respectively. One can see that all the above properties of these functions remain. Specifically, the ZEPCs are

$$\sum_j \int_{\varphi^*_{ij}(\delta_j)}^{+\infty} \left\{ \left[\frac{1}{1 - r_u(q_{ij}(\delta_j, \frac{\tau_{ij}w_i}{\varphi}))} - 1 \right] \frac{\tau_{ij}}{\varphi} q_{ij}(\delta_j, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij} \right\} dG_i(\varphi) = \sum_j \tilde{\Pi}_{ij}(\delta_j) = F_i. \tag{33}$$

For given wages, we have $\tilde{\Pi}_{ij}(\delta_{ij}) = F_i$. Consider the ZEPC in country F, when $\delta_H \geqslant \frac{L_H\overline{B}}{f_{FH}w_F}$, $\delta_F = \delta_{FF}$ with $\tilde{\Pi}_{FF}(\delta_F) = F_F$ and $\tilde{\Pi}_{FH}(\delta_H) = 0$. When $\delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$, according to the implicit function theorem, we can rewrite δ_H as $\delta^F_H(\delta_F)$ which guarantees that $\tilde{\Pi}_{FF}(\delta_F) + \tilde{\Pi}_{FH}(\delta^F_H(\delta_F)) = F_F$ holds and $\mathrm{d}\delta^F_H(\delta_F)/\mathrm{d}\delta_F = -\frac{\mathrm{d}\tilde{\Pi}_{FF}(\delta_F)/\mathrm{d}\delta_F}{\mathrm{d}\tilde{\Pi}_{FH}(\delta_H)/\mathrm{d}\delta_H} < 0$. When $\delta_F \geqslant \frac{L_F\overline{B}}{f_{FF}w_F}$, we can refine $\delta^F_H(\delta_F) = \delta_{FH}$, ensuring that $\tilde{\Pi}_{FF}(\delta_F) = 0$, $\tilde{\Pi}_{FH}(\delta_H) = F_F$, and $\mathrm{d}\delta^F_H(\delta_F)/\mathrm{d}\delta_F = 0$.

Similarly, consider the ZEPC in country H. when $\delta_H \geqslant \frac{L_H\overline{B}}{f_{HH}w_H}$, we obtain $\delta_F = \delta_{HF}$ with $\tilde{\Pi}_{HF}(\delta_F) = F_H$ and $\tilde{\Pi}_{HH}(\delta_H) = 0$. When $\delta_F \in (\delta_{HF}, +\infty)$, we can express δ_H as $\delta^H_H(\delta_F)$ with similar properties as above. □

Proof of Proposition 1. In this section, we first establish the existence and uniqueness of (δ^*_F, δ^*_H) for a given wage $w_F \in (\underline{w_F}, \overline{w_F})$, representing the partial equilibrium. One can see Figure 1 for intuitions. Afterwards, we prove the general equilibrium with endogenous wages. To simplify the notation in partial equilibrium, we express $\delta_{ij}(w_i)$ simply as δ_{ij}. However, we return to the full notation of $\delta_{ij}(w_i)$ when discussing the general equilibrium.

To begin, note that the partial equilibrium value of δ_F cannot be in $(\delta_{HF}, \delta_{FF})$ or $(\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$. Because when $\delta_F < \delta_{FF}$, $\tilde{\Pi}_{FF}(\delta_F) > F_F$, and when $\delta_F > \frac{L_F\overline{B}}{f_{FF}w_F}$, $\delta^H_H(\delta_F) \geqslant \delta_{HH} > \delta^F_H(\delta_F) = \delta_{FH}$.

We define the distance between two implicit functions as follows: $\forall \delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$,

$$\Delta(\delta_F) \equiv \delta^F_H(\delta_F) - \delta^H_H(\delta_F),$$

and further refine $\Delta(\delta_{FF}) = \max\left\{\lim_{\delta_F \to \delta_{FF}^+} \delta_H^F(\delta_F) - \delta_H^H(\delta_{FF}), 0\right\}$ to cover the scenario of δ_F decreasing to δ_{FF}. Consider the following cases:

1) If $\delta_{FF} < \frac{L_F \overline{B}}{f_{HF} w_H}$: according to Assumption 4, $\forall \delta_F \in (\delta_{FF}, \min\left\{\frac{L_F \overline{B}}{f_{HF} w_H}, \frac{L_F \overline{B}}{f_{FF} w_F}\right\})$: $\frac{\mathrm{d}\Delta(\delta_F)}{\mathrm{d}\delta_F} = \frac{\mathrm{d}\delta_H^F(\delta_F)}{\mathrm{d}\delta_F} - \frac{\mathrm{d}\delta_H^H(\delta_F)}{\mathrm{d}\delta_F} < 0$.

1.1) And if $\frac{L_F \overline{B}}{f_{HF} w_H} < \frac{L_F \overline{B}}{f_{FF} w_F}$: $\forall \delta_F \in (\frac{L_F \overline{B}}{f_{HF} w_H}, \frac{L_F \overline{B}}{f_{FF} w_F})$, $\frac{\mathrm{d}\Delta(\delta_F)}{\mathrm{d}\delta_F} = \frac{\mathrm{d}\delta_H^F(\delta_F)}{\mathrm{d}\delta_F} - 0 < 0$. At the boundary, $\Delta(\frac{L_F \overline{B}}{f_{FF} w_F}) = \delta_{FH} - \delta_{HH} < 0$. Therefore, $\Delta(\delta_F)$ is strictly decreasing in $(\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$. If $\Delta(\delta_{FF}) = 0$, $(\delta_{FF}, \delta_H^H(\delta_{FF}))$ is a solution of the partial equilibrium. Since $\lim_{\delta_F \to \delta_{FF}^+} \Delta(\delta_F) = \lim_{\delta_F \to \delta_{FF}^+} \delta_H^F(\delta_{FF}) - \delta_H^H(\delta_{FF}) \leqslant 0$ and the monotonicity of $\Delta(\delta_F)$, $\Delta(\delta_F) < 0$ holds in $(\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$, the uniqueness of the solution is established. If $\Delta(\delta_{FF}) > 0$, since $\Delta(\frac{L_F \overline{B}}{f_{FF} w_F}) < 0$, there exists a unique δ_F^* in $(\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$ s.t. $\Delta(\delta_F^*) = 0$.

1.2) And if $\frac{L_F \overline{B}}{f_{HF} w_H} \geqslant \frac{L_F \overline{B}}{f_{FF} w_F}$: recall that $\Delta(\delta_F) < 0 \ \forall \delta_F \geqslant \frac{L_F \overline{B}}{f_{FF} w_F}$, and $\Delta(\frac{L_F \overline{B}}{f_{FF} w_F}) < 0$. Since $\Delta(\delta_F)$ is strictly decreasing in $(\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$, as in case 1.1), if $\Delta(\delta_{FF}) = 0$, the unique solution is $(\delta_{FF}, \delta_H^H(\delta_{FF}))$. If $\Delta(\delta_{FF}) > 0$, there exists a unique solution in $(\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$ s.t. $\Delta(\delta_F^*) = 0$.

2) If $\delta_{FF} \geqslant \frac{L_F \overline{B}}{f_{HF} w_H}$, $\forall \delta_F \in (\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$, we have $\frac{\mathrm{d}\Delta(\delta_F)}{\mathrm{d}\delta_F} = \frac{\mathrm{d}\delta_H^F(\delta_F)}{\mathrm{d}\delta_F} - 0 < 0$. At the boundary, $\Delta(\frac{L_F \overline{B}}{f_{FF} w_F}) = \delta_{FH} - \delta_{HH} < 0$. Similarly, we obtain that if $\Delta(\delta_{FF}) = 0$, the unique solution is $(\delta_{FF}, \delta_{HH})$, and if $\Delta(\delta_{FF}) > 0$, there exists a unique $(\delta_F^*, \delta_{HH})$ s.t. $\Delta(\delta_F^*) = 0$.

Now consider the case of zero fixed costs. If $f_{ij} = 0, \forall i, j = H, F$, we directly solve for cutoff quantities $q_{ij}^* = 0$ from the ZCPCs regardless of the value of δ_j. Since $\lim_{\delta_j \to 0} \tilde{\pi}_{ij}(\delta_j, \frac{\tau_{ij} w_i}{\varphi}) = +\infty$ and $\lim_{\delta_j \to +\infty} \tilde{\pi}_{ij}(\delta_j, \frac{\tau_{ij} w_i}{\varphi}) = 0$, any values of δ_j guarantee non-negative profits of firms. From the FOCs $[1 - r_u(0)]\, u'(0) = \frac{\delta_j \tau_{ij} w_i}{\varphi_{ij}^*}$, when δ_j increases, the cutoff productivity φ_{ij}^* increases. With the ZEPCs $\tilde{\Pi}_{HH}(\delta_H^H(\delta_F)) + \tilde{\Pi}_{HF}(\delta_F) = F_H$ and $\tilde{\Pi}_{FF}(\delta_F) + \tilde{\Pi}_{FH}(\delta_H^F(\delta_F)) = F_F$, we obtain the following limit properties $\lim_{\delta_F \to +\infty} \delta_H^H(\delta_F) = \delta_{HH}$, $\lim_{\delta_F \to \delta_{HF}} \delta_H^H(\delta_F) = +\infty$, $\lim_{\delta_F \to +\infty} \delta_H^F(\delta_F) = \delta_{FH}$, and $\lim_{\delta_F \to \delta_{FF}} \delta_H^F(\delta_F) = +\infty$. Then we arrive at $\lim_{\delta_F \to +\infty} \Delta(\delta_F) = \delta_{FH} - \delta_{HH} < 0, \lim_{\delta_F \to \delta_{FF}} \Delta(\delta_F) = +\infty$ and $\frac{\mathrm{d}\Delta(\delta_F)}{d\delta_F} = \frac{\mathrm{d}\delta_H^F(\delta_F)}{\mathrm{d}\delta_F} - \frac{\mathrm{d}\delta_H^H(\delta_F)}{\mathrm{d}\delta_F} < 0$, $\forall \delta_F \in (\delta_{FF}, +\infty)$. With continuity, there exists a unique δ_F^* s.t. $\Delta(\delta_F^*) = 0$ and $\delta_H^F(\delta_F^*) = \delta_H^H(\delta_F^*) = \delta_H^*$.

Up to this point, we have demonstrated that $\forall w_F \in (\underline{w_F}, \overline{w_F})$, a unique solution (δ_F^*, δ_H^*) exists. From this, unique solutions for $\varphi_{ij}^*(\delta_j^*), q_{ij}(\delta_j^*, \frac{\tau_{ij} w_i}{\varphi}), M_i, \forall i, j = H, F$ follow.

In order to treat wages as endogenous, we express the solution (δ_F^*, δ_H^*) as a function of the given $w_F \in (\underline{w_F}, \overline{w_F})$, as $(\delta_F^*(w_F), \delta_H^*(w_F))$. We express the implicit $\delta_H^H(\delta_F)$ and $\delta_H^F(\delta_F)$ as $\delta_H^H(\delta_F, w_H)$ and $\delta_H^F(\delta_F, w_F)$, respectively, and the distance function $\Delta(\delta_F)$ as $\Delta(\delta_F, w_F)$.

The unique solution $(\delta_F^*(w_F), \delta_H^*(w_F))$ must then satisfy:

$$\begin{cases} \tilde{\Pi}_{HH}(\delta_H^*(w_F), w_H) + \tilde{\Pi}_{HF}(\delta_F^*(w_F), w_H) = F_H \\ \tilde{\Pi}_{FH}(\delta_H^*(w_F), w_F) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F) = F_F. \end{cases}$$

and $\Delta(\delta_F^*(w_F), w_F) = \delta_H^F(\delta_F^*(w_F), w_F) - \delta_H^H(\delta_F^*(w_F), w_H) = 0$. Furthermore, one can see $\forall \delta_F \in (\delta_{FF}, \delta_F^*(w_F))$, $\Delta(\delta_F, w_F) > 0$, and $\forall \delta_F \in (\delta_F^*(w_F), +\infty)$, $\Delta(\delta_F, w_F) < 0$.

Consider that w_F marginally decreases to a value w_F^-. When $\delta_F^*(w_F) \neq \delta_{FF}(w_F)$, since $\frac{\partial \tilde{\Pi}_{FH}}{\partial w_F} < 0$ and $\frac{\partial \tilde{\Pi}_{FF}}{\partial w_F} < 0$, $\tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-) > F_F$. Define the new implicit function of δ_H^F as $\delta_H^F(\delta_F, w_F^-)$ s.t. $\tilde{\Pi}_{FH}(\delta_H^F(\delta_F, w_F^-), w_F^-) + \tilde{\Pi}_{FF}(\delta_F, w_F^-) = F_F$. We evaluate the latter at $\delta_F = \delta_F^*(w_F)$ and obtain $\tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F^-), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-) = F_F < \tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-)$, which implies that $\delta_H^F(\delta_F^*(w_F), w_F^-) > \delta_H^F(\delta_F^*(w_F), w_F)$, because $\frac{\partial \tilde{\Pi}_{FH}}{\partial \delta_H} < 0$. Recall that $\delta_H^H(\delta_F, w_H)$ is independent of w_F, and deduce that: $\Delta(\delta_F^*(w_F), w_F^-) = \delta_H^F(\delta_F^*(w_F), w_F^-) - \delta_H^H(\delta_F^*(w_F), w_H) > \delta_H^F(\delta_F^*(w_F), w_F) - \delta_H^H(\delta_F^*(w_F), w_H) = 0$. Therefore, $\delta_F^*(w_F^-) > \delta_F^*(w_F)$ s.t. $\Delta(\delta_F^*(w_F^-), w_F^-) = 0$. Also, when $\delta_F^*(w_F) = \delta_{FF}(w_F)$, since $\delta_{FF}(w_F^-) > \delta_{FF}(w_F)$, $\delta_F^*(w_F^-) > \delta_F^*(w_F)$.

