# Market versus Optimum Allocation in Open Economies*

Peter H. Egger
ETH Zürich, CEPR, CESifo

Ruobing Huang†
ETH Zürich

February 23, 2024

**Abstract**

A large body of theoretical and quantitative work concerns models of heterogeneous firms and monopolistic competition. But most of it relies on strong assumptions regarding demand structure, firm-productivity distribution, and country heterogeneity. This paper studies a general equilibrium model with general additive preferences, productivity distributions, and asymmetric countries, for which much less is known. We derive a measure for origin-destination-specific aggregate costs to provide sufficient conditions for the existence and uniqueness of an equilibrium. We then explore the market allocation mechanism and provide a baseline comparison between the market and a utilitarian optimum from a global planner's perspective. We show that misallocation in open economies can be decomposed into two effects, driven by country asymmetry and the variable elasticity of substitution. We present two examples exhibiting constant and variable markups, respectively, to illustrate how to apply our general theorem.

*Keywords:* Monopolistic competition, Heterogeneous firms, Trade, Misallocation
*JEL codes:* F12, L11, D61

*The authors gratefully acknowledge Arnaud Costinot, Carsten Eckel, Katharina Erhardt, Isabelle Méjean, Giordano Mion, John Morrow, Gianmarco Ottaviano, Matthieu Parenti, Jens Südekum, Davide Suverato, 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, the 24th Göttinger Workshop, CITP Inaugural Academic Conference, and the 1st Summer School by Journal of International Economics for helpful comments. We also thank Yuchen Zhong for research assistance. 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

Models of monopolistic competition with heterogeneous firms (Melitz, 2003) are widely used in international economics and beyond. Specifically, they are used to study effects of productivity, trade policy, or globalisation in general on the average effects as well as their distribution (inequality). For reasons of tractability, relatively restrictive assumptions regarding the demand structure are typically imposed. Of those, a constant elasticity of substitution (CES) leading to constant markups, Pareto-distributed firms, country symmetry, and an absence of general-equilibrium effects through, e.g., the presence of an outside sector, are prominent. However, these assumptions are not innocuous regarding the effects in focus. In particular, without these assumptions, much less is known about the existence and uniqueness of equilibria in models with heterogeneous firms and their productivity-related selective market entry.

A recent body of work in this domain studies questions of allocative efficiency. One key result there is that an optimal allocation of output across firms is not being guaranteed: in a single-sector closed economy, it will only materialize under very restrictive model assumptions such as constant markups resulting from CES preferences (Dhingra and Morrow, 2019). That is, firms generally overproduce or underproduce compared to the optimal quantity desirable from a utilitarian social planner's perspective. While the mechanism behind market efficiency is now well understood in a closed economy, this is not the case with international trade, especially, among asymmetric countries. However, a setting with asymmetric open economies is particularly important in the context of studying the effects of trade liberalization or optimal policy. In the absence of allocative efficiency in the decentralized market equilibrium, counterfactual changes may induce effects that perturb the allocation of output. This calls for the actions of a planner to reduce or avoid altogether a misallocation of output through the existence of firms that are too small or too large.

We propose an analysis of a general equilibrium model of monopolistic competition and heterogeneous firms with two asymmetric countries and general additive preferences. We (i) conduct a systematic analysis of the properties of the model, including the existence and uniqueness of an equilibrium, and (ii) provide a comparison between market equilibrium and a social optimum from the perspective of a utilitarian global social planner.

As to the first point, analysing asymmetric-countries models with general demand structures is challenging. Departing from CES preferences, parametric firm-productivity distributions, and country symmetry substantially reduce the analytical tractability of models. As a result, we know little about the existence and uniqueness of equilibria and the pattern of allocation in general equilibrium under such general conditions. In this paper, we establish sufficient conditions for the existence and uniqueness with a simple measure of aggregate cost, summarising all exogenous parameters in the model. Our sufficient conditions, which require that the product of importing (exporting) aggregate costs exceeds that of the aggregate domestic costs, are consistent with customary assumptions in models of heterogeneous firms with selective market entry. Our strategy permits considering a broad range of demand structures in the general-additive-preference domain, nonparametric productivity distributions as well as general country asymmetry. Specifically, we establish and illustrate how destination-specific competition levels and wages are determined, how they interact with each other, and how they affect market outcomes such as production schedules, cutoff productivities, and masses of entrants.

As to the second point, we provide a systematic analysis of misallocation from a utilitarian global social planner's view. The latter may serve as a reasonable baseline for further research on an inefficient allocation under monopolistic competition in open economies. We establish that the gap in the allocation (including production at the intensive margin as well as firm selection at the extensive margin) between the decentralized market and the utilitarian global optimum can be decomposed into two effects. The first effect is driven by the heterogeneous degree of competition in the consumer markets, which originates from the fundamental asymmetry of countries. Specifically, this effect is fundamentally opposite for the domestic versus the foreign market. The second effect originates from the variable elasticity of substitution (VES) demand structure and, hence, variable markups. Although it is qualitatively similar to what we know from closed-economy models, the impact at the extensive margin remains ambiguous due to the option of exporting. The second effect systematically affects the sales of firms from a given country of origin in their consumer markets at home and abroad. We establish that the overall pattern of misallocation in asymmetric open economies is elusive and depends on which one of the mentioned two effects dominates. Furthermore, we establish that a planner who weighs consumers across countries differently can achieve an outcome on the Pareto frontier only with CES but not generally with VES preferences.

We provide two specific examples for the general framework, a CES demand and a VES demand example with Constant Absolute Risk Aversion (CARA), both in the case of Pareto-distributed productivity. We derive an explicit expression for the proposed measure of aggregate cost and illustrate how sufficient conditions guarantee the existence and uniqueness of the market equilibrium in those examples. Besides, we demonstrate that misallocation occurs even with CES demand from the viewpoint of a utilitarian global planner.

With this research agenda, the present paper contributes to the growing literature on the implication of VES demand in the closed economy.[1] Particularly, the present paper is closely related to the earlier work analyzing allocation efficiency under general additive preferences. As a generalization of Dixit and Stiglitz (1977), Dhingra and Morrow (2019) show how demand-side elasticities 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.[2] They highlight two types of inefficiency: one pertaining to an extensive margin associated with the cutoff productivity, and one pertaining to the allocation of quantities across producing firms in terms of over- or under-production, an intensive margin. They show that a CES demand is necessary for the equivalence between the market equilibrium and the social optimum. Behrens et al. (2020) consider a multi-sector context with inter-sectoral labor mobility. In that case, the market generates inefficient selection and firm-level output, as in a single-sector setup. On top of it, it generates inefficient masses of firm entrants.[3] They quantify the sector-level misallocation assuming Constant Absolute Risk Aversion (CARA) preferences,[4] and shows an aggregate welfare loss of about 6-10% of GDP for France and the United Kingdom.[5]

While the inefficiency discussion is well-understood in a closed economy, it is still unclear with international trade among open economies. Related work relies on particular demand structures and/or relatively strong assumptions regarding the parameterization of economies. With the non-additive structure preference and an outside sector proposed by Melitz and Ottaviano (2008), Nocco et al. (2019) study market distortions (on cutoff

[1]See, e.g., Zhelobodko et al. (2012) for the relation between relative love for variety and pro-competitive effects with general additive preferences in a closed economy. Bertoletti and Etro (2017) consider a tractable model with indirectly additive preferences exhibiting variable markups in a closed economy and a Krugman (1979) model with two identical countries. Mrázová and Neary (2017) establish a "demand manifold" characterized by the elasticity and convexity of inverse demand under a similar environment. Baqaee et al. (2023) demonstrate how an increase in market size affects welfare and allocational efficiency under a generalized demand system introduced by Matsuyama and Ushchev (2017).

[2]More recent research extended their results to cover other demand structures (e.g., Nocco et al., 2014, Bertoletti and Etro, 2021; Macedoni and Weinberger, 2022; Bagwell and Lee, 2023).

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

[4]We also employ CARA preferences as an example in our illustration. For the application of CARA preferences in the literature, see, e.g., Behrens and Murata (2012) and Behrens et al. (2014). The former paper also discusses the impact of trade on welfare and efficiency in the framework of Krugman (1979).

[5]Another relevant quantitative study is the one by Mrázová et al. (2021). They introduce Constant Revenue Elasticity of Marginal Revenue (CREMR) preferences and 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.

productivities, quantities, and masses of entrants) and consider the question of a globally optimal multilateral trade policy with asymmetric countries. With the assumption of symmetric countries and Melitz (2003) model, Melitz and Redding (2015) and Kokovin et al. (2022) consider the market misallocation from the perspective of a world planner who maximizes global welfare.

The present paper is also broadly related to work on the equilibrium characterization in two-country models, e.g., ones analyzing optimal trade policy.[6] The respective literature suggests that, when the market equilibrium is inefficient, the customary primal approach for equilibrium analysis is not applicable.

Overall, for tractability, most of the mentioned work employs CES or quasi-linear preferences, a specific functional form of the distribution of firm productivity, country symmetry or a small open economy, and partial equilibrium analysis (through the presence of an outside sector). In contrast, we provide a characterization under general additive VES preferences, nonparametric firm productivity, and country asymmetry with large open economies in general equilibrium.[7] For this case, we establish results including the existence and uniqueness of the equilibrium. And we illustrate that the conclusions drawn from closed-economy settings do not simply carry over to the open-economy case, even with identical countries, because domestic and foreign producers decide about their production for the respective domestic versus foreign markets.

The remainder of the paper is organized as follows. Section 2 outlines the general theoretical framework of market equilibrium. Section 3 describes the social optimum from the perspective of a world planner. Section 4 compares two equilibria. Sections 5 and 6 provide CES and VES examples, respectively. Section 7 further discusses the setup of the social planner. The last section concludes.

# 2 Decentralized Market Equilibrium

We develop a general equilibrium model of monopolistic competition and Melitz (2003) heterogeneous firms, where the consumer preferences are additively separable as in Zhelobodko et al. (2012), Dhingra and Morrow (2019), and Behrens et al. (2020). The demand

[6]See, for example, Demidova and Rodríguez-Clare (2009), Demidova and Rodriguez-Clare (2013), Felbermayr et al. (2013), Haaland and Venables (2016), Demidova (2017), Bagwell and Lee (2020), and Costinot et al. (2020).

[7]Our general model, especially, the discussion of the competition level, is related to Mayer et al. (2021). Developing a model with general additive preferences and multi-product firms, these authors illustrate how demand shocks induce increases in market-specific competition and generate cross-product reallocations within firms.

structure covers both CES and VES cases, indicating the possibility of variable markups at the equilibrium.

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 $u$, $p$, $q$, and $\pi$ to refer to utility, price, quantity, and profit, respectively. We index firms by their unique productivity $\varphi$, which in country $i$ is distributed with c.d.f. $G_i(\varphi)$. Finally, we refer to the mass of potential producers by $M$. Moreover, we use $w_i$ to denote the prevailing wage rate in origin $i$ and $x_{ij}$ to denote any generic double-indexed variable (a placeholder for prices, quantities, profits, etc.) pertaining to producers in $i$ in their sales to customers in $j$. Then, we will frequently normalize $x_{ij}$ by $w_i$ and indicate it by a tilde:

$$\tilde{x}_{ij} \equiv \frac{x_{ij}}{w_i}.$$

The consumers in each country have homogeneous 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 workers 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.**
*1.* $u(\cdot)$ *is additive, (strictly)concave, triple continuously differentiable, and* $\lim_{q\to+\infty} u'(q) = 0$.
*2.* $r_{u'}(q) \equiv -\frac{q \cdot u'''(q)}{u''(q)} < 2$.
*3. All integrals that occur and are composed of* $u(\cdot)$ *and* $G_i(\cdot)$ *converge.*

The first sub-assumption ensures the love for variety of consumers and one of the Inada conditions holds. The second sub-assumption ensures that the second-order condition for profit maximization holds under constant unit cost in our model. The third assumption ensures the convergence of the integrals discussed in this paper.

For later use, we define two demand-side elasticities, the elasticity of utility and of marginal utility, both with respect to the quantity $q$ consumed, as follows:

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

As we show later, $\frac{1}{1-r_u(q)}$ 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 associated social optimum. In the case of increasing markups, $(\frac{1}{1-r_u(q)})' > 0$, We further define $\overline{q} \equiv \min\{q \geqslant 0 \text{ s.t. } r_u(q) = 1\}$, which can be finite or not, ensuring that firms set a positive price.[8] With CES preferences, $1 - r_u(q) = \varepsilon_u(q)$, $\forall q \geq 0$, while the equality does not hold for other VES preferences. To establish the existence and uniqueness of equilibria, we further make the following assumption about the elasticity of utility:

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

Assumption 2 eliminates the case where the 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 equilibrium.