Recall that $\frac{\partial \delta_H^H(\delta_F, w_H)}{\partial \delta_F} < 0, \forall \delta_F \in (\delta_{HF}, \frac{L_F \overline{B}}{f_{HF} w_H})$ and $\frac{\partial \delta_H^H(\delta_F, w_H)}{\partial \delta_F} = 0, \forall \delta_F \in (\frac{L_F \overline{B}}{f_{HF} w_H}, +\infty)$. If $\delta_F^*(w_F) < \frac{L_F \overline{B}}{f_{HF} w_H}$, $\delta_H^*(w_F^-) = \delta_H^H(\delta_F^*(w_F^-), w_H) < \delta_H^H(\delta_F^*(w_F), w_H) = \delta_H^*(w_F)$ and, thus, $\frac{d\delta_H^*(w_F)}{dw_F} < 0$. If $\delta_F^*(w_F) \geqslant \frac{L_F \overline{B}}{f_{HF} w_H}$, $\delta_H^*(w_F^-) = \delta_H^H(\delta_F^*(w_F^-), w_H) = \delta_{HH} = \delta_H^H(\delta_F^*(w_F), w_H) = \delta_H^*(w_F)$ and, thus, $\frac{d\delta_H^*(w_F)}{dw_F} = 0$. At the limit, when $w_F \to \underline{w_F}$, $\delta_{FH}(w_F) \to \delta_{HH}$ and $\lim_{w_F \to \underline{w_F}} \delta_H^*(w_F) = \delta_{HH} = \delta_{FH}(\underline{w_F})$, indicating that firms in both countries only sell to country H.

Similarly, one can show when w_F increases to w_F^+, $\delta_F^*(w_F^+) < \delta_F^*(w_F)$. Besides, when $\delta_F^* > \frac{L_F \overline{B}}{f_{HF} w_H}$, $\delta_H^*(w_F^+) = \delta_H^*(\delta_F^*(w_F^+)) = \delta_{HH} = \delta_H^H(\delta_F^*(w_F)) = \delta_H^*(w_F)$. On the other hand, when $\delta_F^* \leqslant \frac{L_F \overline{B}}{f_{HF} w_H}$, $\delta_H^*(w_F^+) = \delta_H^*(\delta_F^*(w_F^+)) > \delta_H^*(\delta_F^*(w_F)) = \delta_H^*(w_F)$. At the limit, when $w_F \to \overline{w_F}$, $\delta_{FF}(w_F) \to \delta_{HF}$, $\lim_{w_F \to \overline{w_F}} \delta_F^*(w_F) = \delta_{FF}(\overline{w_F}) = \delta_{HF}$, showing that firms in both countries only sell to country F.

Now consider the trade-balance condition (TBC):

$$\begin{aligned} &M_H(\delta_F^*(w_F), \delta_H^*(w_F), w_H) \int_{\varphi_{HF}^*(\delta_F^*(w_F), w_H)}^{+\infty} p_{HF}(\delta_F^*(w_F) w_H, \frac{\tau_{HF}}{\varphi}) q_{HF}(\delta_F^*(w_F) w_H, \frac{\tau_{HF}}{\varphi}) L_F dG_H(\varphi) \\ =&M_F(\delta_F^*(w_F), \delta_H^*(w_F), w_F) \int_{\varphi_{FH}^*(\delta_H^*(w_F), w_F)}^{+\infty} p_{FH}(\delta_H^*(w_F) w_F, \frac{\tau_{FH}}{\varphi}) q_{FH}(\delta_H^*(w_F) w_F, \frac{\tau_{FH}}{\varphi}) L_H dG_F(\varphi). \end{aligned}$$

With the ZEPCs and the resource constraints (RCs), we obtain:

$$M_i(\delta_F^*(w_F),\delta_H^*(w_F),w_i)=\frac{L_i}{\sum_j\int_{\varphi_{ij}^*(\delta_j^*(w_F)w_i)}^{+\infty}\frac{1}{1-r_u(q_{ij}(\delta_j^*(w_F)w_i,\frac{\tau_{ij}}{\varphi}))}\frac{\tau_j}{\varphi}q_{ij}(\delta_j^*(w_F)w_i,\frac{\tau_{ij}}{\varphi})L_jdG_i(\varphi)}$$

Therefore, the TBC can be re-written as

$$\frac{L_Hw_H\int_{\varphi_{HF}^*(\delta_F^*(w_F)w_H)}^{+\infty}\frac{1}{1-r_u(q_{HF}(\delta_F^*(w_F)w_H,\frac{\tau_{HF}}{\varphi}))}\frac{\tau_{HF}}{\varphi}q_{HF}(\delta_F^*(w_F)w_H,\frac{\tau_{HF}}{\varphi})L_FdG_H(\varphi)}{\sum_j\int_{\varphi_{Hj}^*(\delta_j^*(w_F)w_H)}^{+\infty}\frac{1}{1-r_u(q_{Hj}(\delta_j^*(w_F)w_H,\frac{\tau_{Hj}}{\varphi}))}\frac{\tau_{Hj}}{\varphi}q_{Hj}(\delta_j^*(w_F)w_H,\frac{\tau_{Hj}}{\varphi})L_jdG_H(\varphi)}$$
$$=\frac{L_Fw_F\int_{\varphi_{FH}^*(\delta_H^*(w_F)w_F)}^{+\infty}\frac{1}{1-r_u(q_{FH}(\delta_H^*(w_F)w_F,\frac{\tau_{FH}}{\varphi}))}\frac{\tau_{FH}}{\varphi}q_{FH}(\delta_H^*(w_F)w_F,\frac{\tau_{FH}}{\varphi})L_HdG_F(\varphi)}{\sum_j\int_{\varphi_{Fj}^*(\delta_H^*(w_F)w_F)}^{+\infty}\frac{1}{1-r_u(q_{Fj}(\delta_j^*(w_F)w_F,\frac{\tau_{Fj}}{\varphi}))}\frac{\tau_{Fj}}{\varphi}q_{Fj}(\delta_j^*(w_F)w_F,\frac{\tau_{Fj}}{\varphi})L_jdG_F(\varphi)}$$

Recall that $\frac{d\delta_F^*(w_F)}{dw_F}<0$ and $\frac{d\delta_H^*(w_F)}{dw_F}\geqslant 0$. Furthermore, when $w_F\to\underline{w_F}$, $\lim_{w_F\to\underline{w_F}}\delta_H^*(w_F)=\delta_{HH}=\delta_{FH}(\underline{w_F})$, and when $w_F\to\overline{w_F}$, $\lim_{w_F\to\overline{w_F}}\delta_F^*(w_F)=\delta_{FF}(\overline{w_F})=\delta_{HF}$. Therefore, as w_F increases from $\underline{w_F}$ to $\overline{w_F}$, the LHS of the TBC increases from 0 t L_Hw_H, while the RHS of the TBC decreases from $L_F\underline{w_F}$ to 0.

Combining these results, we find a unique w_F^* s.t. the TBC holds. Given the wage vector (w_F^*,w_H), we can determine the solution $(\delta_F^*(w_F^*),\delta_H^*(w_F^*))$, as well as the corresponding quantities, cutoffs, and masses of entries. □

Proof of Proposition 2. For brevity, we refer to λ_i^{opt} by λ_i without ambiguity in this section. Recall that we require the social planner to reallocate resources under the market-equilibrium wages. Because of the concavity of the utility function, $\varepsilon_u(q)\in(0,1]$, $\forall q\in[0,+\infty)$. Consider the FOCs of the social planner's problem in country i:

$$u'(q_{ij})=\frac{\lambda_i\tau_{ij}w_i}{\varphi},\tag{34}$$

where we obtain $\frac{\partial LHS}{\partial q_{ij}}<0$ by concavity and the assumption of $\lim_{q_{ij}\to+\infty}u'(q_{ij})=0$. The LHS may be bounded or not. If $\lim_{q_{ij}\to 0}u'(q_{ij})=+\infty$, the LHS is unbounded and there exists a unique quantity function $q_{ij}(\lambda_i,\frac{\tau_{ij}w_i}{\varphi})$ that is decreasing in λ_i and $\frac{\tau_{ij}w_i}{\varphi}$. If $\lim_{q_{ij}\to 0}u'(q_{ij})<+\infty$, then the LHS is bounded and $u'(q_{ij})\in(0,u'(0)]$. So if $\frac{\tau_{ij}w_i}{\varphi}\in(0,\frac{u'(0)}{\lambda_i}]$, $q_{ij}(\lambda_i,\frac{\tau_{ij}w_i}{\varphi})$ is uniquely determined and is decreasing in λ_i and $\frac{\tau_{ij}w_i}{\varphi}$. But if $\frac{\tau_{ij}w_i}{\varphi}>\frac{u'(0)}{\lambda_i}$, $u'(q_{ij})<\frac{\lambda_i\tau_{ij}w_i}{\varphi},\forall q_{ij}\geqslant 0$. Hence, firms with productivity of $\varphi<\frac{\lambda_i\tau_{ij}w_i}{u'(0)}$ should not produce from a social planner's view so that their $q_{ij}(\lambda_i,\frac{\tau_{ij}w_i}{\varphi})=0$.

When (34) holds, we can write the social profit as

$$\begin{aligned}\pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) &= \left[\frac{1}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1\right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i \\ &= \frac{L_j}{\lambda_i}\left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) - f_{ij}w_i,\end{aligned}$$

where $\left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) = u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) - u'(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})$. Note that $\frac{\partial\left[1-\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right]u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})} = -q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})u''(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) > 0$ and $q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})$ is decreasing in $\frac{\tau_{ij}w_i}{\varphi}$ and λ_i. Therefore, we obtain

$$\frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \lambda_i} < 0 \ , \ \frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \frac{\tau_{ij}w_i}{\varphi}} < 0. \tag{35}$$

Besides, since $0 < \varepsilon_u(q_{ij}) \leqslant 1$ and $u(q_{ij}) = 0$, we obtain

$$\lim_{q_{ij}(\frac{\tau_{ij}w_i}{\varphi}, \lambda_i) \to 0} \left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) = 0,$$

and

$$\lim_{\frac{\tau_{ij}w_i}{\varphi} \to +\infty} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = \lim_{\lambda_i \to +\infty} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = -f_{ij}w_i. \tag{36}$$

We can further define

$$\overline{B}^s \equiv \lim_{q_{ij}(\frac{\tau_{ij}w_i}{\varphi}, \lambda_i) \to +\infty} \left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))$$

and obtain

$$\lim_{\frac{\tau_{ij}w_i}{\varphi} \to 0} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = \frac{L_j\overline{B}^s}{\lambda_i} - f_{ij}w_i \ , \ \lim_{\lambda_i \to 0} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty. \tag{37}$$

When $\lambda_i \to 0$, all firms in country i make strictly positive social profits from selling in country j, and the social cutoff $\varphi^*_{ij}(\lambda_i) \to 0$. Combing (35), (36), and (37), if $\lambda_i \in (0, \frac{L_j\overline{B}^s}{f_{ij}w_i}]$, then a unique cutoff $\varphi^*_{ij}(\lambda_i)$ s.t. $\pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}}) = 0$ exists. Below, we refer to $\pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}}) = 0$ as the zero-cutoff-social-profit condition, ZCSPC. With (35), we can apply the implicit function theorem $\frac{\mathrm{d}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\mathrm{d}\lambda_i} = -\frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\partial \lambda_i} / \frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\partial (\frac{\tau_{ij}w_i}{\varphi^*_{ij}})} < 0$ and obtain $\frac{d\varphi^*_{ij}}{d\lambda_i} > 0$. If $\lambda_i > \frac{L_j\overline{B}^s}{f_{ij}w_i}$, $\pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}}) < \frac{L_j}{\lambda_i}\left[1 - \varepsilon_u(q_{ij})\right] u(q_{ij}) - \frac{L_j\overline{B}^s}{\lambda_i} \leqslant 0$ for all possible quantities so that no firm in i will sell to j.