Overall, as in Behrens et al. (2020), our assumptions are less restrictive and cover a broader range of preferences compared to the ones stated in Dhingra and Morrow (2019). In particular, we relax the Inada condition of $\lim_{q \to 0} u'(q) = +\infty$ as well as the assumption of bounded elasticities, $0 < r_u(q), \varepsilon_u(q) < 1$.

## 2.1 General Model

Consider a world with two countries, labelled by $H$(Home) and $F$(Foreign), each with a single sector. At this stage, we do not make any assumptions about the relationship (e.g. exporting fixed cost is greater than the domestic one) or symmetry (e.g., the iceberg trade costs for exporting are the same for both countries) between exogenous variables. That is, the heterogeneity of countries and production decisions are carefully preserved. Instead, when discussing existence and uniqueness later, we derive a simple measure to summarize all impacts of these parameters. In what follows, we address the problems specific to each type of agent separately.

**Consumers**. The utility maximization by a representative consumer in destination $j$ can be cast in terms of the Lagrangian:

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

[8] When $f_{ij} = 0$, we require the marginal utility $u'(q)$ to be bounded and $\overline{q} < +\infty$ to guarantee the existence and uniqueness of equilibria. Specifically, with bounded marginal utility, the high-cost firms will eventually not be able to survive even without fixed costs, and variable marginal costs are sufficient to induce a selection of high-productivity firms into production.

where $p_{ij}(\varphi)$ and $q_{ij}(\varphi)$ are the price and consumption of a good produced by a supplier in origin $i$ with productivity $\varphi$. Of the $M_i$ potential producers in $i$, only $M_i \int_{\varphi_{ij}^*}^{+\infty} dG_i(\varphi) \leqslant M_i$ actually produce for/sell in $j$, where $\varphi_{ij}^*$ is the cutoff productivity of producers in $i$ selling in $j$.

This section uses $\delta_j$ to denote the Lagrange multiplier from the utility maximization under the decentralized market. In later sections, when comparing different equilibria, we use $\delta_j^{dmkt}$ instead when necessary. It is the marginal utility of income and is specific to the customer market $j$ and invariant to the suppliers' location.

The first-order conditions to the above maximization problem yield

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

The concavity of $u(q)$ ensures the second-order condition of utility maximization to be satisfied, and thus, the first-order condition is sufficient for a global solution.

**Firms**. Labor is the only factor of production, and its supply is inelastic. A firm in origin $i$ faces factor prices of $w_i$ per worker, and it delivers output to a customer's door in $j$ at (iceberg) transport costs of $\tau_{ij} \geq 1$. The production plus delivery costs per unit of shipment for a firm with productivity $\varphi$ are then $\frac{\tau_{ij} w_i}{\varphi}$. Besides, firms in origin $i$ incur a fixed cost $f_{ij} w_i$ to sell in destination $j$. Regarding to both $\tau_{ij}$ and $f_{ij}$, $j$ could be the same as $i$ or not. That is, the shopping costs and fixed costs exist for domestic and exporting production without further constraint. The corresponding profit maximization with respect to $q_{ij}(\varphi)$ then reads

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

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

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

$\frac{1}{1 - r_u(q_{ij}(\varphi))}$ is the private markup charged by a firm producing output $q_{ij}(\varphi)$. The second condition in Assumption 1 guarantees the second-order condition of profit maximization to be met (Zhelobodko et al., 2012).

Equ. (2) equates marginal costs and marginal real revenues under profit maximization. $\delta_j$ can be viewed as a demand shifter and a measure of the competition intensity in market $j$ (Mrázová et al., 2021; Mayer et al., 2021). While individual firms are atomistic and take $\delta_j$ as given, $\delta_j$ is endogenous to aggregate changes. From Equ. (2), we can obtain an implicit solution for $q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})$ and henceforth use the notation of $q_{ij}(\delta_j w_i, \varphi)$ for it without ambiguity. We can further express the corresponding profit as $\pi(\delta_j w_i, \varphi)$. Both quantity $q_{ij}(\delta_j w_i, \varphi)$ and profit $\pi(\delta_j w_i, \varphi)$ decrease with destination-specific competition intensity $\delta_j$ and with origin-specific $w_i$, and increase with firm-specific productivity $\varphi$.

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

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

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

which can be rewritten as $\tilde{\pi}_{ij}(\delta_j w_i, \varphi_{ij}^*) = 0$. Given that profit $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$ decreases with the competition intensity $\delta_j$, a higher $\delta_j$ leads to a higher cutoff productivity $\varphi_{ij}^*$ because it is less profitable for all firms in origin $i$ selling to market $j$ given an exogenous fixed cost $f_{ij}$.

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 face the zero expected profit condition (ZEPC):

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

We can reformulate the ZEPC as $\sum_j \tilde{\Pi}_{ij}(\delta_j w_i) = \mathcal{F}_i$. The expected profit $\tilde{\Pi}_{ij}(\delta_j w_i)$ decreases with $\delta_j$ and $w_i$ as well since the tougher competition level (higher wage) reduces profits for all operating firms (intensive margin) and increase the cutoff productivity (extensive margin). Note that the ZEPC in each country incorporates the competition intensities in both countries, $\delta_H$ as well as $\delta_F$.

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

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

From the resource constraints, we can express the masses of entrants $M_i$ as a function of the competition intensity $\delta_j$ and wage $w_i$.

Finally, with a numeraire, the relative wages between countries are established by the trade-balance condition (TBC):

$$\begin{aligned} &M_H \int_{\varphi_{HF}^*}^{+\infty} \frac{1}{1 - r_u(q_{HF}(\delta_F w_H, \varphi))} \frac{\tau_{HF} w_H}{\varphi} q_{HF}(\delta_F w_H, \varphi) L_F dG_H(\varphi) \\ =&M_F \int_{\varphi_{FH}^*}^{+\infty} \frac{1}{1 - r_u(q_{FH}(\delta_H w_F, \varphi))} \frac{\tau_{FH} w_F}{\varphi} q_{FH}(\delta_H w_F, \varphi) L_H dG_F(\varphi). \end{aligned} \tag{6}$$

Without loss of generality, we choose $w_H = 1$ as the numeraire, and thus Foreign's wage $w_F$ is endogenous. In what follows, for clarity, we might keep or omit $w_H$ in different cases.

## 2.2 Existence and Uniqueness

In the context of examining a general additive demand and a general productivity distribution without imposing constraints on exogenous parameters (e.g., avoiding a focus on free trade or symmetric economics), establishing the existence and uniqueness of the equilibrium is challenging. Except for losing the tractability of CES demand, the alternative choice for export/import and the heterogeneity of countries lead to the exponential increase of exogenous parameters, which further interact with each other at different stages of models. Therefore, before diving into the details, we first outline our approach to proof as follows:

1) *Unique outcome variables given $\delta_j w_i$*: From equ. (2)-(5), we show that all outcome variables, including quantity $q_{ij}(\varphi)$, profit $\tilde{\pi}_{ij}(\varphi)$, average profit $\tilde{\Pi}_{ij}$, and mass of trants $M_i$ can be expressed as functions of destination-specific competition intensity $\delta_j$ and origin-specific wage $w_i$.

2) *Unique $\delta_j(w_F)$ given $w_F$*: We then prove that for a given foreign wage $w_F$, there exists a unique solution of $(\delta_H^*(w_F), \delta_F^*(w_F))$ such that in both countries, the zero-expected-profit conditions in equ. (4) hold. Therefore, we also obtain the corresponding solutions of quantity, profit, and mass of entrants such that equ. (2), (3), and (5) jointly hold.

3) *Unique $w_F$*: We show that there exists a unique foreign wage $w_F^*$ such that the trade-balance condition in equ. (6) holds. Overall, we prove that the decentralized market equilibrium is uniquely determined.

**Step 1)**. From the inverse demand in equ. (2), we obtain an implicit function of quantity $q_{ij}(\delta_j w_i, \varphi)$ and naturally, profit $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$. Combine $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$ with the ZCPC in equ. (3), we deduce the cutoff productivity $\varphi_{ij}^*(\delta_j w_i)$. Similarly, average profit can be computed as $\tilde{\Pi}_{ij}(\delta_j w_i) = \int_{\varphi_{ij}^*(\delta_j w_i)}^{+\infty} \tilde{\pi}_{ij}(\delta_j w_i, \varphi) dG_i(\varphi)$. Finally, with all these expressions on hand, we can determine the mass of entrants $M_i(\delta_j w_i)$ from the resource constraint in equ. (5). We summarize these results in Lemma 1 and provide detailed proof in the Appendix.

**Lemma 1** (**Market equilibrium outcomes**). *Given $\delta_j w_i$, we can uniquely determine the equilibrium quantities $q_{ij}(\delta_j w_i, \varphi)$, cutoff productivity $\varphi_{ij}^*(\delta_j w_i)$, profit $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$, average profit $\tilde{\Pi}_{ij}(\delta_j w_i)$, and the mass of entrants $M_i(\delta_j w_i)$. Specifically, when $\delta_j w_i$ increases, $q_{ij}(\delta_j w_i, \varphi)$, $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$, and $\tilde{\Pi}_{ij}(\delta_j w_i)$ decrease, and $\varphi_{ij}^*(\delta_j w_i)$ increases.*

*Proof.* See Appendix A.1. □

Lemma 1 illustrates the relation between equilibrium outcomes and destination-specific competition intensity $\delta_j$ as well as origin-specific $w_i$. Consider the case of an increase of $\delta_j w_i$. From the inverse demand in equ. (2), all sales from origin $i$ to destination $j$, $q_{ij}(\delta_j w_i, \varphi)$, decrease since the product of competition intensity in $j$ and factor price in $i$ increases. Therefore, all corresponding profits $\tilde{\pi}_{ij}(\delta_j w_i, \varphi)$ decrease as well, which leads to a higher cutoff productivity $\varphi_{ij}^*(\delta_j w_i)$, since firms now will find it harder to gain positive profit for a given fixed cost $f_{ij}$. As a result, the average profits of all firms, $\tilde{\Pi}_{ij}(\delta_j w_i)$, are lower.

**Step 2)**. In this part, we show the existence and uniqueness of $\delta_j$ conditional on a given foreign wage $w_F$. The zero-expected-profit conditions in both countries represent a system of two equations in the two unknowns $\delta_H$ and $\delta_F$. To establish the existence and uniqueness of the solution, we need to analyze the properties of two functions, including monotonicity, differentiability, the existence of implicit functions, and their limits. In what follows, we start with the limits.

To provide a lucid illustration, let us examine the ZEPC in the home country, given by $\tilde{\Pi}_{HH}(\delta_H) + \tilde{\Pi}_{HF}(\delta_F) = \mathcal{F}_H$. When $\delta_F$ increases to infinity, a situation arises where nearly all firms in the home country cannot generate positive profits from exporting, causing the average profit from exporting, $\tilde{\Pi}_{HF}(\delta_F)$, to approach zero. To fulfill the ZEPC, $\delta_H$ must converge to a positive value such that the average profit from domestic sales equals the sunk cost, i.e., $\tilde{\Pi}_{HH}(\delta_H) = \mathcal{F}_H$. One can see that the limit case coincides with the market equilibrium under a closed economy. Similarly, when $\delta_H$ tends towards infinity, $\delta_F$ converges to a positive value that ensures the ZEPC is satisfied. In this scenario, all firms in Home exclusively sell to Foreign. In summary, each origin can only sell to a single destination at the limit. Consequently, we introduce the following counterfactual equilibrium to chatacterize these limits.

**Lemma 2** (**Counterfactual equilibrium**). *A counterfactual equilibrium characterizes the case where firms in origin $i$ are restricted to sell only to destination $j$. Specifically, $\forall i, j = H, F$, $\left\{\delta_{ij} w_i, \varphi^*_{ij}, q_{ij}(\varphi)\right\}$ is the counterfactual equilibrium solution to the following conditions:*

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

*The counterfactual equilibrium is uniquely determined. Furthermore, the solution of $\delta_j w_i$ decreases with $\tau_{ij}$, $f_{ij}$, and $\mathcal{F}_i$ and increases with $L_j$.*

*Proof.* See Appendix A.1. □

While the equilibrium value of $\delta_j w_i$ is endogenously determined by a combination of all exogenous parameters, we assume it as a given constant for the purposes of proof related to open economies. Given the definition of the counterfactual equilibrium, the average profits $\tilde{\Pi}_{ij}(\delta_{ij} w_i)$ equal the sunk cost in origin $i$, which is the limit case of the standard market equilibrium with international trade. As we show in Lemma 2, the equilibrium value of $\delta_{ij} w_i$ decreases with all types of production costs and increases with the market size, and thus, it can be considered as a measure of inverse aggregate cost for firms in origin $i$ to sell to destination $j$. In what follows, we make the assumption about the product of aggregate costs for domestic sales being lower than that of imports and further discuss the properties of such aggregate costs.

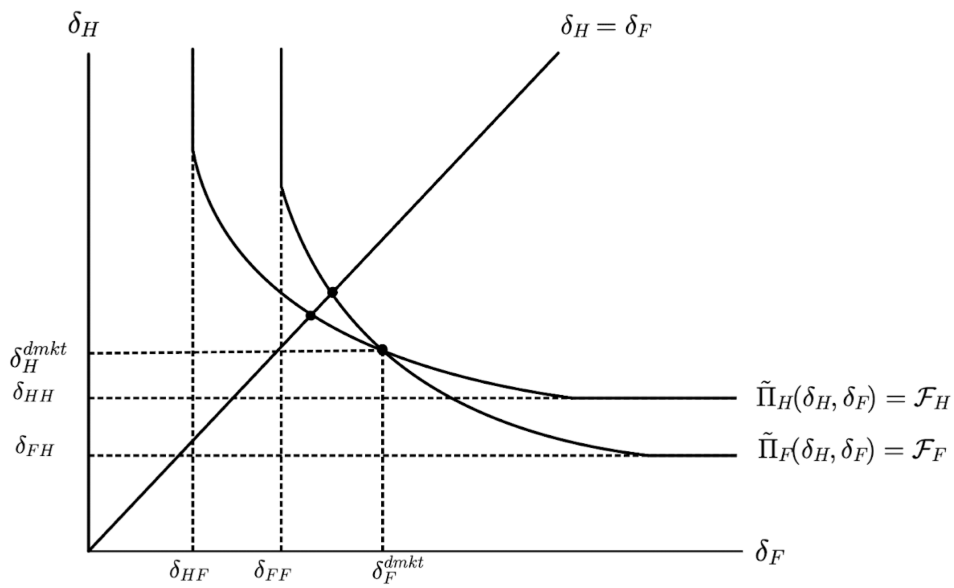
**Assumption 3.** *Define the measure of aggregate cost for firms in origin $i$ to sell to destination $j$ as $\mathcal{C}_{ij} = (\delta_{ij} w_i)^{-1}$. Assume $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$.*

According to Lemma 2, $\mathcal{C}_{ij}$ are uniquely determined by exogenous parameters such as market size and trade, fixed, and sunk costs. From the perspective of firms, they take $\mathcal{C}_{ij}$ as given and maximize their own profits. Therefore, the aggregate cost measure $\mathcal{C}_{ij}$ is a common shifter for all firms located in $i$ to sell in $j$. Assumption 3 imposes a joint constraint on the space for all exogenous parameters, and, as shown later, guarantees the existence of the market equilibrium.

We have two further comments on Assumption 3. First, $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$ ensures the product of aggregate importing (exporting) costs is greater than that of aggregate domestic costs, which is intuitively assigned with the assumptions made in well-known trade models. (e.g., Krugman, 1980; Eaton and Kortum, 2002; Melitz, 2003; Chaney, 2008).

Second, Assumption 3 guarantees the admissible range for endogenous wage $w_F$ at the partial equilibrium mentioned in Step 2). That is, in the following proof, we consider the existence and uniqueness of $(\delta_F, \delta_H)$ for a given $w_F \in (\underline{w_F}, \overline{w_F}) = (\frac{\mathcal{C}_{HH}}{\mathcal{C}_{FH}}, \frac{\mathcal{C}_{HF}}{\mathcal{C}_{FF}})$. For illustration, we rewrite the assumption in terms of $\delta_{ij}(w_i)$, and one can further see that $w_F > \underline{w_F}$ and $w_F < \overline{w_F}$ ensure $\delta_{FH}(w_F) < \delta_{HH}$ and $\delta_{HF} < \delta_{FF}(w_F)$, respectively. These four values of $\delta_{ij}(w_i)$ correspond to the four limits of two ZEPCs, equ. (4), in both countries, as we show in Figure 1.[9] Therefore, by considering $w_F$ in this admissible range, we guarantee the limit of imported competition intensity, $\delta_{ij}(w_i)$, is lower than the corresponding limit of domestic competition intensity, $\delta_{jj}(w_j)$, $\forall i, j = H, F$ and $i \neq j$.

Figure 1: Zero-Expected-Profit Conditions for a Given $w_F \in (\underline{w_F}, \overline{w_F})$




After discussing the limit properties of ZEPCs, we now return to the standard decentralized market equilibrium and delve into the remaining properties. These properties en-

[9]The curves in Figure 1 include kinks, because, for some preferences, there exists an upper bound for quantity, $\overline{q}$, to guarantee the corresponding markups and prices to be positive. We do not provide labels for these kinks to maintain simplicity. Corresponding details are discussed in the Appendix. Besides, the intersections between curves and the 45-degree line characterize the centralized market equilibrium, a topic we explore in detail in Section 3.1.

compass monotonicity, differentiability, and the existence of implicit functions. Specifically, when considering a fixed endogenous foreign wage $w_F \in (\underline{w_F}, \overline{w_F})$, the zero-expected-profit conditions take the following form:

$$\begin{bmatrix} \tilde{\Pi}_H(\delta_H, \delta_F) \\ \tilde{\Pi}_F(\delta_H, \delta_F) \end{bmatrix}_{|w_F} = \begin{bmatrix} \tilde{\Pi}_{HH}(\delta_H) + \tilde{\Pi}_{HF}(\delta_F) \\ \tilde{\Pi}_{FH}(\delta_H) + \tilde{\Pi}_{FF}(\delta_F) \end{bmatrix}_{|w_F} = \begin{bmatrix} \mathcal{F}_H \\ \mathcal{F}_F \end{bmatrix}.$$

**Assumption 4.** *$\forall w_F \in (\underline{w_F}, \overline{w_F})$, when $\tilde{\Pi}_H(\delta_H, \delta_F|w_H)$ and $\tilde{\Pi}_F(\delta_H, \delta_F|w_F)$ are both first-order continuously differentiable in $\delta_H$ and $\delta_F$,*[10] *the conditional Jacobian determinant is positive:*

$$|\mathbf{J}(\delta_H, \delta_F|w_F)| = \begin{vmatrix} \frac{\mathbf{d}\tilde{\Pi}_{HH}(\delta_H)}{\mathbf{d}\delta_H}, & \frac{\mathbf{d}\tilde{\Pi}_{HF}(\delta_F)}{\mathbf{d}\delta_F} \\ \frac{\mathbf{d}\tilde{\Pi}_{FH}(\delta_H)}{\mathbf{d}\delta_H}, & \frac{\mathbf{d}\tilde{\Pi}_{FF}(\delta_F)}{\mathbf{d}\delta_F} \end{vmatrix}_{|w_F} > 0.$$

We further define the wage-adjusted average revenue per capita for firms in origin $i$ selling to destination $j$ as $\tilde{\mathcal{R}}_{ij}(\delta_j) = \int_{\varphi^*_{ij}(\delta_j)}^{+\infty} \tilde{p}_{ij}(\delta_j, \varphi)\, q_{ij}(\delta_j, \varphi)\, dG_i(\varphi)$. Then Assumption 4 can be rewritten as

$$\tilde{\mathcal{R}}_{HH}(\delta_H)\tilde{\mathcal{R}}_{FF}(\delta_F) - \tilde{\mathcal{R}}_{FH}(\delta_H)\tilde{\mathcal{R}}_{HF}(\delta_F) > 0.^{11}$$

Therefore, Assumption 4 requires that, conditional on $w_F$, for any values of competition intensities $(\delta_H, \delta_F)$ such that all average profits are positive, the product of domestic revenues is greater than that of export revenues.

With Assumptions 3 and 4, we establish the existence and uniqueness of the solution $(\delta^*_H(w_F), \delta^*_F(w_F))$ for a given $w_F \in (\underline{w_F}, \overline{w_F})$, as in Figure 1, such that the utility and profit maximization (2), ZCPCs (3), and ZEPCs (4) hold in both countries. The corresponding masses of entrants can then be solved from the labor-market-clearing conditions (5). Specifically, Assumption 3 ensures existence and Assumption 4 ensures uniqueness. Besides, these two assumptions jointly guarantee that there is no solution when $w_F < \underline{w_F}$ or $w_F > \overline{w_F}$.[12]

**Step 3)**. Combining the resource constraints (5) with the trade-balanced condition (6),

[10] The case that $\tilde{\Pi}_H(\delta_H, \delta_F)$ is indifferentiable with respect to $\delta_H$ or $\delta_F$ only happens when the demand functions exhibit an upper bound for production $\overline{q}$ (for example, CARA preference). In such instances, when the competition intensity $\delta_j$ becomes excessively high, firms can not gain positive profits at the maximum production $\overline{q}$. Consequently, aggregate profits remain at zero and do not respond to changes in $\delta_j$. See the Appendix for more details.

[11] To derive this, we employ the formula of elasticity of operating profits with respect to $\delta_j$, $\varepsilon_{\tilde{\pi}_{ij}+f_{ij},\delta_j} = -(\frac{1}{r_u(q_{ij}(\delta_j,\varphi))})$, which is showed by Mayer et al. (2021).

[12] When $w_F = \underline{w_F}$ or $w_F = \overline{w_F}$, an infinite number of solutions emerges, where all products from different origins are exclusively sold to a single destination. These solutions fail to satisfy the trade-balance condition and do not align with real-world observations.

we can re-write the condition (6) as:

$$L_H w_H \frac{\tilde{\mathcal{R}}_{HF}(\delta_F w_H) L_F}{\sum_j \tilde{\mathcal{R}}_{Hj}(\delta_j w_H) L_j} = L_F w_F \frac{\tilde{\mathcal{R}}_{FH}(\delta_H w_F) L_H}{\sum_j \tilde{\mathcal{R}}_{Fj}(\delta_j w_F) L_j}.$$

In Step 2), we showed that, for a given $w_F \in (\underline{w_F}, \overline{w_F})$, a unique solution $(\delta_H^*(w_F), \delta_F^*(w_F))$ exists. Consider a scenario where $w_F$ decreases to $\underline{w_F}$. The equilibrium value of $\delta_H^*(w_F)$ then converges to $\delta_{HH}$, while $\delta_{FH}(w_F)$ simultaneously converges to $\delta_{FH}(\underline{w_F}) = \delta_{HH}$. Consequently, all firms in country $H$ only sell domestically, and all firms in country $F$ only export to $H$. For the trade-balance condition, as $w_F$ approaches $\underline{w_F}$, the LHS converges to zero and the RHS converges to $L_F \underline{w_F}$. Similarly, as $w_F$ increases to $\overline{w_F}$, the LHS of the trade-balance condition converges to $L_H w_H = L_H$, and the RHS converges to zero. Therefore, with the monotonicity of $(\delta_H^*(w_F), \delta_F^*(w_F))$ with respect to $w_F$, there exists a unique solution $w_F^*$, such that trade between the two countries is balanced, and $(\delta_H^*(w_F^*), \delta_F^*(w_F^*))$ guarantees that all other equilibrium conditions hold.

Therefore, our three-step proof culminates in the establishment of the existence and uniqueness of the decentralized market equilibrium, as stated in the following proposition:

**Proposition 1.** *Under Assumptions 1-4, the (decentralized) market equilibrium* $\{\delta_j,\ w_i,\ \varphi_{ij}^*,\ q_{ij}(\varphi),\ M_i,\ \forall i, j = H, F\}$ *is uniquely determined.*

*Proof.* See Appendix A.2. □

To sum up, Proposition 1 establishes the existence and uniqueness of market equilibrium without making particular assumptions on exogenous parameters, such as $\tau_{ij}$, $f_{ij}$, $\mathcal{F}_i$, and $L_j$. Instead, we introduce a simple measure of origin-destination-specific aggregate cost, $\mathcal{C}_{ij}$, and further employ this measure along with another monotonicity condition to derive sufficient conditions. Our methodology allows the full interaction between parameters at the different stages of the model and captures the general equilibrium effect without the commonly used assumptions, such as symmetric countries, zero fixed cost, or an outside sector. In other words, our sufficient conditions encompass a significantly larger space for exogenous parameters, as we later show examples in Section 5 and 6. Before providing two examples of our general model, we first discuss the efficiency of market equilibrium in the next section.