The ZESPC reads:

$$\sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} \left\{ \left[\frac{1}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1 \right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i \right\} dG_i(\varphi) = F_i w_i \quad (38)$$

With the ZCSPCs, $\frac{dLHS}{d\varphi^*_{ij}(\lambda_i)}$=0. Further differentiate the LHS w.r.t λ_i:

$$\begin{aligned}
\frac{\partial LHS}{\partial \lambda_i} =& \sum_j \frac{dLHS}{d\varphi^*_{ij}(\lambda_i)} \frac{d\varphi^*_{ij}(\lambda_i)}{d\lambda_i} + \sum_j \frac{dLHS}{dq_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})} \frac{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \lambda_i} \\
=& \sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} L_j \frac{\tau_{ij}w_i}{\varphi} \left\{ - \frac{q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\left[\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right]^2} \frac{\partial \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})} + \frac{1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} \right\} \\
& \times \frac{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \lambda_i} dG_i(\varphi).
\end{aligned}$$

Note that $-\frac{q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\left[\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))\right]^2} \frac{\partial \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})} + \frac{1-\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} = \frac{r_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} > 0$. Therefore, with $\frac{\partial q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \lambda_i} < 0$, we obtain $\frac{\partial LHS}{\partial \lambda_i} < 0$. To explore the range of the LHS of (38), rewrite it as

$$\sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} \frac{L_j}{\lambda_i} \left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) \right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) - f_{ij}w_i dG_i(\varphi) = F_i w_i.$$

When $\lambda_i \to 0$, $\forall j = H, F$, $\varphi^*_{ij} \to 0$, $q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) \to +\infty$ and, hence,

$$\begin{aligned}
& \lim_{\lambda_i \to 0} \sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} \frac{L_j}{\lambda_i} \left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) \right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) - f_{ij}w_i dG_i(\varphi) \\
=& \lim_{\lambda_i \to 0} \sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} \frac{L_j \overline{B}^s}{\lambda_i} - f_{ij}w_i dG_i(\varphi) = +\infty.
\end{aligned}$$

We then define

$$\underline{j} \equiv \left\{ j \mid \frac{L_j}{f_{ij}} = \min \left\{ \frac{L_j}{f_{ij}}, \forall j = H, F \right\} \right\}, \overline{j} \equiv \left\{ j \mid \frac{L_j}{f_{ij}} = \max \left\{ \frac{L_j}{f_{ij}}, \forall j = H, F \right\} \right\}.$$

When $\lambda_i \to \frac{L_{\underline{j}} \overline{B}^s}{f_{i\underline{j}} w_i}$, $\varphi^*_{i\underline{j}}(\lambda_i) \to +\infty$, $q_{i\underline{j}(\lambda_i, \frac{\tau_{i\underline{j}} w_i}{\varphi})} \to 0$ and $\left[1 - \varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{\underline{i}j}w_i}{\varphi})) \right] u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})) \to$

0. Therefore,

$$\lim_{\lambda_i \to \frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}} \int_{\varphi^*_{i\underline{j}}(\lambda_i)}^{+\infty} \frac{L_{\underline{j}}}{\lambda_i} \left[1 - \varepsilon_u(q_{i\underline{j}}(\lambda_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))\right] u(q_{i\underline{j}}(\lambda_i, \frac{\tau_{i\underline{j}}w_i}{\varphi})) - f_{i\underline{j}}w_i dG_i(\varphi) = 0.$$

However, the LHS of (38) does not converge to 0, since $\frac{L_{\bar{j}}\overline{B}^s}{f_{i\bar{j}}w_i} > \frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}$, $\lim_{\delta_i \to \frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}} \varphi^*_{i\bar{j}}(\lambda_i) < +\infty$, and $\lim_{\lambda_i \to \frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}} q_{i\bar{j}}(\lambda_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}) > 0$.

When $\lambda_i > \frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}$ and $\lambda_i \to \frac{L_{\bar{j}}\overline{B}^s}{f_{i\bar{j}}w_i}$,

$$\int_{\varphi^*_{i\underline{j}}(\lambda_i)}^{+\infty} \frac{L_{\underline{j}}}{\lambda_i} \left[1 - \varepsilon_u(q_{i\underline{j}}(\lambda_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))\right] u(q_{i\underline{j}}(\lambda_i, \frac{\tau_{i\underline{j}}w_i}{\varphi})) - f_{i\underline{j}}w_i dG_i(\varphi) = 0.$$

Then, firms in country i do not sell to country $\underline{j}$, as their profits would be negative. Besides, $\varphi^*_{i\bar{j}}(\lambda_i) \to +\infty$, $q_{i\bar{j}}(\lambda_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}) \to 0$ and $\left[1 - \varepsilon_u(q_{ij}(\delta_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}))\right] u(q_{i\bar{j}}(\delta_i, \frac{\tau_{i\bar{j}}w_i}{\varphi})) \to 0$. Hence, $\int_{\varphi^*_{i\bar{j}}(\lambda_i)}^{+\infty} \frac{L_{\bar{j}}}{\lambda_i} \left[1 - \varepsilon_u(q_{i\bar{j}}(\lambda_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}))\right] u(q_{i\bar{j}}(\lambda_i, \frac{\tau_{i\bar{j}}w_i}{\varphi})) - f_{i\bar{j}}w_i dG_i(\varphi) \to 0$ and the $LHS \to 0$.

In sum, as λ_i goes from 0 to $\frac{L_{\bar{j}}\overline{B}^s}{f_{i\bar{j}}w_i}$, the LHS of (38) decreases from $+\infty$ to 0. Therefore, $\forall F_i w_i > 0$, there exists a unique $\lambda_i^{opt} > 0$ s.t. the ZESPC (38) for country i being satisfied, and the associated cutoff $\varphi^*_{ij}(\lambda_i^{opt})$ and $q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij}w_i}{\varphi})$ are determined. The masses of entrants can then be solved from the resource constraint.

Lastly, consider the case of zero fixed costs. If $f_{ij} = 0, \forall i, j = H, F$, we can directly solve for the cutoff quantity $q^*_{ij} = 0$ from the ZCSPCs for any i, j. Since $\lim_{\lambda_i \to 0} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty$ and $\lim_{\lambda_i \to +\infty} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = 0$, all values of λ_i guarantee non-negative social profits. From the FOCs, we obtain $u'(0) = \frac{\lambda_i \tau_{ij} w_i}{\varphi^*_{ij}}$, indicating that the cutoff $\varphi^*_{ij}(\lambda_i)$ is positively related to λ_i. Consider the ZESPCs in (38), where the LHS decreases with λ_i. We have LHS $\to +\infty$ when $\lambda_i \to 0$, and LHS $\to 0$ when $\lambda_i \to +\infty$. Therefore, $\forall F_i w_i > 0$, there exists a unique λ_i^{opt} s.t. the ZESPC is satisfied, and the corresponding equilibrium outcomes can be obtained. □

Proof of Proposition 3. For brevity, we refer to δ_i^{cmkt} by δ_i without ambiguity in this section. By design, all FOCs of the centralized problem are the same as those of the decentralized problem, except for the Lagrange multipliers being indexed by the location of the producers i instead of the consumers j. By design, we require the centralized 8planner to reallocate resources under the market-equilibrium wages. Therefore, we obtain the following properties with similar steps from the proof of Lemma 1.

From the FOCs $[1-r_u(q_{ij})]\,u'(q_{ij}) = \frac{\delta_i \tau_{ij} w_i}{\varphi}$, we obtain a unique quantity function $q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})$ satisfying (1) $\frac{\partial q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \delta_i} < 0$ and $\frac{\partial q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \frac{\tau_{ij}w_i}{\varphi}} < 0$; (2) $\lim_{\frac{\tau_{ij}w_i}{\varphi}\to+\infty} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \lim_{\delta_i\to+\infty} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = 0$; (3) $\lim_{\frac{\tau_{ij}w_i}{\varphi}\to 0} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \lim_{\delta_i\to 0} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \overline{q}$.

When the FOCs hold, we can further write a firm's profit in the centralized equilibrium as $\pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \left[\frac{1}{1-r_u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1\right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i$ and recall the definition $\overline{B} = r_u(\overline{q})u'(\overline{q})\overline{q}$. Then, we obtain the following properties of $\pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})$: (1) $\frac{\partial \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})} > 0$, $\frac{\partial \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \delta_i} < 0$ and $\frac{\partial \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \frac{\tau_{ij}w_i}{\varphi}} < 0$; (2) $\lim_{\frac{\tau_{ij}w_i}{\varphi}\to+\infty} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \lim_{\delta_i\to+\infty} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = -f_{ij}w_i$; (3) $\lim_{\frac{\tau_{ij}w_i}{\varphi}\to 0} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \frac{L_j\overline{B}}{\delta_i} - f_{ij}w_i$, $\lim_{\delta_j\to 0} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty$. Further, for $\delta_i \in (0, \frac{L_j\overline{B}}{f_{ij}w_i}]$, we can solve for a unique cutoff $\varphi^*_{ij}(\delta_i)$ s.t. $\pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}}) = 0$ and $\frac{d\varphi^*_{ij}}{d\delta_i} > 0$.

Consider the ZEPCs, $\forall i = H, F$:

$$\sum_j \Pi_{ij}(\delta_i) = \sum_j \int_{\varphi^*_{ij}(\delta_i)}^{+\infty} \left\{ \left[\frac{1}{1-r_u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1\right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i \right\} dG_i(\varphi) = F_i w_i. \tag{39}$$

We can obtain $\frac{\partial \Pi_{ij}(\delta_i)}{\partial \delta_i} < 0$. Thus, the LHS of (39) is decreasing in δ_i. We further explore the possible range of the LHS by rewriting the ZEPC (39) as:

$$\sum_j \int_{\varphi^*_{ij}(\delta_i)}^{+\infty} \frac{L_j}{\delta_i} r_u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}))\varepsilon_u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}))u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi})) - f_{ij}w_i dG_i(\varphi) = F_i w_i.$$

When $\delta_i \to 0$, the cutoff productivity $\varphi^*_{ij}(\delta_i) \to 0$, the cutoff quantity $q(\delta_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}}) \to \overline{q}$, and $r_u(\overline{q})\varepsilon_u(\overline{q})u(\overline{q}) \to \overline{B}$. Therefore, we obtain $\lim_{\delta_i\to 0} \sum_j \int_{\varphi^*_{ij}(\delta_i)}^{+\infty} (\frac{L_j\overline{B}}{\delta_i} - f_{ij}w_i) dG_i(\varphi) = +\infty$, indicating that the LHS $\to +\infty$.

Recall that

$$\underline{j} = \left\{ j \mid \frac{L_j}{f_{ij}} = \min\left\{\frac{L_j}{f_{ij}}, \forall j = H, F\right\}\right\}, \overline{j} = \left\{ j \mid \frac{L_j}{f_{ij}} = \max\left\{\frac{L_j}{f_{ij}}, \forall j = H, F\right\}\right\}.$$

When $\delta_i \to \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}$, $\varphi^*_{i\underline{j}} \to +\infty$, $q_{i\underline{j}} \to 0$ and $r_u(q_{i\underline{j}})\varepsilon_u(q_{i\underline{j}})u(q_{i\underline{j}}) \to 0$. Therefore,

$$\lim_{\delta_i \to \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}} \int_{\varphi^*_{i\underline{j}}(\delta_i)}^{+\infty} \frac{L_{\underline{j}}}{\delta_i} r_u(q_{ij}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))\varepsilon_u(q_{i\underline{j}}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))u(q_{i\underline{j}}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi})) - f_{i\underline{j}}w_i dG_i(\varphi) = 0.$$

However, the LHS will not converge to 0 since $\frac{L_{\bar{j}}\overline{B}}{f_{i\bar{j}}w_i} > \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}$, $\lim_{\delta_i \to \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}} \varphi^*_{i\bar{j}} < +\infty$, and $\lim_{\delta_i \to \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}} q_{i\bar{j}} > 0$. When $\delta_i > \frac{L_{\underline{j}}\overline{B}}{f_{i\underline{j}}w_i}$ and $\delta_i \to \frac{L_{\bar{j}}\overline{B}}{f_{i\bar{j}}w_i}$,

$$\int_{\varphi^*_{i\underline{j}}(\delta_i)}^{+\infty} \frac{L_{\underline{j}}}{\delta_i} r_u(q_{ij}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))\varepsilon_u(q_{i\underline{j}}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi}))u(q_{i\underline{j}}(\delta_i, \frac{\tau_{i\underline{j}}w_i}{\varphi})) - f_{i\underline{j}}w_i dG_i(\varphi) = 0,$$

indicating that firms in country i will not sell to country $\underline{j}$, as their profits would be negative. Besides, $\varphi^*_{i\bar{j}} \to +\infty$, $q_{i\bar{j}} \to 0$ and $r_u(q_{i\bar{j}})\varepsilon_u(q_{i\bar{j}})u(q_{i\bar{j}}) \to 0$. Hence,

$$\lim_{\delta_i \to \frac{L_{\bar{j}}\overline{B}}{f_{i\bar{j}}w_i}} \int_{\varphi^*_{i\bar{j}}(\delta_i)}^{+\infty} \frac{L_{\bar{j}}}{\delta_i} r_u(q_{ij}(\delta_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}))\varepsilon_u(q_{i\bar{j}}(\delta_i, \frac{\tau_{i\bar{j}}w_i}{\varphi}))u(q_{i\bar{j}}(\delta_i, \frac{\tau_{i\bar{j}}w_i}{\varphi})) - f_{i\bar{j}}w_i dG_i(\varphi) = 0,$$

and the LHS of (39) $\to 0$.