# 3 Utilitarian Global Planner Social Optimum

In this section, we delve into the comparison between the market equilibrium and the social optimum, examining it through the lens of a global social planner. In the literature, a global planner is often employed to emphasize the efficiency properties of market equilibrium, while abstracting from the incentive of national planners to influence the terms-of-trade effect (Behrens and Murata, 2012; Melitz and Redding, 2015; Kokovin et al., 2022).

However, defining an objective function for a global social planner poses a pivotal and sometimes contentious challenge, particularly when the task involves aggregating welfare across countries. As a reasonable reference point, We utilize a benevolent utilitarian global planner whose objective is to maximize the global aggregate households' utility. As articulated in Welch (1987), "utilitarianism dominates the landscape of contemporary thought in the social sciences." In the field of economics, Burk (1938) is often credited as the first treatise to introduce a social welfare function that aggregates all individuals' utility. Although the utilitarian approach had been criticized by Rawlsian (Rawls, 1958; Rawls, 1974) and Libertarianist contenders (Hayek, 1944; Nozick, 1974), the utilitarian social welfare function has remained a widely used baseline for evaluating efficiency and distributional considerations (Sen, 1997).[13]

To be specific, we consider the global planner who maximizes the global aggregate utility with the ability to choose quantities, cutoffs, and the masses of entrants. By being focused on the aggregate, the planner is agnostic about the distribution of welfare across countries and treats households from different destinations equally.[14] Toward the end of the paper, we will delve into the scenario, where a utilitarian global social planner is allowed to assign different weights to countries. However, for now, our focus remains on the following planner's problem:

**Proposition 2 (utilitarian global social optimum).** *A (global) social optimum is the solution of the following Lagrangian:*

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

[13]For a detailed discussion of utilitarianism in the context of welfare economics and international trade, see Shelburne (2006).

[14]One could consider an isomorphic alternative problem instead, where a national planner maximizes the aggregate welfare of a country with two regions.

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

*Proof.* See Appendix A.3. □

Throughout our analysis of the planner's problem, we assume the sufficiency of the optimal conditions.[15] Note that 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 concerning quantities are as follows:

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

which equates marginal utility for consumers and marginal cost for firms. $\lambda_i^{opt}$ is the marginal utility of the 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

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

and

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

$\frac{1}{\varepsilon_u(q_{ij}(\lambda_i^{opt}, \frac{\tau_{ij}w_i}{\varphi}))}$ is the social markup a utilitarian global social planner assigns a firm with productivity $\varphi$ to charge. Equs. (8) and (9) are the zero-cutoff-social-profit condition (ZCSPC) and the zero-expected-social-profit condition (ZESPC). With these equations in place, we can establish 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. First, $\lambda_i^{opt}$ serves as a demand shifter and an amplifier of firm productivity from the social planner's perspective. Firms' production and social profits decrease with $\lambda_i^{opt}$, similar to firms' production and profits decrease with $\delta_j^{dmkt}$ in the market equilibrium. However, the impact of $\lambda_i^{opt}$ is origin-specific rather than destination-specific and, importantly, it can be determined from the ZESPC of origin $i$, independent of that of destination $j$. In contrast, the demand shifters in the

[15] See Online Appendix B.1 for a detailed discussion.

decentralized equilibrium, $\delta_j^{dmkt}$, are destination-specific and have to be jointly solved from the ZEPCs in both countries.

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 the social optimum. Consequently, we can require the social planner to reallocate resources under the market-equilibrium wages.

Last, it is essential to note that the pricing strategies differ between the decentralized market equilibrium versus the social optimum. In the decentralized market equilibrium, a firm with productivity $\varphi$ sets a private markup of $\frac{1}{1-r_u(q_{ij}(\delta_j^{dmkt},\frac{\tau_{ij}w_i}{\varphi}))}$. In contrast, the firm 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 should further note that the Lagrange multipliers misalignment also affects the markup misalignment.

In relation to the existing literature, it is worth noting that the first two channels, namely shifters and wages, are missing in the models of a closed economy (e.g., Dhingra and Morrow, 2019; Behrens et al., 2020; Bagwell and Lee, 2021; Baqaee et al., 2023; Bagwell and Lee, 2023) and also the models of open economies with an outside sector or country symmetry (e.g., Behrens and Murata, 2012; Nocco et al., 2019; Bagwell and Lee, 2020). Additionally, the last channel associated with markups is absent in models of monopolistic competition and CES preferences,[16] as discussed in Zhelobodko et al. (2012), Parenti et al. (2017), Dhingra and Morrow (2019), and Behrens et al. (2020).

Therefore, comparing the two types of equilibria is indirect and relatively more complicated than in earlier work. An ideal way to tackle this complexity is to dissect and analyze the misalignment channels separately. Our solution is to employ the method of Dhingra and Morrow (2019) by considering an auxiliary planner's problem, in which case a planner has a real-revenue-maximizing objective. This planner's problem guarantees that the corresponding demand shifters are origin-specific, and the real-revenue maximization ensures that the planner uses the pricing strategies akin to those used by firms in the decentralized market equilibrium.

## 3.1 Centralized Market Equilibrium

To construct the auxiliary equilibrium, we evaluate equ. (2) at the equilibrium and express it as $u'(q_{ij}(\varphi))\,[1-r_u(q_{ij}(\varphi))] = \frac{\delta_j^{dmkt}\tau_{ij}w_i}{\varphi}$. The latter equates the marginal real revenue to

[16] The CES preferences, along with oligopolistic competition, can induce variable markups. See Atkeson and Burstein (2008) for details.

the marginal cost in real terms from the perspective of firms. Integrating the marginal real revenue, we arrive at $\int u'(q_{ij}(\varphi))\,[1 - r_u(q_{ij}(\varphi))]\,\mathrm{d}\,q_{ij}(\varphi) = u'(q_{ij}(\varphi))q_{ij}(\varphi)$, which represents the real revenue for a firm selling quantity $q_{ij}(\varphi)$. We dub the solution to a planner's problem of maximizing global aggregate real revenue a centralized market equilibrium. In this scenario, the planner is entitled and able to choose quantities, cutoff productivities, and masses of entrants.

**Proposition 3** (**Centralized market equilibrium**). *A (global) centralized market equilibrium is the solution of the following Lagrangian:*

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

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

*Proof.* See Appendix A.4. □

In a closed economy, the outcomes of decentralized and centralized market equilibria align perfectly. However, the equivalence does not hold in the open economies. Under the decentralized market equilibrium, firms maximize 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 contrast, the centralized market equilibrium emerges when the planner maximizes aggregate real revenues for firms subject to resource constraints in origins. In realization, a real-revenue maximizing firm with productivity $\varphi$ is indifferent between domestic and export production. Thus, the centralized market equilibrium is characterized by origin-specific multipliers. We will use the same acronym $\delta_i^{cmkt}$ 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 $\delta_i^{cmkt}$, which is origin-specific − see equs. (2), (3), (4), (5). As a result, the pricing strategies of firms and the shapes of the production schedules are the same between the centralized and decentralized market equilibria. In Figure 1, the intersections between the curves of zero-expected profit conditions and the 45-degree line characterize the solutions of $\delta_i^{cmkt}$.

From a mathematical standpoint, solving the centralized market equilibrium is distinct from solving the decentralized market equilibrium in that it does not require the simultaneous consideration of zero-expected-profit conditions (ZEPCs) for all countries. With the existence and uniqueness of the centralized market equilibrium established, it serves as an ideal auxiliary 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 the same, except for their respective 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.

# 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. 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 will discuss each of the three pairwise comparisons separately: the decentralized market equilibrium versus the centralized market equilibrium, the centralized market equilibrium versus the social optimum, and the decentralized market equilibrium versus the social optimum.

## 4.1 Decentralized vs. Centralized Market: Quantity-locus Shift

First, it is important to note that when countries are identical, or in cases where an outside sector exists, the competition intensities take on identical values: $\delta_H^{dmkt} = \delta_F^{dmkt} = \delta_H^{cmkt} = \delta_F^{cmkt}$. Consequently, competition levels remain consistent across different countries, leading the associated open-economy models to reflect a partial equilibrium effect.

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 sole distinction between the two equilibria lies in the properties of these 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: $q_{\ell i}^{dmkt}(\varphi) \leqslant q_{\ell i}^{cmkt}(\varphi)$, $q_{\ell j}^{dmkt}(\varphi) \geqslant q_{\ell j}^{cmkt}(\varphi)$.*
- *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, all inequalities strictly hold.*[17]

*Proof.* See Appendix A.5. □

Proposition 4 indicates, 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. These results are also evident from an inspection of Figure 1.

For intuition, consider that a real-revenue-maximizing planner faces the realization of the decentralized market equilibrium $\{\delta_H^{dmkt}, \delta_F^{dmkt}\}$ with $\delta_H^{dmkt} < \delta_F^{dmkt}$. 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 $H$, $\tilde{\Pi}_{HH}(\delta_H^{dmkt}) + \tilde{\Pi}_{HF}(\delta_F^{dmkt}) = \mathcal{F}_H$, the planner will increase aggregate real revenue in origin H by adjusting

[17] All inequalities also strictly hold when $\overline{q} = +\infty$ or $f_{ij} = 0$, $\forall i, j$.

the shifters to equalize the marginal real revenue of resources for firms in $H$ for both domestic and exporting sales. We denote the resulting marginal real revenue of resources in the ZEPC of country H as $\delta_H^{cmkt}$, which guarantees $\tilde{\Pi}_{HH}(\delta_H^{cmkt}) + \tilde{\Pi}_{HF}(\delta_H^{cmkt}) = \mathcal{F}_H$, and explore its properties.

**Boundedness**. To adjust the resource allocation between domestic and export production while maintaining the ZEPC, $\delta_H^{cmkt}$ must be between $\delta_H^{dmkt}$ and $\delta_F^{dmkt}$, with the properties of production shown in the proof of Lemma 2. The planner will adjust $\delta_H^{dmkt}$ and $\delta_F^{dmkt}$ towards the common equilibrium value and thus, the domestic and export output move in opposite directions.

**Consistency**. The boundedness property allows us to view $\delta_H^{cmkt}$ as a weighted average of $\delta_H^{dmkt}$ and $\delta_F^{dmkt}$. 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 $\delta_H^{dmkt}$ should be greater than that of $\delta_F^{dmkt}$ in order to obtain the equilibrium value of $\delta_H^{cmkt}$. This means $\delta_i^{cmkt}$ should be closer to $\delta_i^{dmkt}$ than to $\delta_j^{dmkt}$ for all $i \neq j$. The boundedness property implies that the ranking of demand shifters remains consistent in both equilibria: if $\delta_H^{dmkt} < \delta_F^{dmkt}$, then $\delta_H^{cmkt} < \delta_F^{cmkt}$. As for the outcomes, the centralized planner's adjustment on the market equilibrium will be greater for exports than for the domestic market.

In summary, we show the ranking of demand shifters in the decentralized and centralized market equilibria. Unlike models with the assumptions of a closed economy or symmetric countries, the country-level fundamental differences generally result in differing marginal utilities of income, and thus, the centralized planner will adjust domestic and export production to equalize the marginal real revenue of resources. We further show that these adjustments are opposite across different destinations and are greater for exports than for domestic production. As a consequence, trade generally induces an extra channel of difference when comparing the decentralized market equilibrium with a utilitarian global social optimum.

## 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 utilitarian global social optimum. Both are optimization problems from the planner's perspective, subjective to the same resource constraints but with different objectives. Consequently, the comparison between the centralized market and the socially optimal equilibrium can be conducted in the spirit of Dhingra and Morrow (2019). While most of the analysis in this section is similar to the discussion in the closed economy, we emphasize that the results in the closed economy can not be simply carried over to the open economies because now firms have an additional decision on exporting. To start with, we make the following assumptions about markups.

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

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 assumptions of interior markups and allows our analysis to incorporate a larger group of preferences. 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.

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 | $0 < 1-\eta < 1-\theta < 1$ | $\alpha > 0, 0 < \rho < 1$ | $a > 0, 0 < p < 1$ |
| $u'(q)$ | $> 0$ | $> 0$ | $> 0$ |
| $\lim_{q\to+\infty} u'(q)$ | $0$ | $0$ | $0$ |
| $u''(q)$ | $< 0$ | $< 0$ | $< 0$ |
| $\lim_{q\to 0}\varepsilon_u(q)$ | $1-\eta$ | $1$ | $1-\rho$ |
| $[1-r_u(q)]'\,[\varepsilon_u(q)]'$ | $> 0$ | $> 0$ | $> 0$ |
| $\lim_{q\to+\infty}\varepsilon_u(q)$ | $1-\theta$ | $\rho$ | $0$ |
| $\lim_{q\to+\infty} 1-r_u(q)$ | $1-\theta$ | $\rho$ | $-\infty$ |

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.