Overall, as δ_i increases from 0 to $\frac{L_{\bar{j}}\overline{B}}{f_{i\bar{j}}w_i}$, the LHS strictly decreases from $+\infty$ to 0. Therefore, $\forall F_i w_i > 0$, there exists a unique $\delta_i^{cmkt} > 0$ s.t. $\sum_j \Pi_{ij}(\delta_i^{cmkt}) = F_i w_i$, which implies a unique cutoff productivity $(\varphi^*_{ij})^{cmkt} = \varphi^*_{ij}(\delta_i^{cmkt})$ and quantity function $q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}) = q_{ij}(\delta_i^{cmkt}, \frac{\tau_{ij}w_i}{\varphi})$, $\forall j = H, F$. We can further solve for the masses of entrants M_i from the resource constraints.

Let us consider the case of zero fixed costs. If $f_{ij} = 0, \forall i, j = H, F$, we can directly solve for the cutoff quantity $q^*_{ij} = 0$ from ZCPCs for any i, j. Since $\lim_{\delta_i \to 0} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty$ and $\lim_{\delta_i \to +\infty} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = 0$, all values of δ_i guarantee non-negative profits. From the FOCs, $[1 - r_u(0)]\, u'(0) = \frac{\delta_i \tau_{ij} w_i}{\varphi^*_{ij}}$. Hence, when δ_i increases, the cutoff $\varphi^*_{ij}(\delta_i)$ increases. Recall that the LHS of the ZEPCs (39) decreases with δ_i, the LHS $\to +\infty$ when $\delta_i \to 0$, and the LHS $\to 0$ when $\delta_i \to +\infty$. Therefore, $\forall F_i w_i > 0$, there exists a unique δ_i^{cmkt} s.t. the ZEPC being satisfied, and then we can solve for the cutoff $\varphi^*_{ij}(\delta_i^{cmkt})$, quantity function $q_{ij}(\delta_i^{cmkt}, \frac{\tau_{ij}w_i}{\varphi})$, and the masses of entrants M_i, $\forall i, j = H, F$. □

Proof of Proposition 4. We assume that the solution of the decentralized equilibrium is $(\delta_F^{dmkt}, \delta_H^{dmkt})$ where $\delta_H^{dmkt} > \delta_F^{dmkt}$. From above, we know $\delta_F^{dmkt} \in [\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$ and $\delta_H^H(\delta_F^{dmkt}) = \delta_H^F(\delta_F^{dmkt}) = \delta_H^{dmkt}$. At the wage of the decentralized equilibrium (w_F, w_H), the centralized equilibrium $(\delta_F^{cmkt}, \delta_H^{cmkt})$ satisfies $\delta_H^H(\delta_H^{cmkt}) = \delta_H^{cmkt}$ and $\delta_H^F(\delta_F^{cmkt}) = \delta_F^{cmkt}$.

Since $\delta_H^H(\delta_F^{dmkt}) = \delta_H^{dmkt} > \delta_F^{dmkt}$ and $\delta_H^H(\delta_F)$ is strictly decreasing in $(\delta_{HF}, \frac{L_F\overline{B}}{f_{HF}w_H})$ and constant in $(\frac{L_F\overline{B}}{f_{HF}w_H}, +\infty)$, $\delta_H^{cmkt} > \delta_F^{dmkt}$ must hold such that $\delta_H^H(\delta_H^{cmkt}) = \delta_H^{cmkt}$. Besides, $\delta_H^H(\delta_H^{cmkt}) = \delta_H^{cmkt} \leqslant \delta_H^H(\delta_F^{dmkt}) = \delta_H^{dmkt}$.

If $\delta_F^{dmkt} = \delta_{FF}$, with the TBC, neither country exports, $\delta_H^H(\delta_{FF}) = \delta_H^{dmkt} = \delta_{HH} > \delta_{FF}$

and $\Delta(\delta_{FF}) = 0$. Then, $\forall \delta_F > \delta_{FF}$, $\Delta(\delta_{FF}) < 0$. Therefore, we obtain $\Delta(\delta_H^{cmkt}) = \delta_H^F(\delta_H^{cmkt}) - \delta_H^H(\delta_H^{cmkt}) = \delta_H^F(\delta_H^{cmkt}) - \delta_H^{cmkt} < 0$. Since $\delta_H^F(\delta_F)$ is strictly decreasing in $(\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$ and constant in $(\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$, $\delta_F^{cmkt} < \delta_H^{cmkt}$ must hold s.t. $\delta_H^F(\delta_F^{cmkt}) = \delta_F^{cmkt}$. Besides, $\delta_F^{cmkt} \geqslant \delta_{FF} = \delta_F^{dmkt}$ by the definition of $\delta_H^F(\delta_F)$. Hence, we show that $\delta_H^{dmkt} \geqslant \delta_H^{cmkt} > \delta_F^{cmkt} \geqslant \delta_F^{dmkt}$.

If $\delta_F^{dmkt} > \delta_{FF}$, with the TBC, $\delta_H^{dmkt} > \delta_{HH}$. Recall that $\delta_H^F(\delta_F^{dmkt}) = \delta_H^{dmkt} > \delta_F^{dmkt}$. Since $\delta_H^F(\delta_F)$ is strictly decreasing in $(\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$ and constant in $(\frac{L_F\overline{B}}{f_{FF}w_H}, +\infty)$, $\delta_F^{cmkt} > \delta_F^{dmkt}$. Similarly, $\delta_H^H(\delta_F)$ is strictly decreasing in $(\delta_{HF}, \frac{L_F\overline{B}}{f_{HF}w_H})$ and constant in $(\frac{L_F\overline{B}}{f_{HF}w_H}, +\infty)$, and $\delta_H^H(\delta_F^{dmkt}) = \delta_H^{dmkt} > \delta_H^H(\delta_H^{cmkt}) = \delta_H^{cmkt}$. Furthermore, since $\delta_F^{dmkt} \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$, in which $\Delta(\delta_F)$ is strictly decreasing, $\Delta(\delta_F^{cmkt}) = \delta_H^F(\delta_F^{cmkt}) - \delta_H^H(\delta_F^{cmkt}) = \delta_F^{cmkt} - \delta_H^H(\delta_F^{cmkt}) < \Delta(\delta_F^{dmkt}) = 0$. We then know $\delta_F^{cmkt} < \delta_H^{cmkt}$ from $\delta_F^{cmkt} < \delta_H^H(\delta_F^{cmkt})$ and $\delta_F^{cmkt} > \delta_F^{dmkt}$ from $\Delta(\delta_F^{cmkt}) < \Delta(\delta_F^{dmkt})$. Hence, we show that $\delta_H^{dmkt} > \delta_H^{cmkt} > \delta_F^{cmkt} > \delta_F^{dmkt}$.

From the viewpoint of country j, let us generically refer to δ_j^{dmkt} and δ_i^{cmkt} by δ in the following statement. Then, for both the decentralized and centralized equilibria, the cutoff behaves as $\frac{d\varphi_{ij}^*(\delta)}{d\delta} > 0$ for $\delta \in (0, \frac{L_j\overline{B}}{f_{ij}w_i})$ and $\frac{d\varphi_{ij}^*(\delta)}{d\delta} = 0$ for $\delta \in (\frac{L_j\overline{B}}{f_{ij}w_i}, +\infty)$, and the quantity behaves as $\frac{dq_{ij}(\delta,\varphi)}{d\delta} < 0$ for $\delta \in (0, \frac{L_j\overline{B}}{f_{ij}w_i})$ and $\frac{dq_{ij}(\delta,\varphi)}{d\delta} = 0$ for $\delta \in (\frac{L_j\overline{B}}{f_{ij}w_i}, +\infty)$. Hence, the comparisons in quantity and cutoff productivity can be directly obtained. $\square$

Proof of Proposition 5. Define $\overline{\sigma} \equiv \sup\{\varepsilon_u(q)|q \geqslant 0\}$ and $\underline{\sigma} \equiv \inf\{\varepsilon_u(q)|q \geqslant 0\}$. From the FOCs of the centralized market equilibrium, we have:

$$\begin{aligned}
\delta_i^{cmkt} w_i &= \frac{(M_i)^{cmkt}}{L_i} \sum_j L_j \int_{(\varphi_{ij}^*)^{cmkt}}^{+\infty} u'(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi})dG_i(\varphi) \\
&= \frac{(M_i)^{cmkt}}{L_i} \sum_j L_j \int_{(\varphi_{ij}^*)^{cmkt}}^{+\infty} \varepsilon_u(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))u(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\delta_i^{cmkt} w_i}{\overline{\sigma}} &= \frac{(M_i)^{cmkt}}{L_i} \sum_j L_j \int_{(\varphi_{ij}^*)^{cmkt}}^{+\infty} \frac{\varepsilon_u(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))u(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))}{\overline{\sigma}}dG_i(\varphi) \\
&< \frac{(M_i)^{cmkt}}{L_i} \sum_j L_j \int_{(\varphi_{ij}^*)^{cmkt}}^{+\infty} u(q_{ij}^{cmkt}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi) \\
&< \frac{(M_i)^{opt}}{L_i} \sum_j L_j \int_{(\varphi_{ij}^*)^{opt}}^{+\infty} u(q_{ij}^{opt}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi) = (\lambda_i)^{opt} w_i,
\end{aligned}$$

and we obtain $\frac{\delta_i^{cmkt}}{\overline{\sigma}} < \lambda_i^{opt}$.

Similarly, from the FOCs of the social planner's problem, we have

$$\underline{\sigma}\lambda_i^{opt}w_i = \frac{(M_i)^{opt}}{L_i}\sum_j L_j \int_{(\varphi_{ij}^*)^{opt}}^{+\infty} u(q_{ij}^{opt}(\frac{\tau_{ij}w_i}{\varphi}))\underline{\sigma}dG_i(\varphi)$$
$$< \frac{(M_i)^{opt}}{L_i}\sum_j L_j \int_{(\varphi_{ij}^*)^{opt}}^{+\infty} u'(q_{ij}^{opt}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}^{opt}(\frac{\tau_{ij}w_i}{\varphi})dG_i(\varphi) < \delta_i^{cmkt}w_i.$$

We, hence, obtain

$$0 \leqslant \underline{\sigma} < \frac{\delta_i^{cmkt}}{\lambda_i^{opt}} < \overline{\sigma} \leqslant 1,$$

which shows that $0 < \delta_i^{cmkt} < \lambda_i^{opt} < 1$.

To start with, consider $(1 - r_u(q))' < 0$ and $\varepsilon_u'(q) < 0$, indicating that both the market markup and the social markup increase with quantity. Since $\lim_{q\to 0}\varepsilon_u(q) > 0$, employing L'Hôpital's rule, we obtain $\lim_{q\to 0}\varepsilon_u(q) = \lim_{q\to 0} 1 - r_u(q) > 0$ and, hence, $\sup_{q\geqslant 0}(1 - r_u(q)) = \sup_{q\geqslant 0}\varepsilon_u(q)$. Besides, if $\lim_{q\to+\infty}\varepsilon_u(q) > 0$, we obtain $\lim_{q\to+\infty} 1 - r_u(q) = \lim_{q\to+\infty}\varepsilon_u(q)$ by L'Hôpital's rule. If $\lim_{q\to+\infty}\varepsilon_u(q) = 0$, we have $\lim_{q\to+\infty} 1 - r_u(q) \leqslant 0$ by assumption. If $\lim_{q\to+\infty} 1 - r_u(q) < 0$, we define $\overline{q} \equiv \min\{q \geqslant 0 \text{ s.t.} r_u(q) = 1\}$. Because of the positive-markup assumption, firms will produce less than $\overline{q}$ in a centralized market equilibrium. If $\lim_{q\to+\infty} 1 - r_u(q) = 0$, then $\overline{q} = +\infty$ and $q \in [0, +\infty)$. Hence, we obtain $\lim_{q\to+\infty}\varepsilon_u(q) = \lim_{q\to\overline{q}} 1 - r_u(q)$ and $\inf_{q\in[0,\overline{q})}(1 - r_u(q)) = \inf_{q\geqslant 0}\varepsilon_u(q)$.

In what follows, we consider full support of φ irrespective of the cutoff and purely rely on the FOCs w.r.t. quantity of two equilibria. $\forall i, j = H, F$ and $\forall\varphi > 0$:

$$\begin{cases} \left[u''(q_{ij}^{cmkt})q_{ij}^{cmkt} + u'(q_{ij}^{cmkt})\right] = \dfrac{\delta_i^{cmkt}\tau_{ij}w_i}{\varphi} \\ u'(q_{ij}^{opt}) = \dfrac{\lambda_i^{opt}\tau_{ij}w_i}{\varphi}, \end{cases} \tag{40}$$

from where we obtain the two implicit quantity functions, which we refer to as $q_{ij}^{cmkt}(\varphi)$ and $q_{ij}^{opt}(\varphi)$ without ambiguity.