**Proposition 5** (**Quantity distortions**). *$\forall i, j = H, F$, $q_{ij}^{cmkt}(\varphi)$ and $q_{ij}^{opt}(\varphi)$ have a unique intersection $\tilde{\varphi}_{ij}$:*[18]

- *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.*

*Proof.* See Appendix A.6. □

Proposition 5 relies on a portrait of the quantity schedules $q_{ij}^h$ with $h \in \{cmkt, opt\}$ as a function of productivity $\varphi$, each. As with a closed economy, 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 also be noted that the direction of rotation is independent of the fundamentals of the countries.

We do not repeat the explanation of the 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 from a social planner's perspective; whether a firm overproduces or underproduces depends on the preference of consumers and its specific productivity. However, one should note that the extent of rotation differs across the destinations of sales, because the variable markups 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 selection 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 for firms in origin $i$ to sell. For simplicity, we rewrite the cutoff productivity corre-

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

sponding to $\underline{j}^i$ as $\varphi^*_{i\underline{j}}$ rather than $\varphi^*_{i\underline{j}^i}$. This definition allows us to obtain the ranking of inverse fixed costs per consumer, which are the key parameters for cutoff distortions.

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

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

*Specially, when all fixed costs are zero:*

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

*Proof.* See Appendix A.7. □

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 costs per consumer dominates the business-stealing effect, encouraging the production of the marginal variety and decreasing the cutoffs in the centralized market equilibrium. However, 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 of asymmetric market sizes, and when two countries are identical without further assumptions on the ranking of fixed costs. 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 as in the closed economy.

Overall, the comparison between the quantity loci in the centralized market equilibrium and the utilitarian global social optimum can be summarized as a rotation effect, which depends on the demand-side elasticities and on the fixed cost per capita of domestic and export production. However, the centralized market equilibrium is constructed as an intermediate case to compare 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 do not simply extend to asymmetric 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. Put differently, as we show in Proposition 4, the dispersion between $\delta_i^{dmkt}$ and $\delta_j^{dmkt}$ is greater than that between $\delta_i^{cmkt}$ and $\delta_j^{cmkt}$. 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.[19] Below, we will combine the shift and the rotation effects to discuss the distortions in a world with asymmetric countries.

## 4.3 Decentralized Market vs. Social Optimum: Elusiveness

In the open-economy case, the comparison of outcomes is elusive between the decentralized market equilibrium and the utilitarian global 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.

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 utilitarian 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.

[19] 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.

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 firms in one origin-destination pair may all overproduce or underproduce.

Regarding entry, there are no explicit general results in comparing the decentralized market and the social planner's equilibrium. 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 similarly, because $M_i^{dmkt}$ now depends on the competition intensities of all destinations. In what follows, we discuss some specific cases for illustration, here.

**Free trade**. With free trade among two asymmetric countries, the rotation effects including the intersections on domestic and export quantities are the same, when disregarding the cutoff productivities. However, cutoff distortions in the rotation effects persist due to 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-shifting effects persist.

**Free trade and symmetric countries with positive fixed market-access costs.** In this case, the rotation effects on domestic and export quantities are identical. However, the cutoff distortions in the rotation effects are elusive since the fixed costs for domestic and export production are different. The shift effects driven by destination-specific competition disappear with the symmetry of countries.

This analysis presents three examples that highlight the fundamental contrasts in conclusions drawn from models of general open economies as opposed to those based on closed or simplified open economies. When we align our broader analysis with existing literature, we observe that distortions brought about by varying intensities of competition are notably absent in closed economy models (e.g., Dhingra and Morrow, 2019; Behrens et al., 2020) as well as in models of open economies that feature either an external sector (Nocco et al.,

2019) or symmetric countries (Melitz and Redding, 2015; Baqaee et al., 2023), in which obtains the conclusion that market inefficiency is independent of countries. As discussed earlier, heterogeneity in competition intensities across destinations renders the misallocation in open economies generally ambiguous. The following two sections will delve into this complexity by examining cases of CES and VES preferences.

# 5 Example: CES Preferences

## 5.1 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) or symmetric open economies (see Melitz and Redding, 2015). In this section, we prove that CES preferences do not guarantee efficient cutoff levels and outputs in our single-sector open economy model.

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

*Proof.* See Appendix A.8. □

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, $\delta_j^{dmkt}$, 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 reasons why our conclusion for open economies differs from the closed-economy one are as follows. With a CES demand, all firms charge a constant markup, which is the same in both the market equilibrium and the utilitarian global 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 marginal real revenues of resources are equal across asymmetric destinations, leading to reallocations in market-pair-specific quantities and cutoff productivities.

### 5.2 A Special Case with Pareto-distributed Productivity

We further provide an example with CES demand and Pareto productivity to explain the results of the existence and uniqueness of decentralized market equilibrium in Proposition 1 and of efficiency in Proposition 7 explicitly. The detailed proof is provided in Appendix A.9. 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 $\frac{1}{\rho}$. The productivity distribution is $G(\varphi) = 1 - (\frac{1}{\varphi})^{\gamma}$ with $\gamma > 1$.

Assumption 3 requires that $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$, where $\mathcal{C}_{ij} = (\delta_{ij} w_i)^{-1}$ is the measure of aggregate cost. With the CES demand structure and the Pareto productivity distribution, we can obtain the expression of $\mathcal{C}_{ij}$ and the admissible range for endogenous wage $w_F$:

$$w_F \in (\underline{w_F}, \overline{w_F}) = \left(\frac{\mathcal{C}_{HH}}{\mathcal{C}_{FH}}, \frac{\mathcal{C}_{HF}}{\mathcal{C}_{FF}}\right) = \left( (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{\rho}{\gamma}} (\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{HH}}{\tau_{FH}})^{\rho}, (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{\rho}{\gamma}} (\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}} (\frac{\tau_{HF}}{\tau_{FF}})^{\rho} \right). \tag{10}$$

As for Assumption 4, the corresponding Jacobian determinant can be simplified as:

$$\begin{vmatrix} (\frac{1}{f_{HH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{HH}})^{\gamma}, & (\frac{1}{f_{HF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{HF}})^{\gamma} \\ (\frac{1}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{FH}})^{\gamma}, & (\frac{1}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}} (\frac{1}{\tau_{FF}})^{\gamma} \end{vmatrix} > 0. \tag{11}$$

Both assumptions can be simplified as:

$$(\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HH}}{\tau_{FH}})^{\gamma} < (\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{HF}}{\tau_{FF}})^{\gamma}.\tag{12}$$ [20]

Under condition (12), we can establish the uniqueness and existence of the decentralized market equilibrium. To condense the report of solutions, we use $w_i$ and $w_j$ instead of normalizing and ignoring $w_H$ and summarize the results in Table 2.

Table 2: Explicit Solution with CES preferences and demand

| | *dmkt* | *cmkt, opt* |
|---|---|---|
| $\varphi^*_{ii}$ | $\left\{\frac{\rho f_{ii}\left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma}-1\right]}{[(1-\rho)\gamma-\rho]\mathcal{F}_i\left[(\frac{f_{ij}f_{ji}}{f_{jj}f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ij}\tau_{ji}}{\tau_{jj}\tau_{ii}})^{\gamma}-\frac{\mathcal{F}_jw_j}{\mathcal{F}_iw_i}(\frac{f_{ji}w_j}{f_{ii}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^{\gamma}\right]}\right\}^{\frac{1}{\gamma}}$ | $\left\{\frac{\rho f_{ii}}{[(1-\rho)\gamma-\rho]\mathcal{F}_i}\left[1+(\frac{\tau_{ii}}{\tau_{ij}})^{\gamma}(\frac{f_{ii}}{f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{L_j}{L_i})^{\frac{(1-\rho)\gamma}{\rho}}\right]\right\}^{\frac{1}{\gamma}}$ |
| $\varphi^*_{ij}$ | $\left\{\frac{\rho f_{ij}\left[1-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]}{[(1-\rho)\gamma-\rho]\mathcal{F}_i\left[\frac{\mathcal{F}_jw_j}{\mathcal{F}_iw_i}(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}\right]}\right\}^{\frac{1}{\gamma}}$ | $\left\{\frac{\rho f_{ij}}{[(1-\rho)\gamma-\rho]\mathcal{F}_i}\left[1+(\frac{\tau_{ij}}{\tau_{ii}})^{\gamma}(\frac{f_{ij}}{f_{ii}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{L_i}{L_j})^{\frac{(1-\rho)\gamma}{\rho}}\right]\right\}^{\frac{1}{\gamma}}$ |
| $q_{ii}$ | $\frac{\rho}{1-\rho}\frac{f_{ii}}{L_i\tau_{ii}}\left(\frac{1}{\varphi^*_{ii}}\right)^{\frac{\rho}{1-\rho}}\varphi^{\frac{1}{1-\rho}}$ | |
| $q_{ij}$ | $\frac{\rho}{1-\rho}\frac{f_{ij}}{L_j\tau_{ij}}\left(\frac{1}{\varphi^*_{ij}}\right)^{\frac{\rho}{1-\rho}}\varphi^{\frac{1}{1-\rho}}$ | |
| $M_i$ | $\frac{L_i\rho}{\mathcal{F}_i\gamma}$ | |
| $\frac{w_j}{w_i}$ | $\frac{L_i}{L_j}\frac{\left[(\frac{\mathcal{F}_jw_j}{\mathcal{F}_iw_i})(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^{\gamma}\right]}{\left[(\frac{\mathcal{F}_iw_i}{\mathcal{F}_jw_j})(\frac{f_{ii}w_i}{f_{ji}w_j})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{ii}w_i}{\tau_{ji}w_j})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ij}\tau_{ji}})^{\gamma}\right]}$ | |

We use $\varphi^*_{ij}$ 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 parameters of the CES demand and Pareto technology. Second, $\frac{f_{ij}}{\mathcal{F}_i}$, 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. Condition (11) ensures this term is strictly positive. Finally, $\frac{\mathcal{F}_jw_j}{\mathcal{F}_iw_i}(\frac{f_{jj}w_j}{f_{ij}w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^{\gamma}-(\frac{f_{jj}f_{ii}}{f_{ji}f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^{\gamma}$, measuring the relative advantage of domestic sales to imports in country $j$ adjusted for global trade frictions, matters. With condition (10), we restrict the admissible range of endogenous $w_F$, and one can verify it guarantees the positivity of this term and of the equilibrium wage rate $\frac{w_j}{w_i}$.

Given the equivalence between the centralized market equilibrium and the utilitarian global 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

[20] As one might realize in our two samples with the CES and CARA demand under symmetric Pareto-distributed distribution, the explicit expressions of Assumption 3 and 4 are the same. However, such a conclusion does not hold in the general case, as we show in Proposition 1, since Assumption 3 guarantees the existence and Assumption 4 guarantees the uniqueness.

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*

$$(L_i)^{\frac{(1-\rho)\gamma}{\rho}} \left[ (\frac{1}{\mathcal{F}_i w_i})(\frac{1}{f_{ii} w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{ii} w_i})^{\gamma} - (\frac{1}{\mathcal{F}_j w_j})(\frac{1}{f_{ji} w_j})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{ji} w_j})^{\gamma} \right]$$
$$> (L_j)^{\frac{(1-\rho)\gamma}{\rho}} \left[ (\frac{1}{\mathcal{F}_j w_j})(\frac{1}{f_{jj} w_j})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{jj} w_j})^{\gamma} - (\frac{1}{\mathcal{F}_i w_i})(\frac{1}{f_{ij} w_i})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{ij} w_i})^{\gamma} \right],$$

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

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

*Proof.* See Appendix A.8. □

Lemma 3 is an application of Proposition 7. As discussed in Section 4, under the CES demand, the equivalence between the centralized market equilibrium and the utilitarian global social optimum eliminates the rotation effect, but the shift effect persists when competition intensities vary across destinations. With the explicit measures of competition intensities on hand, one can quantify the shift effect with data. In the presence of only the shift effect, distortions in sales to different destinations behave in opposite ways. Specifically, the utilitarian global planner with a CES demand will systematically reduce all sales to one destination and increase all sales to the other.