Combining the two FOCs (40), we obtain:

$$\frac{\left[1 - r_u(q_{ij}^{cmkt}(\varphi))\right] u'(q_{ij}^{cmkt}(\varphi))}{u'(q_{ij}^{opt}(\varphi))} = \frac{\delta_i^{cmkt}}{\lambda_i^{opt}}.$$

Since $\sup_{q\in[0,\overline{q})}(1 - r_u(q)) = \sup_{q\geqslant 0}\varepsilon_u(q)$ and $\inf_{q\in[0,\overline{q})}(1 - r_u(q)) = \inf_{q\geqslant 0}\varepsilon_u(q)$, we obtain

$\forall q \in [0, \overline{q})$:

$$\sup_{q\in[0,\overline{q})} (1 - r_u(q)) > \frac{\delta_i^{cmkt}}{\lambda_i^{opt}} > \inf_{q\in[0,\overline{q})} (1 - r_u(q)).$$

Since $q_{ij}^{cmkt}(\varphi)$ strictly increases in φ and $1 - r_u(q)$ is monotonic in q, there exists a unique $\tilde{\varphi}_{ij}$ s.t. $1-r_u(q_{ij}^{cmkt}(\tilde{\varphi}_{ij})) = \frac{\delta_i^{cmkt}}{\lambda_i^{opt}}$, indicating that $\frac{u'(q_{ij}^{cmkt}(\tilde{\varphi}_{ij}))}{u'(q_{ij}^{opt}(\tilde{\varphi}_{ij}))} = 1$ and $q_{ij}^{cmkt}(\tilde{\varphi}_{ij}) = q_{ij}^{opt}(\tilde{\varphi}_{ij})$. $\forall \varphi > \widetilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) > q_{ij}^{cmkt}(\tilde{\varphi}_{ij})$, $1 - r_u(q_{ij}^{cmkt}(\varphi)) < 1 - r_u(q_{ij}^{cmkt}(\tilde{\varphi}_{ij})) = \frac{\delta_i^{cmkt}}{\lambda_i^{opt}}$, and, hence, $\frac{u'(q_{ij}^{cmkt}(\varphi))}{u'(q_{ij}^{opt}(\varphi))} > 1$ and $q_{ij}^{cmkt}(\varphi) < q_{ij}^{opt}(\varphi)$. Similarly, $\forall \varphi < \widetilde{\varphi}_{ij}$ we obtain $q_{ij}^{cmkt}(\varphi) > q_{ij}^{opt}(\varphi)$.

Now we can show the location of the intersections of the domestic-sales versus the exporting-sales implicit quantity for the centralized and the social optimum equilibrium in q-φ-space. Let us take country H as an example, and combine its two FOCs in the centralized equilibrium to obtain

$$\left[1 - r_u(q_{HF}^{cmkt}(\varphi))\right] u'(q_{HF}^{cmkt}(\varphi)) = \frac{\tau_{HF}}{\tau_{HH}} \left[1 - r_u(q_{HH}^{cmkt}(\varphi))\right] u'(q_{HH}^{cmkt}(\varphi)).$$

Since $\forall q \in [0, \overline{q})$, $\frac{\partial[1-r_u(q)]u'(q)}{\partial q} < 0$, we obtain $q_{HH}^{cmkt}(\varphi) > q_{HF}^{cmkt}(\varphi)$. Since $(1 - r_u(q))' < 0$, we obtain $1 - r_u(q_{HH}^{cmkt}(\tilde{\varphi}_{HF})) < 1 - r_u(q_{HF}^{cmkt}(\tilde{\varphi}_{HF})) = \frac{\delta_H^{cmkt}}{\lambda_H^{opt}}$ and $\tilde{\varphi}_{HH} < \tilde{\varphi}_{HF}$. The latter means that the intersection of the implicit domestic-sales quantity function is located at a lower productivity level ($\tilde{\varphi}_{HH}$ for country H) than that of the exporting-sales ones ($\tilde{\varphi}_{HF}$ for country H).

When $(1 - r_u(q))' > 0$ and $\varepsilon_u'(q) > 0$, $\lim_{q\to 0} \varepsilon_u(q) = \lim_{q\to 0} 1 - r_u(q) > 0$ holds and, hence, we obtain $\inf_{q\geqslant 0}(1 - r_u(q)) = \inf_{q\geqslant 0} \varepsilon_u(q)$. Because $\varepsilon_u'(q) > 0$, $\lim_{q\to +\infty} \varepsilon_u(q) > \lim_{q\to 0} \varepsilon_u(q) > 0$, and, hence, we obtain $\lim_{q\to +\infty} \varepsilon_u(q) = \lim_{q\to +\infty} 1 - r_u(q)$ and $\sup_{q\geqslant 0}(1 - r_u(q)) = \sup_{q\geqslant 0} \varepsilon_u(q)$ by L'Hôpital's rule. By identical arguments, we obtain that there exists a unique $\tilde{\varphi}_{ij}$ s.t. $q_{ij}^{cmkt}(\widetilde{\varphi}_{ij}) = q_{ij}^{opt}(\widetilde{\varphi}_{ij})$. $\forall \varphi > \widetilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) > q_{ij}^{opt}(\varphi)$. $\forall \varphi < \widetilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) < q_{ij}^{opt}(\varphi)$. Besides, the intersection of the implicit domestic-sales quantity function is located at a lower productivity level than that of the exporting-sales one, so that $\tilde{\varphi}_{HH} < \tilde{\varphi}_{HF}$. □

Proof of Proposition 6. For $\alpha \in [0, 1]$, we define:

$$v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) \equiv \alpha u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi}) + (1 - \alpha)u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))$$

and

$$\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) \equiv u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi}) - u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) = u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) \left[\varepsilon_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) - 1\right], \tag{41}$$

then

$$v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) = u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) + \alpha\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})). \tag{42}$$

Consider the weighted-average Lagrangian

$$\begin{aligned}\mathcal{L} = M_i &\left\{\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)\right\} \\ &+ \beta_i(\alpha)\left\{L_i w_i - M_i\left\{\sum_j\left[\int_{\varphi_{ij}^*}^{+\infty}\frac{\tau_{ij}w_i}{\varphi}q_{ij}(\frac{\tau_{ij}w_i}{\varphi})L_j + f_{ij}w_i dG_i(\varphi)\right] + F_i w_i\right\}\right\},\end{aligned}$$

where $\beta_i(\alpha)$ is the Lagrange multiplier depending on weight α. This Lagrangian is identical to the centralized market problem when $\alpha = 1$ and is identical to the social optimum problem when $\alpha = 0$. $\forall i, j = H, F$, consider the FOCs w.r.t. the masses of entrants M_i and quantity $q_{ij}(\frac{\tau_{ij}w_i}{\varphi})$: $\beta_i(\alpha)w_i = \frac{M_i}{L_i}\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)$ and $\left\{\alpha\left[1 - r_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))\right] + (1-\alpha)\right\}u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) = \beta_i(\alpha)\frac{\tau_{ij}w_i}{\varphi}$.

When fixed costs are zero, differentiating $\beta_i(\alpha)w_i$ w.r.t. α, we obtain: $\frac{d\beta_i(\alpha)w_i}{d\alpha} = \frac{M_i}{L_i}\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty}\left[\varepsilon_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) - 1\right]u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi) < 0$, indicating that $\lambda_i^{opt} > \delta_i^{cmkt}$. Besides, $\forall i, j = H, F$, $f_{ij} = 0$ leads to $(q_{ij}^*)^{cmkt} = (q_{ij}^*)^{opt} = 0$. Evaluating the FOCs of the centralized market and the socially optimal equilibrium at the productivity cutoff, we obtain $1 - r_u(0) = \frac{(\delta_i)^{cmkt}}{(\lambda_i)^{opt}}\frac{(\varphi_{ij}^*)^{opt}}{(\varphi_{ij}^*)^{cmkt}}$. Recall that we have $\sup_{q\in[0,\bar{q})}(1 - r_u(q)) > \frac{\delta_i^{cmkt}}{\lambda_i^{opt}} > \inf_{q\in[0,\bar{q})}(1 - r_u(q))$. Consider the case of aligned preferences so that $\varepsilon_u'(q)(1 - r_u(q))' > 0$. Then, if $\varepsilon_u'(q) > 0$ and $(1-r_u(q))' > 0$, $\frac{\delta_i^{cmkt}}{\lambda_i^{opt}} > \inf_{q\in[0,\bar{q})}(1-r_u(q)) = 1-r_u(0)$, implying that $(\varphi_{ij}^*)^{cmkt} > (\varphi_{ij}^*)^{opt}$. In contrast, if $\varepsilon_u'(q) < 0$ and $(1 - r_u(q))' < 0$, then $(\varphi_{ij}^*)^{cmkt} < (\varphi_{ij}^*)^{opt}$.

If fixed costs are greater than zero, the FOCs w.r.t the cutoff productivity of the weighted-average Lagrangian yields:

$$v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*})) = \beta_i(\alpha)\left[\frac{\tau_{ij}w_i}{\varphi_{ij}^*}q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}) + \frac{f_{ij}w_i}{L_j}\right]. \tag{43}$$

Differentiate the LHS w.r.t. α to obtain:

$$\frac{dv_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))}{d\alpha} = \frac{d\beta_i(\alpha)}{d\alpha}\frac{v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))}{\beta_i(\alpha)} + \beta_i(\alpha)\left[\frac{d(\frac{1}{\varphi_{ij}^*})}{d\alpha}\tau_{ij}w_i q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}) + \frac{\tau_{ij}w_i}{\varphi_{ij}^*}\frac{dq_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*})}{d\alpha}\right].$$

We can express $v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))$ alternatively as $v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*})) = \alpha u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}) + (1-\alpha)u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))$. Differentiation of the latter obtains $\frac{dv_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*}))}{d\alpha} = \omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi_{ij}^*})) +$

$\frac{\beta_i(\alpha)\tau_{ij}w_i}{\varphi^*_{ij}}\frac{dq_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{d\alpha}$. Therefore, by equating the two expressions, we obtain

$$\begin{aligned}&\beta_i(\alpha)\frac{d(\frac{1}{\varphi^*_{ij}})}{d\alpha}\tau_{ij}w_iq_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})\\=&\frac{\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)-v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)}{\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)}.\end{aligned}$$

Note that the sign of $\frac{d(\frac{1}{\varphi^*_{ij}})}{d\alpha}$ depends on the numerator. With (41) and (42), we can simplify the numerator as

$$\begin{aligned}&u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})\left[\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)\right]\\&-u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\left[\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi})dG_i(\varphi)\right]\gtrless 0.\end{aligned}\tag{44}$$

With the definition of $\underline{j}$ and $\overline{j}$, we can rewrite (44) as

$$\varepsilon_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\gtrless\frac{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}\varepsilon_u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi}))u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi}))dG_i(\varphi)+L_{\overline{j}}\int_{\varphi^*_{i\overline{j}}}^{+\infty}\varepsilon_u(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi}))u(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi}))dG_i(\varphi)}{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi}))dG_i(\varphi)+L_{\overline{j}}\int_{\varphi^*_{i\overline{j}}}^{+\infty}u(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi}))dG_i(\varphi)}.\tag{45}$$

Combining (43) with the FOC $\left\{\alpha\left[1-r_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\right]+(1-\alpha)\right\}u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))=\beta_i(\alpha)\frac{\tau_{ij}w_i}{\varphi^*_{ij}}$, we obtain

$$\begin{aligned}&\beta_i(\alpha)f_{ij}w_i\\=&L_j\left\{(1-\alpha)\left[u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))-u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})\right]+\alpha\left[-u''(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))^2\right]\right\}.\end{aligned}\tag{46}$$

Since $\left[u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))-u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})\right]$ and $\left[-u''(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))^2\right]$ both increase in $q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})$, the RHS of (46) increases in $q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})$. Combine the conditions w.r.t $\underline{j}$ and $\overline{j}$:

$$\frac{\left\{(1-\alpha)\left[u(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi^*_{i\overline{j}}}))-u'(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi^*_{i\overline{j}}}))q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi^*_{i\overline{j}}})\right]+\alpha\left[-u''(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi^*_{i\overline{j}}}))(q_{i\overline{j}}(\frac{\tau_{i\overline{j}}w_i}{\varphi^*_{i\overline{j}}}))^2\right]\right\}}{\left\{(1-\alpha)\left[u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi^*_{i\underline{j}}}))-u'(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi^*_{i\underline{j}}}))q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi^*_{i\underline{j}}})\right]+\alpha\left[-u''(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi^*_{i\underline{j}}}))(q_{i\underline{j}}(\frac{\tau_{i\underline{j}}w_i}{\varphi^*_{i\underline{j}}}))^2\right]\right\}}=\frac{f_{i\overline{j}}L_{\underline{j}}}{L_{\overline{j}}f_{i\underline{j}}}.$$

By definition, $\frac{f_{i\underline{j}}}{L_{\underline{j}}} > \frac{f_{i\bar{j}}}{L_{\bar{j}}}$, and we obtain $q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi^*_{i\underline{j}}}) > q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi^*_{i\bar{j}}})$.