# 6 Example: CARA Preferences

In the previous section, we have established a set of results based on CES preferences 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 than in the general case even with specific variable-elasticity-of-substitution (VES) preferences, which incorporate both shift and rotation effects. Specifically, for an illustration we rely on the Constant Absolute Risk Aversion (CARA) preferences with parameter $a > 0$, an absence of fixed market-access costs ($f_{ij} = 0\ \forall i, j = H, F$), 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 (De Loecker et al., 2016) and Marshall (1920)'s Second Law of Demand is satisfied.

## 6.1 Solutions under Different Equilibria

We relegate most of the analytical details to Appendix A.10. One can verify that CARA preferences satisfy Assumptions 1, 2, and 5. We now derive the explicit expressions for Assumptions 3 and 4 and show how these assumptions guarantee the existence and uniqueness of the decentralized market equilibrium.

Assumption 3 requires that $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$, where $\mathcal{C}_{ij} = (\delta_{ij}w_i)^{-1}$ is the measure of aggregate cost. With the CARA demand structure and the Pareto productivity distribution, we can obtain the expression of $\mathcal{C}_{ij}$ and the admissible range for endogenous wage $w_F$:

$$w_F \in (\underline{w_F}, \overline{w_F}) = \left(\frac{\mathcal{C}_{HH}}{\mathcal{C}_{FH}}, \frac{\mathcal{C}_{HF}}{\mathcal{C}_{FF}}\right) = \left( (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HH}}{\tau_{FH}})^{\frac{\gamma}{\gamma+1}}, (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HF}}{\tau_{FF}})^{\frac{\gamma}{\gamma+1}} \right) \tag{13}$$

Assumption 3 becomes $\frac{\tau_{HH}}{\tau_{FH}} < \frac{\tau_{HF}}{\tau_{FF}}$ and guarantees the existence of the range for $w_F$.

As for Assumption 4, the positivity of the Jacobian determinant requirement can be simplified to:

$$\begin{vmatrix} \frac{L_H}{\tau_{HH}^{\gamma}}, & \frac{L_F}{\tau_{HF}^{\gamma}} \\ \frac{L_H}{\tau_{FH}^{\gamma}}, & \frac{L_F}{\tau_{FF}^{\gamma}} \end{vmatrix} > 0. \tag{14}$$

Both two expressions can be simplified to

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

After establishing the conditions for existence and uniqueness, 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 $\mathcal{W}$ is the Lambert function (Corless et al., 1996), which satisfies $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$.

Starting with the decentralized market equilibrium, we first discuss how our assumptions guarantee the uniqueness and existence and then illustrate the properties of equilibrium

Table 3: Explicit Solution of CARA Demand

| | *dmkt* | *cmkt* | *opt* |
|---|---|---|---|
| $\varphi_{ii}^*$ | $\left\{\frac{\gamma\kappa_1 L_i\tau_{ii}\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma-1\right]}{a\mathcal{F}_i\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma-(\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{ji}w_j}{\tau_{ii}w_i})^\gamma\right]}\right\}^{\frac{1}{\gamma+1}}$ | $\left\{\frac{\gamma\kappa_1 L_i\tau_{ii}\left[1+\frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{a\mathcal{F}_i}\right\}^{\frac{1}{\gamma+1}}$ | $\left\{\frac{L_i\tau_{ii}\left[1+\frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{a(\gamma+1)^2\mathcal{F}_i}\right\}^{\frac{1}{\gamma+1}}$ |
| $\varphi_{ij}^*$ | $\left\{\frac{\gamma\kappa_1 L_j\tau_{ij}\left[1-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma\right]}{a\mathcal{F}_i\left[(\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma\right]}\right\}^{\frac{1}{\gamma+1}}$ | $\left\{\frac{\gamma\kappa_1 L_j\tau_{ij}\left[1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^\gamma\right]}{a\mathcal{F}_i}\right\}^{\frac{1}{\gamma+1}}$ | $\left\{\frac{L_j\tau_{ij}\left[1+\frac{L_i}{L_j}(\frac{\tau_{ij}}{\tau_{ii}})^\gamma\right]}{a(\gamma+1)^2\mathcal{F}_i}\right\}^{\frac{1}{\gamma+1}}$ |
| $q_{ii}$ | $\frac{1}{a}\left[1-\mathcal{W}(e\frac{\varphi_{ii}^*}{\varphi})\right]$ | | $\frac{1}{a}\ln(\frac{\varphi}{\varphi_{ii}^*})$ |
| $q_{ij}$ | $\frac{1}{a}\left[1-\mathcal{W}(e\frac{\varphi_{ij}^*}{\varphi})\right]$ | | $\frac{1}{a}\ln(\frac{\varphi}{\varphi_{ij}^*})$ |
| $M_i$ | $\frac{L_i}{\mathcal{F}_i(\gamma+1)}$ | | |
| $\frac{w_j}{w_i}$ | $\frac{L_i}{L_j}\frac{(\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}{(\frac{\mathcal{F}_i w_i}{\mathcal{F}_j w_j})(\frac{\tau_{ii}w_i}{\tau_{ji}w_j})^\gamma-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}$ | | |

outcomes by using $\varphi_{ij}^*$ as an example. First, $\gamma\kappa_1$, which only depends on the Pareto shape parameter, as well as the CARA parameter $a$ matter. Second, $\frac{L_j\tau_{ij}}{\mathcal{F}_i}$, 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. Condition (14) ensures this term is positive. Finally, $(\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{jj}w_j}{\tau_{ij}w_i})^\gamma-(\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma$, measuring the relative advantage of domestic sales to imports in country $j$ adjusted for global trade frictions, enter. With condition (13), we guarantee the positivity of this term and of the equilibrium wage rate $\frac{w_j}{w_i}$. 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 trade-off 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 utilitarian global 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 $\frac{L_j\tau_{ij}}{\mathcal{F}_i}$, and the preference parameter $\frac{1}{a}$. What differs is that the utilitarian global 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, the utilitarian global planner assigns markups to firms differently, resulting in different quantity functions 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.[21] In the following sections, we demonstrate how these differences between the solutions correspond to our general results.

## 6.2 Comparisons of Cutoffs

With CARA preferences and Pareto productivities, we are able to derive an explicit expression for the destination-specific demand shifters $\delta_j^{dmkt}$. 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*

$$L_i\left[(\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ii} w_i})^{\gamma} - (\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{ji} w_j})^{\gamma}\right] > L_j\left[(\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{jj} w_j})^{\gamma} - (\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ij} w_i})^{\gamma}\right], \quad (15)$$

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

*Proof.* See Appendix A.11. □

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.

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

*Proof.* See Appendix A.12. □

Lemma 5 exemplifies Proposition 6 and summarizes the rotation effect discussed in Section 4.2. Due to the increasing private and social markups associated with CARA

[21] Note that the efficiency of entry is not guaranteed in the general case, as we discussed above. However, efficient entry is also reported in Behrens et al. (2020) and Bagwell and Lee (2023).

preferences, the selection effect under the centralized equilibrium is weaker than that in the social optimum.[22]

**Proposition 8.** *$\forall i, j = H, F$ and $i \neq j$, if condition (15) 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 Lemmata 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 cost variables. 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 productivity. Since the mass of entrants is efficient in the market equilibrium, distortions in the mass of producing firms are determined by the cutoff distortions.

## 6.3 Comparisons of 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 $\varphi < \tilde{\varphi}^{cmkt}_{ij}$. $\forall i \neq j$, $\tilde{\varphi}^{cmkt}_{ij} = \frac{\tau_{ij}}{\tau_{ii}} \tilde{\varphi}^{cmkt}_{ii}$.*

*Proof.* see Appendix A.13. □

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 $\varphi > \tilde{\varphi}^{cmkt}_{ij}$ underproduce, while the less productive firms with $\varphi < \tilde{\varphi}^{cmkt}_{ij}$ 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 (15) 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, $q^{dmkt}_{\ell i}(\varphi) < q^{opt}_{\ell i}(\varphi)$ for $\varphi > \tilde{\varphi}^{dmkt}_{\ell i}$, and $q^{dmkt}_{\ell i}(\varphi) > q^{opt}_{\ell i}(\varphi)$ for $\varphi < \tilde{\varphi}^{dmkt}_{\ell i}$.*

[22] Since we assume all fixed costs are zero for tractability, the CARA example does not show the impact of fixed costs per capita, characterized by $\underline{j}^i$ and $\overline{j}^i$ in Proposition 6.

- *If $q_{\ell i}^{dmkt}(\varphi)$ and $q_{\ell i}^{opt}(\varphi)$ do not intersect, then $q_{\ell i}^{opt}(\varphi) > q_{\ell i}^{dmkt}(\varphi)$ for all $\varphi$.*

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 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.

In summary, most of the results are general and do not permit a uniform conclusion regarding the differences between the market equilibrium and global social optimum. More specifically, the misallocation patterns depend not only on the demand-side elasticities but also on the fundamentals of countries, which represents a critical distinction between open and closed economies. In our working paper version (Egger and Huang, 2023), we further provide a quantitative example under CARA preferences and discuss the relative strength of two types of effects.

We aim to provide a baseline for comparison and highlight the inefficiency discussion in open economies. Our paper contributes to the theory of future policy design, particularly in the global context. We emphasize the influence of country heterogeneity and the VES demand. In such scenarios, where laissez-faire equilibria lead to inefficiencies and the traditional primary method proves inadequate, our findings provide crucial insights. However, we acknowledge that the concept of a global social planner, as proposed in our paper and discussed in the literature, is subject to debate. In the following section, we will explore this issue further by examining a more generalized setting.

# 7 Discussion: The utilitarian global Social Planner

As we discussed in Section 3, we consider a utilitarian global social planner who weights two countries equally to maximize the aggregate global welfare as a reasonable benchmark. Such a planner puts equal utilitarian weight to any customer at home and abroad. However, we understand that global planners might be endowed with other intentions. We therefore reflect on a more general setup here. To be specific, we consider a utilitarian setup but allow the planner to assign positive Pareto weights, $\omega_{ij}$, to the origin-destination-specific welfare gains, $M_i L_j \int_{\varphi_{ij}^*}^{+\infty} u(q_{ij}(\varphi)) dG_i(\varphi)$ associated with reallocation. Different values of Pareto weights generate the Pareto frontier of global welfare. We then consider the question of whether the market allocation is on the Pareto frontier from a utilitarian global's perspective.

In Section 4, we show that the distortions between the decentralized market equilibrium and utilitarian global social optimum can be decomposed into shift and rotation effects. The shift effect is driven by the difference between destination-specific marginal utility of income, which is a constant, while the rotation effect is driven by the demand-side elasticities and is distributional. One can now observe that the planner, when given access to market-specific constant weights, can only shift the quantity schedules instead of rotating them. Hence, market-specific Pareto weights can be viewed as adjustments to competition intensities $\delta_j^{dmkt}$. Given that the CES demand is the only case without a rotation effect, we can therefore obtain the following proposition about the relationship between the decentralized market equilibrium and the social optimum.

**Proposition 10.** *The necessary and sufficient condition for the market allocation to be on the Pareto frontier is CES demand.*

With Proposition 4, we can determine the origin-destination-specific weight $\omega_{ij} = \delta_i^{cmkt}/\delta_j^{dmkt}$ such that the market allocation lies on the Pareto frontier. Therefore, Proposition 10 shows that CES preferences are necessary and – only when using market-specific Pareto weights with asymmetric markets – also sufficient for the Second Welfare Theorem to be preserved. However, for general VES preferences, the market allocation never resides on the Pareto frontier, due to the rotation effect discussed in Section 4.

# 8 Conclusions

We introduce a general equilibrium model of monopolistic competition and nonparametrically distributed heterogeneous firms with two asymmetric countries and general additive preferences. We systematically characterize the model and introduce a novel measure of market-pair-specific aggregate cost to establish sufficient conditions for the existence and uniqueness of the market equilibrium. This measure captures all the (direct and indirect) impacts of exogenous parameters, thus allowing the model to feature general country asymmetry. Our results extend to a large class of demand structures and general productivity distributions.

Furthermore, we provide a detailed comparison between the decentralized market equilibrium and social optimum, the latter from the perspective of a utilitarian global social planner. This may serve as a baseline in the misallocation discussion in open-economy models. We show that distortions can be characterized as shift and rotation effects emanating from the country asymmetry and VES demand, respectively. The combination of the two effects does not permit unambiguous conclusions, because the shift effects counteract the rotation effects in specific markets. We further provide two examples to illustrate our sufficient conditions and misallocation discussions.

One key conclusion we can draw from our analysis is the critical role of both country asymmetry and variable demand elasticity in open-economy models, particularly when devising global policies. Potentially fruitful future research directions include quantitative analyses based on the model, comparisons of market equilibrium with the social optimum from a national social planner's perspective, and the exploration of optimal policies from the standpoints of global versus national social planners.

# References


Amann, H., J. Escher, and G. Brookfield (2005). *Analysis.* Springer.

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

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.

Bagwell, K. and S. H. Lee (2023). Monopolistic competition and efficiency under firm heterogeneity and non-additive preferences. *American Economic Journal: Microeconomics*.

Baqaee, D. R., E. Farhi, and K. Sangani (2023). The Darwinian Returns to Scale. *The Review of Economic Studies*, rdad061.

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 (2017). Monopolistic competition when income matters. *The Economic Journal 127*(603), 1217–1243.

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

Burk, A. (1938). A reformulation of certain aspects of welfare economics. *The Quarterly Journal of Economics 52*(2), 310–334.

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.

Costinot, A., A. Rodríguez-Clare, and I. Werning (2020). Micro to macro: Optimal trade policy with firm heterogeneity. *Econometrica 88*(6), 2739–2776.

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

Demidova, S. (2017). Trade policies, firm heterogeneity, and variable markups. *Journal of International Economics 108*, 260–273.

Demidova, S. and A. Rodríguez-Clare (2009). Trade policy under firm-level heterogeneity in a small economy. *Journal of International Economics 78*(1), 100–112.

Demidova, S. and A. Rodriguez-Clare (2013). The simple analytics of the melitz model in a small economy. *Journal of International Economics 90*(2), 266–272.

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.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Egger, P. and R. Huang (2023). *Market Versus Optimum Diversity in Open Economies: Theory and Quantitative Evidence*. Centre for Economic Policy Research.

Felbermayr, G., B. Jung, and M. Larch (2013). Optimal tariffs, retaliation, and the welfare loss from tariff wars in the melitz model. *Journal of International Economics 89*(1), 13–25.

Gale, D. and H. Nikaido (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen 159*(2), 81–93.

Haaland, J. I. and A. J. Venables (2016). Optimal trade policy with monopolistic competition and heterogeneous firms. *Journal of International Economics 102*, 85–95.

Hayek, F. A. (1944). *The Road to Serfdom*. University of Chicago Press.

Kokovin, S., P. Molchanov, and I. Bykadorov (2022). Increasing returns, monopolistic competition, and international trade: Revisiting gains from trade. *Journal of International Economics 137*, 103595.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review 70*(5), 950–959.

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics 9*(4), 469–479.

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

Marshall, A. (1920). *Principles of Economics* (8th ed.). London: Macmillan.

Matsuyama, K. and P. Ushchev (2017). Beyond ces: three alternative classes of flexible homothetic demand systems. *Global Poverty Research Lab Working Paper* (17-109).

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.

Melitz, M. J. and S. J. Redding (2015). New trade models, new welfare implications. *American Economic Review 105*(3), 1105–1146.

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.

Nozick, R. (1974). *Anarchy, state, and utopia.* John Wiley & Sons.

Parenti, M., P. Ushchev, and J.-F. Thisse (2017). Toward a theory of monopolistic competition. *Journal of Economic Theory 167*, 86–115.

Rawls, J. (1958). Justice as fairness. *The Philosophical Review 67*(2), 164–194.

Rawls, J. (1974). Some reasons for the maximin criterion. *The American Economic Review 64*(2), 141–146.

Sen, A. (1997). *On economic inequality.* Oxford University Press.

Shelburne, R. C. (2006). A utilitarian welfare analysis of trade liberalization. *UNECE Discussion Paper Series No. 2006.4*.

Welch, C. (1987). Utilitarianism. In M. M. John Eatwell and P. Newman (Eds.), *The New Palgrave: A Dictionary of Economics*, pp. 1–10. Macmillan Press,.

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 Appendix

## A.1 Proof of Lemmata 1 & 2

*Proof.* We first show the properties of the optimality conditions in general equilibrium and then apply them with the counterfactual equilibrium. For brevity, we refer to $\delta_j^{dmkt}$ by $\delta_j$ without ambiguity. Recall that 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 exceed unity 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 $\overline{q} = +\infty$.[23]

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

$$[1 - r_u(q_{ij})]\, u'(q_{ij}) = \frac{\delta_j \tau_{ij} w_i}{\varphi}. \tag{16}$$

Taking the derivative of the LHS w.r.t. $q_{ij}$ obtains $\frac{\partial\{[1-r_u(q_{ij})]\cdot u'(q_{ij})\}}{\partial q_{ij}} = u''(q_{ij})\cdot[2 - r_{u'}(q_{ij})] < 0$,[24] where $r_{u'}(q_{ij}) \equiv -\frac{q_{ij}\cdot u'''(q_{ij})}{u''(q_{ij})}$. Recall that we have $\lim_{q_{ij}\to\overline{q}} [1 - r_u(q_{ij})]\, u'(q_{ij}) = 0$ and $\lim_{q_{ij}\to+\infty} u'(q_{ij}) = 0$.

The LHS of (16) 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

$$\lim_{\frac{\tau_{ij}}{\varphi}\to+\infty} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i\to+\infty} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = 0, \tag{17}$$

$$\lim_{\frac{\tau_{ij}}{\varphi}\to 0} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i\to 0} q_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \overline{q}. \tag{18}$$

When (16) holds, we can write a firm's profit divided by the wage in the decentralized

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

[24]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

$$
\frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \delta_j w_i} < 0 \text{ and } \frac{\partial \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi})}{\partial \frac{\tau_{ij}}{\varphi}} < 0. \tag{19}
$$

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 (17), we obtain

$$
\lim_{\frac{\tau_{ij}}{\varphi} \to +\infty} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \lim_{\delta_j w_i \to +\infty} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = -f_{ij}. \tag{20}
$$

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 (18), we obtain

$$
\lim_{\frac{\tau_{ij}}{\varphi} \to 0} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = \frac{L_j \overline{B}}{\delta_j w_i} - f_{ij} \text{ and } \lim_{\delta_j w_i \to 0} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = +\infty. \tag{21}
$$

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

Combine (19), (20), and (21), 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 (19) 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

$$\lim_{\delta_j w_i \to 0} \varphi^*_{ij}(\delta_j w_i) = 0, \text{ and } \lim_{\delta_j w_i \to \frac{L_j \overline{B}}{f_{ij}}} \varphi^*_{ij}(\delta_j w_i) = +\infty. \tag{22}$$

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 we obtain using (20), (21), and (22):

$$\lim_{\delta_j w_i \to 0} \tilde{\Pi}_{ij}(\delta_j w_i) = +\infty \text{ and } \lim_{\delta_j w_i \to \frac{L_j \overline{B}}{f_{ij}}} \tilde{\Pi}_{ij}(\delta_j w_i) = 0. \tag{23}$$

Now we prove Lemma 2, the existence and uniqueness of the counterfactual equilibrium $\left\{\delta_{ij} w_i, \varphi^*_{ij}, q_{ij}(\varphi)\right\}$, which satisfies the following equilibrium conditions:

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

Since $\frac{d\tilde{\Pi}_{ij}(\delta_j w_i)}{d(\delta_j w_i)} < 0$ and (23) holds, there exists a unique $\delta_{ij} w_i$ s.t. $\tilde{\Pi}_{ij}(\delta_{ij} w_i) = \mathcal{F}_i$, which implies a unique solution for $\varphi^*_{ij}(\delta_{ij} w_i)$ and $q_{ij}(\delta_{ij} w_i, \frac{\tau_{ij}}{\varphi})$. We can show the relation between the solution $\delta_{ij} w_i$ and $\tau_{ij}$, $f_{ij}$, and $\mathcal{F}_i$. Since $\frac{\partial\tilde{\Pi}_{ij}}{\partial\tau_{ij}} < 0$, $\frac{\partial\tilde{\Pi}_{ij}}{\partial f_{ij}} < 0$, and $\frac{\partial\tilde{\Pi}_{ij}}{\partial L_j} > 0$, the equilibrium value of $\delta_{ij} w_i$ must decrease with $\tau_{ij}$ and $f_{ij}$ and increase with $L_j$ s.t. $\tilde{\Pi}_{ij}(\delta_{ij} w_i) = \mathcal{F}_i$. Besides, with an increase of $\mathcal{F}_i$ and $\frac{\partial \Pi_{ij}}{\partial \delta_{ij} w_i} < 0$, $\delta_{ij} w_i$ must decrease to satisfy the counterfactual ZEPC.

Last, we consider the case of zero fixed costs, $f_{ij} = 0$, $\forall i, j = H, F$. We then directly obtain the cutoff quantity $q_{ij}(\varphi^*_{ij}) = 0$. Since now $\overline{B} < +\infty$, we can obtain $\lim_{\delta_j w_i \to +\infty} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = 0$ and $\lim_{\delta_j w_i \to 0} \tilde{\pi}_{ij}(\delta_j w_i, \frac{\tau_{ij}}{\varphi}) = +\infty$ and, thus, all values of $\delta_j w_i$ guarantee non-negative profits. From the FOCs, $[1 - r_u(0)]\, u'(0) = \frac{\delta_i \tau_{ij} w_i}{\varphi^*_{ij}}$, we can obtain the solution of cutoff productivity $\varphi^*_{ij}(\delta_j w_i)$ that increases with $\delta_j w_i$. The rest of the analysis is similar to the case with fixed costs. □

## A.2 Proof of Proposition 1

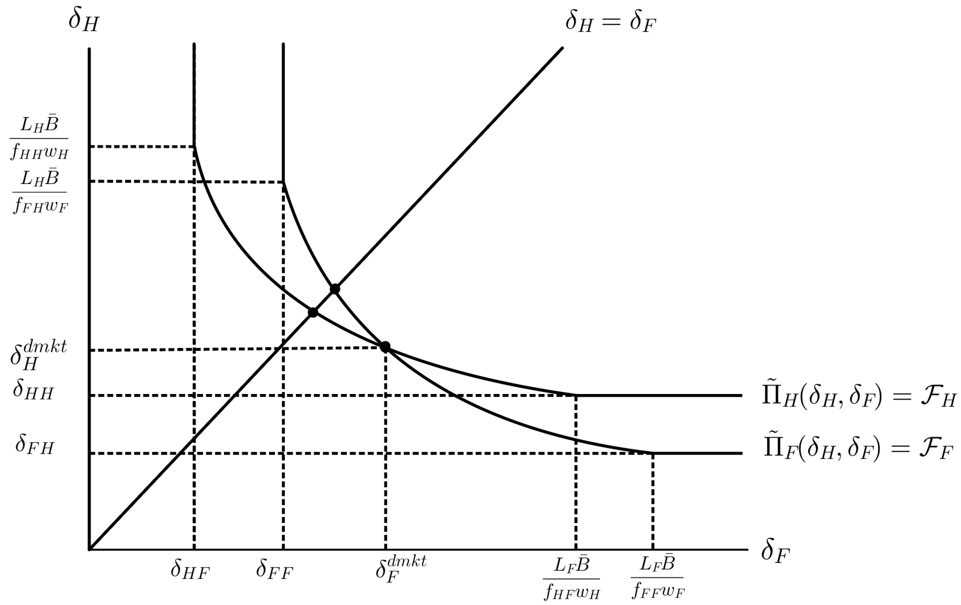
*Proof.* **Step 1. The existence and uniqueness of $\delta_j(w_F)$ conditional on $w_F \in (\underline{w}_F, \overline{w}_F)$.** In this step, we consider the partial equilibrium conditional on endogenous wage $w_F$. That is, all endogenous variables are pinned down conditional on a given wage $w_F$. Recall that $\forall w_F \in (\underline{w}_F, \overline{w}_F)$, we have $\delta_{FF}(w_F) > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}(w_F)$. Although we set $w_H = 1$ as the numeraire, we keep using $w_H$ for consistency of notation. For simplicity of notation, we express the expression of the endogenous variable without the conditional value of the wage (e.g., $\tilde{\Pi}_{ij}(\delta_j)$). However, we return the full notation (e.g., $\tilde{\Pi}_{ij}(\delta_j, w_i)$) when discussing the general equilibrium.