If $\varepsilon'_u(q) > 0$, then $\varepsilon_u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi^*_{i\underline{j}}})) > \varepsilon_u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi^*_{i\bar{j}}}))$. We can rewrite (45) as

$$
\begin{aligned}
&\frac{L_{\underline{j}} \int_{\varphi^*_{i\underline{j}}}^{+\infty} \varepsilon_u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi}))u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi}))dG_i(\varphi) + L_{\bar{j}} \int_{\varphi^*_{i\bar{j}}}^{+\infty} \varepsilon_u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi}))u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi}))dG_i(\varphi)}{L_{\underline{j}} \int_{\varphi^*_{i\underline{j}}}^{+\infty} u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi}))dG_i(\varphi) + L_{\bar{j}} \int_{\varphi^*_{i\bar{j}}}^{+\infty} u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi}))dG_i(\varphi)} \\
> &\frac{L_{\underline{j}}\varepsilon_u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi^*_{i\underline{j}}})) \int_{\varphi^*_{i\underline{j}}}^{+\infty} u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi}))dG_i(\varphi) + L_{\bar{j}}\varepsilon_u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi^*_{i\bar{j}}})) \int_{\varphi^*_{i\bar{j}}}^{+\infty} u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi}))dG_i(\varphi)}{L_{\underline{j}} \int_{\varphi^*_{i\underline{j}}}^{+\infty} u(q_{i\underline{j}}(\frac{\tau_{i\underline{j}} w_i}{\varphi}))dG_i(\varphi) + L_{\bar{j}} \int_{\varphi^*_{i\bar{j}}}^{+\infty} u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi}))dG_i(\varphi)} > \varepsilon_u(q_{i\bar{j}}(\frac{\tau_{i\bar{j}} w_i}{\varphi^*_{i\bar{j}}})).
\end{aligned}
$$

Therefore, we prove that $\frac{d(\frac{1}{\varphi^*_{i\bar{j}}})}{d\alpha} < 0$ and $\frac{d(\varphi^*_{i\bar{j}})}{d\alpha} > 0$, implying that $(\varphi^*_{i\bar{j}})^{cmkt} > (\varphi^*_{i\bar{j}})^{opt}$ but $(\varphi^*_{i\underline{j}})^{cmkt} \gtrless (\varphi^*_{i\underline{j}})^{opt}$.

When $\varepsilon'_u(q) < 0$, by identical steps, we can show that $(\varphi^*_{i\bar{j}})^{cmkt} < (\varphi^*_{i\bar{j}})^{opt}$ but $(\varphi^*_{i\underline{j}})^{cmkt} \gtrless (\varphi^*_{i\underline{j}})^{opt}$. □

Proof of Proposition 7 and Lemma 3. The proof of Proposition 7 is an application of Proposition 4 with the equivalence between the centralized market equilibrium and the social optimum under CES demand. Lemma 3 can be proven in a similar way as Lemma 4 with the equilibrium solutions under CES demand. □

Specialized results for CES preferences. For brevity, we use the notation of $q_{ij}(\varphi)$ for $q^v_{ij}(\frac{\tau_{ij} w_i}{\varphi})$ for $v \in \{dmkt, cmkt, opt\}$ in the following proof.

Decentralized market equilibrium. The FOCs and ZCPCs yield $\forall i, j = H, F$,

$$
\begin{cases}
\rho^2(q_{ij}(\varphi))^{\rho-1} = \dfrac{\delta_j^{dmkt}\tau_{ij} w_i}{\varphi} \\
\rho^2(q_{ij}(\varphi^*_{ij}))^{\rho-1} = \dfrac{\delta_j^{dmkt}\tau_{ij} w_i}{\varphi^*_{ij}} \\
(\dfrac{1}{\rho} - 1)\dfrac{\tau_{ij}}{\varphi^*_{ij}} q_{ij}(\varphi^*_{ij}) L_j = f_{ij}.
\end{cases} \tag{47}
$$

We obtain explicit expressions for the quantity functions as follows:

$$
\begin{cases}
q_{ij}(\varphi) = \dfrac{\rho}{1-\rho}\dfrac{f_{ij}}{L_j \tau_{ij}}(\dfrac{1}{\varphi^*_{ij}})^{\frac{\rho}{1-\rho}} \varphi^{\frac{1}{1-\rho}} \\
\dfrac{\varphi^*_{jj}}{\varphi^*_{ij}} = (\dfrac{f_{jj}}{f_{ij}})^{\frac{1-\rho}{\rho}}(\dfrac{\tau_{jj}}{\tau_{ij}})(\dfrac{w_j}{w_i})^{\frac{1}{\rho}}.
\end{cases} \tag{48}
$$

Then the ZEPCs read:

$$F_i = \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left[(\frac{1}{\rho} - 1) \frac{\tau_{ij}}{\varphi} q_{ij}(\varphi) L_j - f_{ij} \right] dG_i(\varphi)$$
$$= \frac{\rho}{(1-\rho)\gamma - \rho} \sum_j f_{ij} (\frac{1}{\varphi_{ij}^*})^\gamma. \tag{49}$$

Before solving the system, we need to check Assumptions 3 and 4 to obtain the explicit constraints on the parameter space to guarantee the existence and uniqueness of the decentralized market equilibrium. Recall that we choose the domestic wage rate w_H as the numeraire, and the foreign wage rate w_F is endogenous. Assumption 3 requires $\overline{w_F} > \underline{w_F} > 0$ such that $\forall w_F \in (\underline{w_F}, \overline{w_F})$, $\delta_{HH} > \delta_{FH}(w_F)$ and $\delta_{FF}(w_F) > \delta_{HF}$. The counterfactual partial equilibrium is the solution to the following conditions:

$$\begin{cases} \rho^2 (q_{ij}(\varphi_{ij}^*))^{\rho-1} = \dfrac{\delta_{ij}\tau_{ij} w_i}{\varphi_{ij}^*} \\ (\dfrac{1}{\rho} - 1) \dfrac{\tau_{ij}}{\varphi_{ij}^*} q_{ij}(\varphi_{ij}^*) L_j = f_{ij} \\ \dfrac{\rho}{(1-\rho)\gamma - \rho} f_{ij} (\dfrac{1}{\varphi_{ij}^*})^\gamma = F_i. \end{cases}$$

We can obtain the explicit expressions for $\delta_{ij}(w_i)$ as follows:

$$(\delta_{ij}(w_i))^\gamma = \frac{\rho^{\frac{(1+\rho)\gamma+\rho}{\rho}} (1-\rho)^{\frac{(1-\rho)\gamma}{\rho}}}{(1-\rho)\gamma - \rho} L_j^{\frac{(1-\rho)\gamma}{\rho}} \left[(\frac{1}{F_i w_i}) (\frac{1}{f_{ij} w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{ij} w_i})^\gamma \right]$$

The requirement $\overline{w_F} > \underline{w_F} > 0$ then becomes:

$$(\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{\tau_{HF}}{\tau_{FF}})^\gamma > (\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{\tau_{HH}}{\tau_{FH}})^\gamma. \tag{50}$$

Therefore, we can assume $\forall i \neq j$, $f_{ij} > f_{jj}$ and $\tau_{ij} > \tau_{jj}$ as the sufficient conditions for Assumption 3.

As for Assumption 4, since $\overline{B} = +\infty$ under CES preferences, $\delta_{FF}(w_F) < \frac{L_F \overline{B}}{f_{HF} w_H}$ holds. We first rewrite the ZEPC as:

$$\sum_\ell \Pi_{i\ell}(\delta_\ell) = L_i^{\frac{(1-\rho)\gamma}{\rho}} (\frac{1}{f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{ii}})^\gamma (\frac{1}{\delta_i})^{\frac{\gamma}{\rho}} + L_j^{\frac{(1-\rho)\gamma}{\rho}} (\frac{1}{f_{ij}})^{(1-\rho)\gamma-1} (\frac{1}{\tau_{ij}})^\gamma (\frac{1}{\delta_j})^{\frac{\gamma}{\rho}} = \frac{F_i (w_i)^{\frac{\gamma}{\rho}} [(1-\rho)\gamma - \rho]}{\rho^{\frac{(1+\rho)\gamma+\rho}{\rho}} (1-\rho)^{\frac{(1-\rho)\gamma}{\rho}}}$$

and then express the corresponding explicit functions as follows: $\forall \delta_F \in (\delta_{FF}(w_F), +\infty)$,

$$\frac{d\delta_H^F(\delta_F \mid w_F)}{d\delta_F} = -\frac{\partial \Pi_{FF}/\partial \delta_F}{\partial \Pi_{FH}/\partial \delta_H} = -(\frac{L_F}{L_H})^{\frac{(1-\rho)\gamma}{\rho}}(\frac{f_{FH}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{FH}}{\tau_{FF}})^{\gamma}(\frac{\delta_F}{\delta_H})^{-(\frac{\gamma}{\rho}+1)}$$

and
$$\frac{d\delta_H^H(\delta_F)}{d\delta_F} = -\frac{\partial \Pi_{HF}/\partial \delta_F}{\partial \Pi_{HH}/\partial \delta_H} = -(\frac{L_F}{L_H})^{\frac{(1-\rho)\gamma}{\rho}}(\frac{f_{HH}}{f_{HF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HH}}{\tau_{HF}})^{\gamma}(\frac{\delta_F}{\delta_H})^{-(\frac{\gamma}{\rho}+1)}.$$

Assumption 4 requires $\frac{d\delta_H^F(\delta_F|w_F)}{d\delta_F} < \frac{d\delta_H^H(\delta_F)}{d\delta_F}$, which can also be simplified to (50). Therefore, assuming (50) can sufficiently restrict the parameter space such that the decentralized market equilibrium is uniquely determined.

Now we can solve for the cutoff productivities with the cutoff relation (48) and the ZEPCs (49). We need to consider the system of ZEPCs for all countries jointly to obtain the solution for all cutoff productivities.

There, $\forall i \neq j$, the explicit solutions read:

$$\varphi_{ii}^* = \left\{ \frac{\rho f_{ii}\left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma} - 1\right]}{[(1-\rho)\gamma-\rho]\, F_i \left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma} - \frac{F_j w_j}{F_i w_i}(\frac{f_{ji}w_j}{f_{ii}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^{\gamma}\right]} \right\}^{\frac{1}{\gamma}}$$

and

$$\varphi_{ij}^* = \left\{ \frac{\rho f_{ij}\left[1 - (\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]}{[(1-\rho)\gamma-\rho]\, F_i \left[\frac{F_j w_j}{F_i w_i}(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma} - (\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]} \right\}^{\frac{1}{\gamma}}.$$

Consider the resource constraint and simplify as well as rearrange it to obtain

$$M_i \left\{ \frac{\gamma-\rho}{(1-\rho)r-\rho} \sum_j f_{ij}(\frac{1}{\varphi_{ij}^*})^{\gamma} + F_i \right\} = L_i. \tag{51}$$

With the expressions for the cutoffs, we obtain the solution for the masses of entrants:

$$M_i = \frac{L_i \rho}{F_i \gamma}.$$

We further rewrite the TBC as

$$\frac{L_i f_{ij}}{F_i} w_i (\frac{1}{\varphi_{ij}^*})^{\gamma} = \frac{L_j f_{ji}}{F_j} w_j (\frac{1}{\varphi_{ji}^*})^{\gamma},$$

and the implicit solution of the relative wage ratio is:

$$\frac{w_j}{w_i} = \frac{L_i}{L_j} \frac{\left[(\frac{F_j w_j}{F_i w_i})(\frac{f_{jj} w_j}{f_{ij} w_i})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^\gamma - (\frac{f_{jj} f_{ii}}{f_{ji} f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^\gamma\right]}{\left[(\frac{F_i w_i}{F_j w_j})(\frac{f_{ii} w_i}{f_{ji} w_j})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{ii} w_i}{\tau_{ji} w_j})^\gamma - (\frac{f_{jj} f_{ii}}{f_{ji} f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^\gamma\right]}.$$

Centralized market equilibrium and social optimum. By construction, the centralized market equilibrium is the same as the social optimum under CES. The FOCs and ZCPCs for centralized market yield $\forall i, j = H, F$,

$$\begin{cases} \rho^2 (q_{ij}(\varphi))^{\rho-1} = \dfrac{\delta_i^{cmkt}\tau_{ij} w_i}{\varphi} \\ \rho^2 (q_{ij}(\varphi_{ij}^*))^{\rho-1} = \dfrac{\delta_i^{cmkt}\tau_{ij} w_i}{\varphi_{ij}^*} \\ (\dfrac{1}{\rho} - 1)\dfrac{\tau_{ij}}{\varphi_{ij}^*} q_{ij}(\varphi_{ij}^*) L_j = f_{ij}. \end{cases} \tag{52}$$

We obtain explicit expressions for the quantity functions as follows:

$$\begin{cases} q_{ij}(\varphi) = \dfrac{\rho}{1-\rho}\dfrac{f_{ij}}{L_j \tau_{ij}}(\dfrac{1}{\varphi_{ij}^*})^{\frac{\rho}{1-\rho}} \varphi^{\frac{1}{1-\rho}} \\ \quad \dfrac{\varphi_{ii}^*}{\varphi_{ij}^*} = \dfrac{\tau_{ii}}{\tau_{ij}}(\dfrac{f_{ii} L_j}{f_{ij} L_i})^{\frac{1-\rho}{\rho}}. \end{cases} \tag{53}$$

The ZEPCs for the centralized market can also be simplified to (49). Note that, in contrast to the decentralized market equilibrium, the cutoff productivity φ_{ij}^* is proportional to φ_{ii}^* rather than to φ_{jj}^* in the centralized market equilibrium. Therefore, the ZEPC of country i alone pins down the cutoffs φ_{ii}^* and φ_{ij}^*. We can then obtain the explicit cutoffs as follows:

$$\begin{cases} \varphi_{ii}^* = \left\{ \dfrac{\rho f_{ii}}{[(1-\rho)\gamma - \rho] F_i} \left[1 + (\dfrac{\tau_{ii}}{\tau_{ij}})^\gamma (\dfrac{f_{ii}}{f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\dfrac{L_j}{L_i})^{\frac{(1-\rho)\gamma}{\rho}} \right] \right\}^{\frac{1}{\gamma}} \\ \varphi_{ij}^* = \left\{ \dfrac{\rho f_{ij}}{[(1-\rho)\gamma - \rho] F_i} \left[1 + (\dfrac{\tau_{ij}}{\tau_{ii}})^\gamma (\dfrac{f_{ij}}{f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\dfrac{L_i}{L_j})^{\frac{(1-\rho)\gamma}{\rho}} \right] \right\}^{\frac{1}{\gamma}}. \end{cases}$$

With the solutions of cutoffs and (51), we obtain the solution of masses of entrants as $M_i = \frac{L_i \rho}{F_i \gamma}$. □

Specialized results for CARA preferences. With CARA we throughout consider the case of zero fixed costs, $f_{ij} = 0 \ \forall i, j = H, F$.

Decentralized market equilibrium. The FOCs yield $\forall i, j = H, F$,

$$\begin{cases} ae^{-aq_{ij}(\varphi)}(1 - aq_{ij}(\varphi)) = \dfrac{\delta_j^{dmkt}\tau_{ij}w_i}{\varphi} \\ ae^{-aq_{ij}(\varphi_{ij}^*)}(1 - aq_{ij}(\varphi_{ij}^*)) = \dfrac{\delta_j^{dmkt}\tau_{ij}w_i}{\varphi_{ij}^*} \\ q_{ij}(\varphi_{ij}^*) = 0. \end{cases}$$

We can employ the Lambert function $\mathbf{W}$, which satisfies $z = \mathbf{W}(z)e^{\mathbf{W}(z)}$, and obtain explicit expressions for the quantity functions as follows:

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\left[1 - \mathbf{W}(e\dfrac{\varphi_{ij}^*}{\varphi})\right] \\ \quad \varphi_{ij}^* = \dfrac{\tau_{ij}w_i}{\tau_{jj}w_j}\varphi_{jj}^*. \end{cases} \tag{54}$$

Define $z_{ij} \equiv \mathbf{W}(e\frac{\varphi_{ij}^*}{\varphi})$, so as to obtain $\varphi = \frac{\varphi_{ij}^*}{z_{ij}e^{z_{ij}-1}}$. Then, $z_{ij} = 1$ when $\varphi = \varphi_{ij}^*$, and $z_{ij} = 0$ when $\varphi = +\infty$. With the Pareto distribution $G(\varphi) = 1 - (\frac{1}{\varphi})^\gamma$, we obtain

$$d\varphi = \frac{-\varphi_{ij}^*(z_{ij}+1)}{z_{ij}^2 e^{z_{ij}-1}}dz_{ij}, \;\; dG(\varphi) = -\gamma\frac{1}{(\varphi_{ij}^*)^\gamma}\frac{z_{ij}+1}{z_{ij}}(z_{ij}e^{z_{ij}-1})^\gamma dz_{ij}. \tag{55}$$

Then, the ZEPCs can be rewritten as:

$$F_i w_i = \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left\{\left[\frac{1}{1 - r_u(q_{ij}(\varphi))} - 1\right]\frac{\tau_{ij}w_i}{\varphi}q_{ij}(\varphi)L_j - f_{ij}w_i\right\} dG(\varphi) = \frac{\gamma\kappa_1 w_i}{a}\sum_j \frac{L_j\tau_{ij}}{(\varphi_{ij}^*)^{\gamma+1}}, \tag{56}$$

where $\kappa_1 = \int_0^1(\frac{1}{z} + z - 2)\frac{z+1}{z}(ze^{z-1})^{\gamma+1}dz$.[16]

Before solving the system, we need to check Assumptions 3 and 4 to obey the constraints on the parameter space to guarantee the existence and uniqueness of the decentralized market equilibrium. Recall that we choose the domestic wage rate w_H as the numeraire, while the foreign wage rate w_F is endogenous. Assumption 3 requires that $\delta_{FF}(w_F) > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}(w_F)$. According to the definition of a counterfactual partial equilibrium, combining $\delta_{ij} = \frac{a\varphi_{ij}^*}{\tau_{ij}w_i}$ and $\frac{\gamma\kappa_1}{a}\frac{L_j\tau_{ij}}{(\varphi_{ij}^*)^{\gamma+1}} = F_i$, we can obtain the explicit expressions for δ_{ij}.

[16]We can drop the index ij since the limit values of z_{ij} are independent of the index.

Taking $\delta_{HH} > \delta_{FH}(w_F)$ as an example, δ_{HH} and $\delta_{FH}(w_F)$ read:

$$\begin{cases} \delta_{HH} = \dfrac{a}{\tau_{HH} w_H} (\dfrac{L_H \tau_{HH} \gamma \kappa_1}{a F_H})^{\frac{1}{\gamma+1}} \\ \delta_{FH}(w_F) = \dfrac{a}{\tau_{FH} w_F} (\dfrac{L_H \tau_{FH} \gamma \kappa_1}{a F_F})^{\frac{1}{\gamma+1}}. \end{cases}$$

The difference between δ_{HH} and $\delta_{FH}(w_F)$ reads:

$$\delta_{HH} - \delta_{FH}(w_F) = a^{\frac{\gamma}{\gamma+1}} (L_H \gamma \kappa_1)^{\frac{1}{\gamma+1}} \left[(\frac{1}{F_H w_H})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{HH} w_H})^{\frac{\gamma}{\gamma+1}} - (\frac{1}{F_F w_F})^{\frac{1}{\gamma+1}} (\frac{1}{\tau_{FH} w_F})^{\frac{\gamma}{\gamma+1}} \right].$$

In order to ensure $\delta_{HH} > \delta_{FH}(w_F)$, the endogenous wage must satisfy $\frac{w_F}{w_H} > (\frac{F_H}{F_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HH}}{\tau_{FH}})^{\frac{\gamma}{\gamma+1}}$. Similarly, $\delta_{FF}(w_F) > \delta_{HF}$ requires $(\frac{F_H}{F_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HF}}{\tau_{FF}})^{\frac{\gamma}{\gamma+1}} > \frac{w_F}{w_H}$. We combine the two inequalities to ensure the existence of an interval for the relative wage ratio $\frac{\tau_{HF}}{\tau_{FF}} > \frac{\tau_{HH}}{\tau_{FH}}$. Therefore, we can assume $\forall i \neq j$, $\tau_{ij} > \tau_{jj}$ as the sufficient condition for Assumption 3.

For Assumption 4, since $\forall i, j = H, F$, $f_{ij} = 0$ and $\bar{B}$ is finite, $\delta_{FF}(w_F) < \frac{L_F \bar{B}}{f_{HF} w_H}$ always holds. Combing with the FOCs, we can rewrite the ZEPCs in terms of δ_H and δ_F as:

$$\begin{cases} \dfrac{L_H}{\tau_{HH}^{\gamma}} \dfrac{1}{\delta_H^{\gamma+1}} + \dfrac{L_F}{\tau_{HF}^{\gamma}} \dfrac{1}{\delta_F^{\gamma+1}} = \dfrac{F_H(w_H)^{\gamma+1}}{\gamma \kappa_1 a^{\gamma}} \\ \dfrac{L_H}{\tau_{FH}^{\gamma}} \dfrac{1}{\delta_H^{\gamma+1}} + \dfrac{L_F}{\tau_{FF}^{\gamma}} \dfrac{1}{\delta_F^{\gamma+1}} = \dfrac{F_F(w_F)^{\gamma+1}}{\gamma \kappa_1 a^{\gamma}}. \end{cases} \tag{57}$$

According to the implicit function theorem, $\forall \delta_F \in (\delta_{FF}(w_F), +\infty)$, we have $\frac{d\delta_H^F(\delta_F|w_F)}{d\delta_F} = -\frac{\partial \Pi_{FF}/\partial \delta_F}{\partial \Pi_{FH}/\partial \delta_H} = -\frac{L_F}{L_H} \frac{\tau_{FH}^{\gamma}}{\tau_{FF}^{\gamma}} (\frac{\delta_F}{\delta_H})^{-(\gamma+2)}$ and $\frac{d\delta_H^H(\delta_F)}{d\delta_F} = -\frac{\partial \Pi_{HF}/\partial \delta_F}{\partial \Pi_{HH}/\partial \delta_H} = -\frac{L_F}{L_H} \frac{\tau_{HH}^{\gamma}}{\tau_{HF}^{\gamma}} (\frac{\delta_F}{\delta_H})^{-(\gamma+2)}$. Assumption 4 requires $\frac{d\delta_H^F(\delta_F|w_F)}{d\delta_F} < \frac{d\delta_H^H(\delta_F)}{d\delta_F}$, which can be simplified as $\frac{\tau_{HF}}{\tau_{FF}} > \frac{\tau_{HH}}{\tau_{FH}}$. Therefore, assuming $\forall i \neq j$, $\tau_{ij} > \tau_{jj}$ can sufficiently restrict the parameter space such that Assumption 3 and 4 hold, and, thus, the decentralized market equilibrium is uniquely determined.

Now we can solve the cutoff productivities based on (54) and the ZEPCs (56). Note that (56) permits reducing the set of cutoff productivities to those for the domestic market in each country. However, (54) indicates that, when considering the latter, the ZEPC in each country depends on the domestic cutoff productivities in all countries. Hence, the system of ZEPCs for all countries has to be used to determine the country-specific domestic cutoff productivities in an interdependent way.

There, $\forall i \neq j$, we obtain explicit solutions for the cutoffs relevant for domestic sales of

$$(\varphi_{ii}^*)^{\gamma+1} = \frac{\gamma \kappa_1 L_i \tau_{ii} \left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1 \right]}{aF_i \left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{F_j w_j}{F_i w_i})(\frac{\tau_{ji} w_j}{\tau_{ii} w_i})^\gamma \right]}, \tag{58}$$

and for exporting sales of

$$(\varphi_{ij}^*)^{\gamma+1} = (\frac{\tau_{ij} w_i}{\tau_{jj} w_j} \varphi_{jj}^*)^{\gamma+1} = \frac{\gamma \kappa_1 L_j \tau_{ij} \left[1 - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma \right]}{aF_i \left[(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma \right]} \tag{59}$$

One can see that $\tau_{HF} > \tau_{FF}$ and $\tau_{FH} > \tau_{HH}$ jointly guarantee that the numerators of φ_{ii}^* and φ_{ij}^* are positive, and $\delta_{FF}(w_F) > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}(w_F)$ guarantee that the denominators are positive.

The resource constraint can be simplified as:

$$L_i = M_i \left\{ \sum_j \left[L_j \tau_{ij} \int_{\varphi_{ij}^*}^{+\infty} \frac{q_{ij}(\varphi)}{\varphi} dG(\varphi) \right] + F_i \right\} = M_i \left\{ \frac{\gamma \kappa_3}{a} \sum_j \left[\frac{L_j \tau_{ij}}{(\varphi_{ij}^*)^{\gamma+1}} \right] + F_i \right\},$$

where $\kappa_3 = \int_0^1 (ze^{z-1})^{\gamma+1} \frac{1-z^2}{z} dz$ and is independent of the index ij. One can verify that $\frac{\kappa_3}{\kappa_1} = \gamma$ holds. Then we can obtain the solution for the masses of entrants $M_i = \frac{L_i}{(\gamma+1)F_i}$.

Now consider the trade balanced condition (TBC) $\forall i \neq j$:

$$M_i \int_{\varphi_{ij}^*}^{+\infty} \frac{1}{1 - r_u(q_{ij}(\varphi))} \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\varphi) L_j dG(\varphi) = M_j \int_{\varphi_{ji}^*}^{+\infty} \frac{1}{1 - r_u(q_{ji}(\varphi))} \frac{\tau_{ji} w_j}{\varphi} q_{ji}(\varphi) L_i dG(\varphi). \tag{60}$$

With (55) and solutions of $q_{ij}(\varphi)$ and M_i, we can rewrite (60) as

$$\frac{L_i w_i}{F_i} \frac{\tau_{ij} L_j}{(\varphi_{ij}^*)^{\gamma+1}} = \frac{L_j w_j}{F_j} \frac{\tau_{ji} L_i}{(\varphi_{ji}^*)^{\gamma+1}}.$$

With (59), we obtain the implicit solution for the relative wage ratio as

$$\frac{w_j}{w_i} = \frac{L_i}{L_j} \frac{(\frac{F_j w_j}{F_i w_i})(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}{(\frac{F_i w_i}{F_j w_j})(\frac{\tau_{ii} w_i}{\tau_{ji} w_j})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}.$$

Centralized market equilibrium. The FOCs for the centralized market yield $\forall i, j =$

H, F,

$$\begin{cases} ae^{-aq_{ij}(\varphi)}(1 - aq_{ij}(\varphi)) = \dfrac{\delta_i^{cmkt}\tau_{ij}w_i}{\varphi} \\ ae^{-aq_{ij}(\varphi_{ij}^*)}(1 - aq_{ij}(\varphi_{ij}^*)) = \dfrac{\delta_i^{cmkt}\tau_{ij}w_i}{\varphi_{ij}^*} \\ q_{ij}(\varphi_{ij}^*) = 0. \end{cases}$$

As in the decentralized equilibrium, we can apply the Lambert function $\mathbf{W}$ and obtain

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\left[1 - \mathbf{W}(e\dfrac{\varphi_{ij}^*}{\varphi})\right] \\ \varphi_{ij}^* = \dfrac{\tau_{ij}}{\tau_{ii}}\varphi_{ii}^*. \end{cases} \tag{61}$$

Using identical definitions of z_{ij} and κ_1 as in the decentralized equilibrium, we can simplify the ZEPC as

$$F_i w_i = \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left\{ \left[\frac{1}{1 - r_u(q_{ij}(\varphi))} - 1\right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\varphi)L_j - f_{ij}w_i \right\} dG(\varphi) = \sum_j \frac{L_j \gamma \tau_{ij} w_i \kappa_1}{a(\varphi_{ij}^*)^{\gamma+1}}. \tag{62}$$

In this case, since the demand shifter δ_i^{cmkt} is indexed by the origin, we can solve the cutoff productivity of origin i based on the ZEPC (62) and the cutoff relation (61) in country i. Note that, in contrast to the decentralized market equilibrium, the cutoff productivity φ_{ij}^* is proportional to φ_{ii}^* rather than to φ_{jj}^* in the centralized market equilibrium. Therefore, a country's own resource constraint alone pins down the cutoff productivity φ_{ii}^*. Then, $\forall i \neq j$, we can obtain explicit solutions for the productivity cutoffs in country i:

$$\begin{cases} (\varphi_{ii}^*)^{\gamma+1} = \dfrac{\tau_{ii}\gamma\kappa_1\left[L_i + L_j(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{aF_i} \\ (\varphi_{ij}^*)^{\gamma+1} = \dfrac{\tau_{ij}\gamma\kappa_1\left[L_i(\frac{\tau_{ij}}{\tau_{ii}})^\gamma + L_j\right]}{aF_i}. \end{cases} \tag{63}$$

With the resource constraint, we can similarly define κ_3 and obtain the solution for the masses of entrants as $M_i = \frac{L_i}{F_i(\gamma+1)}$.

Social optimum equilibrium. The FOCs $\forall i, j = H, F$, yield

$$\begin{cases} ae^{-aq_{ij}(\varphi)} = \dfrac{\lambda_i^{opt}\tau_{ij}w_i}{\varphi} \\ ae^{-aq_{ij}(\varphi_{ij}^*)} = \dfrac{\lambda_i^{opt}\tau_{ij}w_i}{\varphi_{ij}^*}. \end{cases}$$

We obtain the solutions of the quantity functions and the productivity cutoffs as

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\ln(\dfrac{\varphi}{\varphi_{ij}^*}) & (64) \\ \varphi_{ij}^* = \dfrac{\tau_{ij}}{\tau_{ii}}\varphi_{ii}^*. & (65) \end{cases}$$

Consider the FOC w.r.t. the masses of entrants,

$$\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} u(q_{ij}(\varphi))dG(\varphi) = \lambda_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\varphi) L_j dG(\varphi) \right] + F_i w_i \right\}. \quad (66)$$

With (64), we can rewrite the LHS of (66) as

$$\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} (1 - e^{-aq_{ij}(\varphi)})dG(\varphi) = \sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} (1 - \frac{\varphi_{ij}^*}{\varphi})dG(\varphi) = \sum_j \left[L_j \frac{1}{\gamma+1} (\frac{1}{\varphi_{ij}^*})^\gamma \right].$$

Given that $\lambda_i = \frac{a\varphi_{ii}^*}{\tau_{ii}w_i}$ and (64), the RHS of (66) becomes:

$$\lambda_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\varphi) L_j dG(\varphi) \right] + F_i w_i \right\} = \frac{\varphi_{ii}^*}{\tau_{ii}} \sum_j \left[\tau_{ij} L_j \frac{\gamma}{(\gamma+1)^2 (\varphi_{ij}^*)^{\gamma+1}} \right] + \frac{a\varphi_{ii}^*}{\tau_{ii}} F_i.$$

With (65), equating the LHS to the RHS obtains the solutions for the productivity cutoffs:

$$\begin{cases} (\varphi_{ii}^*)^{\gamma+1} = \dfrac{\tau_{ii}\left[L_i + L_j(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{a(\gamma+1)^2 F_i} \\ (\varphi_{ij}^*)^{\gamma+1} = \dfrac{\tau_{ij}\left[L_i(\frac{\tau_{ij}}{\tau_{ii}})^\gamma + L_j\right]}{a(\gamma+1)^2 F_i}. \end{cases}$$

The resource constraints yield:

$$L_i = M_i \left\{ \sum_j \left[\int_{\varphi_{ij}^*}^{+\infty} \frac{q_{ij}(\varphi)\tau_{ij}L_j}{\varphi} dG(\varphi) \right] + F_i \right\} = M_i \left\{ \frac{\gamma}{a(\gamma+1)^2} \sum_j \left[\frac{\tau_{ij}L_j}{(\varphi_{ij}^*)^{\gamma+1}} \right] + F_i \right\},$$

which yields the explicit solution for the masses of entrants as $M_i = \frac{L_i}{F_i(\gamma+1)}$. □

Proof of Lemma 4. With explicit solutions for cutoffs under the decentralized and centralized market equilibria, we can make the following comparison:

$$\left[\frac{(\varphi_{ii}^*)^{dmkt}}{(\varphi_{ii}^*)^{cmkt}}\right]^{\gamma+1} = \frac{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1\right]}{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{F_j w_j}{F_i w_i})(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^\gamma\right]\left[1 + \frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]} \gtreqless 1,$$

which can be rewritten as:

$$\frac{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1\right]}{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{F_j w_j}{F_i w_i})(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^\gamma\right]} - \left[1 + \frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right] \gtreqless 0.$$

The comparison can be further simplified to

$$(\frac{L_j}{L_i})(\frac{\tau_{ii}}{\tau_{ij}})^\gamma \left\{\frac{L_i}{L_j}\frac{\left[(\frac{1}{F_i w_i})(\frac{1}{\tau_{ii}w_i})^\gamma - (\frac{1}{F_j w_j})(\frac{1}{\tau_{ji}w_j})^\gamma\right]}{\left[(\frac{1}{F_j w_j})(\frac{1}{\tau_{jj}w_j})^\gamma - (\frac{1}{F_i w_i})(\frac{1}{\tau_{ij}w_i})^\gamma\right]} - 1\right\} \gtreqless 0$$
$$\Leftrightarrow L_i\left[(\frac{1}{F_i w_i})(\frac{1}{\tau_{ii}w_i})^\gamma - (\frac{1}{F_j w_j})(\frac{1}{\tau_{ji}w_j})^\gamma\right] \gtreqless L_j\left[(\frac{1}{F_j w_j})(\frac{1}{\tau_{jj}w_j})^\gamma - (\frac{1}{F_i w_i})(\frac{1}{\tau_{ij}w_i})^\gamma\right]. \quad (67)$$

Therefore, we show that, if the LHS of (67) is greater than the RHS, $(\varphi_{ii}^*)^{dmkt} > (\varphi_{ii}^*)^{cmkt}$ and $(\varphi_{jj}^*)^{dmkt} < (\varphi_{jj}^*)^{cmkt}$. One can further see that $(\varphi_{ij}^*)^{dmkt} < (\varphi_{ij}^*)^{cmkt}$ and $(\varphi_{ji}^*)^{dmkt} > (\varphi_{ji}^*)^{cmkt}$ in this case. □

Proof of Lemma 5. With the cutoff solutions of the centralized market equilibrium and the social optimum in Table 3, we obtain the cutoff ratios

$$\left[\frac{(\varphi_{ii}^*)^{cmkt}}{(\varphi_{ii}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{ij}^*)^{cmkt}}{(\varphi_{ij}^*)^{opt}}\right]^{\gamma+1} = (\gamma+1)A,$$

where $A = \int_0^1 z^{\gamma+1}(e^{z-1})^{\gamma+1}dz$. As in Behrens et al. (2020), consider the utility for a representative consumer under market equilibrium and simplify it as:

$$\int_{\varphi_{ij}^*}^{+\infty}(1 - e^{-aq_{ij}(\varphi)})dG(\varphi) = (\frac{1}{\varphi_{ij}^*})^\gamma\left[\frac{1-(\gamma+1)A}{\gamma+1}\right] > 0,$$

indicating that $(\gamma+1)A < 1$ and $\left[\frac{(\varphi_{ii}^*)^{cmkt}}{(\varphi_{ii}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{ij}^*)^{cmkt}}{(\varphi_{ij}^*)^{opt}}\right]^{\gamma+1} < 1$. □

Proof of Lemma 6. With the quantity functions of the centralized market and the social optimum equilibrium, we can define the difference $\Delta q_{ij}(\varphi) = q_{ij}^{opt}(\varphi) - q_{ij}^{cmkt}(\varphi)$. With the properties of the Lambert function $\mathbf{W}$, we can rewrite $q_{ij}^{cmkt}(\varphi) = \frac{1}{a}\ln\left[\frac{\varphi}{(\varphi_{ij}^{*})^{cmkt}}\mathbf{W}(e\frac{(\varphi_{ij}^{*})^{cmkt}}{\varphi})\right]$.

Given that $0 < (\varphi_{ij}^{*})^{cmkt} < (\varphi_{ij}^{*})^{opt} < +\infty$, we obtain the properties:

- $\forall\varphi \in \left[1, (\varphi_{ij}^{*})^{cmkt}\right]$, $\Delta q_{ij}(\varphi) = 0$.
- $\forall\varphi \in \left[(\varphi_{ij}^{*})^{cmkt}, (\varphi_{ij}^{*})^{opt}\right]$, $\Delta q_{ij}(\varphi) = 0 - \frac{1}{a}\ln\left[\frac{\varphi}{(\varphi_{ij}^{*})^{cmkt}}\mathbf{W}(e\frac{(\varphi_{ij}^{*})^{cmkt}}{\varphi})\right] < 0$.
- $\forall\varphi \in ((\varphi_{ij}^{*})^{cmkt}, +\infty)$, $\Delta q_{ij}(\varphi) = \frac{1}{a}\left\{\ln(\frac{(\varphi_{ij}^{*})^{cmkt}}{(\varphi_{ij}^{*})^{opt}}) - \ln\left[\mathbf{W}(e\frac{(\varphi_{ij}^{*})^{cmkt}}{\varphi})\right]\right\}$, which is positive at the limit since $\lim_{\varphi\to+\infty}\ln\left[\mathbf{W}(e\frac{(\varphi_{ij}^{*})^{\text{cmkt}}}{\varphi})\right] = -\infty$ and $\ln(\frac{(\varphi_{ij}^{*})^{cmkt}}{(\varphi_{ij}^{*})^{opt}}) < 0$.

With continuity and monotonicity, we obtain a unique $\tilde{\varphi}_{ij} \in ((\varphi_{ij}^{*})^{cmkt}, +\infty)$ s.t. $q_{ij}^{opt}(\tilde{\varphi}_{ij}) = q_{ij}^{cmkt}(\tilde{\varphi}_{ij})$. The result about $q_{ii}^{cmkt}(\varphi)$ and $q_{ii}^{cmkt}(\varphi)$ can be obtained in the same way.

We further show the relation between $\tilde{\varphi}_{ij}$ and $\tilde{\varphi}_{ii}$. Given that $(\varphi_{ij}^{*})^{cmkt} = \frac{\tau_{ij}}{\tau_{ii}}(\varphi_{ii}^{*})^{cmkt}$ and $(\varphi_{ij}^{*})^{opt} = \frac{\tau_{ij}}{\tau_{ii}}(\varphi_{ii}^{*})^{opt}$, we have

$$\begin{aligned}\Delta(q_{ij}(\tilde{\varphi}_{ij})) =& \frac{1}{a}\left\{\ln(\frac{(\varphi_{ij}^{*})^{cmkt}}{(\varphi_{ij}^{*})^{opt}}) - \ln\left[\mathbf{W}(e\frac{(\varphi_{ij}^{*})^{cmkt}}{\widetilde{\varphi}_{ij}})\right]\right\} \\ =& \frac{1}{a}\left\{\ln(\frac{(\varphi_{ii}^{*})^{cmkt}}{(\varphi_{ii}^{*})^{opt}}) - \ln\left[\mathbf{W}(e\frac{(\varphi_{ii}^{*})^{cmkt}}{\frac{\tau_{ii}}{\tau_{ij}}\widetilde{\varphi}_{ij}})\right]\right\} = \Delta(q_{ii}(\tilde{\varphi}_{ii})) = 0.\end{aligned}$$

Hence, we obtain $\tilde{\varphi}_{ij} = \frac{\tau_{ij}}{\tau_{ii}}\tilde{\varphi}_{ii}$.

Finally, because the quantity functions of the decentralized and centralized market equilibria have the same form, Proposition 9 can be proven in the same way. □