We know that $\forall i, j = H, F$, $\tilde{\Pi}_{ij}(\delta_j)$ is first-order continuously differentiable for $\delta_j \in (0, \frac{L_j \overline{B}}{f_{ij} w_i})$. We further define $\tilde{\Pi}_i(\delta_H, \delta_F) \equiv \tilde{\Pi}_{iH}(\delta_H) + \tilde{\Pi}_{iF}(\delta_F)$ and have $\tilde{\Pi}_F(\frac{L_H \overline{B}}{f_{FH} w_F}, \delta_{FF}) = \mathcal{F}_F$.

***Case i)*** $\tilde{\Pi}_H(\frac{L_H \overline{B}}{f_{FH} w_F}, \delta_{FF}) \geq \mathcal{F}_H$ and $\delta_{FF} < \frac{L_F \overline{B}}{f_{HF} w_H}$.

We can obtain a solution, where $\delta^*_F = \delta^*_{FF}$ and $\delta^*_H \geq \frac{L_H \overline{B}}{f_{FH} w_F}$. Now we prove the uniqueness of the solution in this case. Given that $\delta_{FF} < \frac{L_F \overline{B}}{f_{HF} w_H}$, we consider $\delta_F \in (\delta_{FF}, \min\{\frac{L_F \overline{B}}{f_{HF} w_H}, \frac{L_F \overline{B}}{f_{FF} w_F}\})$. Since within this interval, $\tilde{\Pi}_{FF}(\delta_F)$ and $\tilde{\Pi}_{FH}(\delta_H)$ are strictly decreasing in $\delta_F$ and $\delta_H$, respectively, there exists a unique global implicit function $\delta^F_H(\delta_F)$ such that $\tilde{\Pi}_F(\delta^F_H(\delta_F), \delta_F) = \mathcal{F}_F$. Besides, since both $\tilde{\Pi}_{FF}(\delta_F)$ and $\tilde{\Pi}_{FH}(\delta_H)$ are first-order continuously differentiable, $\delta^F_H(\delta_F)$ is continuously differentiable in the open interval, and

Figure 2: Zero-Expected-Profit Conditions for a Given $w_F \in (\underline{w_F}, \overline{w_F})$




the derivative is given by

$$\delta_H^{F\prime}(\delta_F) = -\tilde{\Pi}'_{FF}(\delta_F)/\tilde{\Pi}'_{FH}(\delta_H^F(\delta_F)) < 0.$$[25]

Next, evaluate $\tilde{\Pi}_H(\delta_H, \delta_F)$ at $\delta_H = \delta_H^F(\delta_F)$. Given that $\delta_H^* \geq \frac{L_H\overline{B}}{f_{FH}w_F}$ implies $\frac{L_H\overline{B}}{f_{HH}w_H} > \frac{L_H\overline{B}}{f_{FH}w_F}$, and that $\forall \delta_F \in (\delta_{FF}, \min\{\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F}\})$ we have $\delta_H^F(\delta_F) \leq \frac{L_H\overline{B}}{f_{FH}w_F}$, $\tilde{\Pi}_{HH}(\delta_H^F(\delta_F))$ and $\tilde{\Pi}_{HF}(\delta_F)$ are both differentiable in $\delta_H^F(\delta_F)$ and $\delta_F$, respectively, within the interval. Therefore, with the positivity of the Jacobian (Assumption 4), the differentiation $\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F)$ w.r.t $\delta_F$ implies

$$\frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F)}{\mathrm{d}\delta_F} = \tilde{\Pi}'_{HH}(\delta_H^F(\delta_F)) \cdot \left(-\frac{\tilde{\Pi}'_{FF}(\delta_F)}{\tilde{\Pi}'_{FH}(\delta_H^F(\delta_F))}\right) + \tilde{\Pi}'_{HF}(\delta_F) = -\frac{|J|_{\delta_H = \delta_H^F(\delta_F)}}{\tilde{\Pi}'_{FH}(\delta_H^F(\delta_F))} > 0,$$

where $|J|_{\delta_H = \delta_H^F(\delta_F)} = \begin{vmatrix} \tilde{\Pi}'_{HH}(\delta_H) & \tilde{\Pi}'_{HF}(\delta_F) \\ \tilde{\Pi}'_{FH}(\delta_H) & \tilde{\Pi}'_{FF}(\delta_F) \end{vmatrix}_{\delta_H = \delta_H^F(\delta_F)}$ is the Jacobian evaluated at $\delta_H = \delta_H^F(\delta_F)$. Therefore,

$$\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F) > \mathcal{F}_H, \quad \forall \delta_F \in (\delta_{FF}, \min\{\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F}\}).$$

[25]We apply a similar method as in Theorem 8(ii) of Gale and Nikaido (1965). Another method is to employ the global inverse function theorem to prove existence, uniqueness, and differentiability of the global implicit function in our cases. For details, see Theorems III.5.7 and IV.1.8 in Amann et al. (2005).

Now consider $\delta_F \geq \min\{\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F}\}$. If $\frac{L_F\overline{B}}{f_{HF}w_H} < \frac{L_F\overline{B}}{f_{FF}w_F}$, $\forall \delta_F \in \Big[\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F}\Big)$, $\delta_H^F(\delta_F)$ is continuous and strictly decreasing and, thus,

$$\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F) = \tilde{\Pi}_H(\delta_H^F(\delta_F), \frac{L_F\overline{B}}{f_{HF}w_H}) > \tilde{\Pi}_H(\delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}), \frac{L_F\overline{B}}{f_{HF}w_H}) > \mathcal{F}_H.$$

Besides, given that $\delta_{HH} > \delta_{FH}$, there is no solution for $\delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$. On the other hand, if $\frac{L_F\overline{B}}{f_{HF}w_H} \geq \frac{L_F\overline{B}}{f_{FF}w_F}$, $\forall \delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$, we know $\tilde{\Pi}_F(\delta_{FH}, \delta_F) = \mathcal{F}_F$. Since $\delta_{FH} < \delta_{HH}$ and $\tilde{\Pi}_{HF}(\delta_F) \geq 0$,

$$\tilde{\Pi}_H(\delta_{FH}, \delta_F) = \tilde{\Pi}_{HH}(\delta_{FH}) + \tilde{\Pi}_{HF}(\delta_F) > \tilde{\Pi}_{HH}(\delta_{HH}) = \mathcal{F}_H.$$

Therefore, there is no solution $\forall \delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$.

***Case ii)*** $\tilde{\Pi}_H(\frac{L_H\overline{B}}{f_{FH}w_F}, \delta_{FF}) \geq \mathcal{F}_H$ and $\delta_{FF} \geq \frac{L_F\overline{B}}{f_{HF}w_F}$.

In this case, there exists a solution, where $\delta_F^* = \delta_{FF}$ and $\delta_H^* = \delta_{HH} \geq \frac{L_H\overline{B}}{f_{FH}w_F}$. Besides, $\forall \delta_F \in (\delta_{FF}, +\infty)$, $\tilde{\Pi}_H(\delta_H, \delta_F) = \mathcal{F}_H$ holds only when $\delta_H = \delta_{HH}$. However, $\tilde{\Pi}_F(\delta_{HH}, \delta_F) < \tilde{\Pi}_F(\delta_{HH}, \delta_{FF}) = \mathcal{F}_F$. Therefore, the solution is unique.

***Case iii)*** $\tilde{\Pi}_H(\frac{L_H\overline{B}}{f_{FH}w_F}, \delta_{FF}) < \mathcal{F}_H, \frac{L_H\overline{B}}{f_{FH}w_F} \leq \frac{L_H\overline{B}}{f_{HH}w_H}$, and $\tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}) < \mathcal{F}_F$.

$\forall \delta_F \in (\delta_{FF}, \frac{L_H\overline{B}}{f_{FF}w_F})$, there exists a differentiable function $\delta_H^F(\delta_F)$ s.t. $\tilde{\Pi}_F(\delta_H^F(\delta_F), \delta_F) = \mathcal{F}_F$. We can further refine $\delta_H^F(\delta_F) = \delta_{FH}$ when $\delta_F \geq \frac{L_H\overline{B}}{f_{FF}w_F}$.

Given that $\frac{L_F\overline{B}}{f_{HF}w_H} \leq \frac{L_F\overline{B}}{f_{FF}w_F}$, since $\tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_F}) < \mathcal{F}_F = \tilde{\Pi}_F(\delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}), \frac{L_F\overline{B}}{f_{HF}w_H})$, we obtain

$$\delta_{HH} > \delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}) \text{ and } \tilde{\Pi}_H(\delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}), \frac{L_F\overline{B}}{f_{HF}w_H}) > \tilde{\Pi}_H(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}) = \mathcal{F}_H.$$

$\forall \delta_F \in (\delta_{FF}, \frac{L_H\overline{B}}{f_{HF}w_H})$, we have $\frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F)}{\mathrm{d}\delta_F} > 0$. Therefore, within the interval, there exists a unique $\delta_F^*$ s.t. $\delta_H^* = \delta_H^F(\delta_F)$ and $\tilde{\Pi}_H(\delta_H^*, \delta_F^*) = \mathcal{F}_H$. Besides, since $\forall \delta_F \in [\frac{L_H\overline{B}}{f_{HF}w_H}, \frac{L_H\overline{B}}{f_{FF}w_F})$, $\delta_H^{F\prime}(\delta_F) < 0$, we obtain

$$\delta_H^F(\delta_F) < \delta_H^F(\frac{L_H\overline{B}}{f_{HF}w_H}) < \delta_{HH}.$$

$\forall \delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$, $\delta_H^F(\delta_F) = \delta_{FH} < \delta_{HH}$, but $\tilde{\Pi}_H(\delta_H, \delta_F) = \mathcal{F}_H$ only holds, when $\delta_H = \delta_{HH}$. Therefore, the solution is unique.

If $\frac{L_F\overline{B}}{f_{HF}w_H} > \frac{L_F\overline{B}}{f_{FF}w_F}$, we obtain

$$\tilde{\Pi}_H(\delta_H^F(\frac{L_F\overline{B}}{f_{FF}w_F}), \frac{L_F\overline{B}}{f_{FF}}) = \tilde{\Pi}_H(\delta_{FH}, \frac{L_F\overline{B}}{f_{FF}}) > \tilde{\Pi}_H(\delta_{HH}, \frac{L_F\overline{B}}{f_{FF}}) > \mathcal{F}_H.$$

$\forall\delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$, since $\frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F),\delta_F)}{\mathrm{d}\delta_F} > 0$, there exists a unique $\delta_F^*$ s.t. $\delta_H^* = \delta_H^F(\delta_F)$ and $\tilde{\Pi}_H(\delta_H^*, \delta_F^*) = \mathcal{F}_H$. $\forall\delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$, $\tilde{\Pi}_F(\delta_H, \delta_F) = \mathcal{F}_F$ holds only when $\delta_H = \delta_{FH}$. However, now $\tilde{\Pi}_H(\delta_{FH}, \delta_F) > \mathcal{F}_H$. Hence, we obtain a unique solution.

***Case iv)*** $\tilde{\Pi}_H(\frac{L_H\overline{B}}{f_{FH}w_F}, \delta_{FF}) < \mathcal{F}_H$, $\frac{L_H\overline{B}}{f_{FH}w_F} \leq \frac{L_H\overline{B}}{f_{HH}w_H}$, and $\tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}) \geq \mathcal{F}_F$.

First consider the case of $\tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}) = \mathcal{F}_F$. Then, we obtain $(\delta_F^*, \delta_H^*) = (\frac{L_F\overline{B}}{f_{HF}w_H}, \delta_{HH})$ as a solution. $\forall\delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{HF}w_H})$, since $\frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F),\delta_F)}{\mathrm{d}\delta_F} > 0$, there is no solution. Besides, since $\forall\delta_F \in (\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F})$, $\delta_H^{F\prime}(\delta_F) < 0$, we have

$$\delta_H^F(\delta_F) < \delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}) = \delta_{HH}.$$

However, $\tilde{\Pi}_H(\delta_H, \delta_F) = \mathcal{F}_H$ holds only when $\delta_H = \delta_{HH}$. A similar proof can be applied for $\delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$. Hence, the solution is unique.

Now consider $\tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}) > \mathcal{F}_F$. $\forall\delta_F \in (\max\{\frac{L_F\overline{B}}{f_{HF}w_H}, \delta_{FF}\}, \frac{L_F\overline{B}}{f_{FF}w_F})$, $\delta_H^F(\delta_F)$ exists and decreases. Specifically, if $\frac{L_F\overline{B}}{f_{HF}w_H} \geq \delta_{FF}$, then

$$\tilde{\Pi}_F(\delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}), \frac{L_F\overline{B}}{f_{HF}w_H}) = \mathcal{F}_F < \tilde{\Pi}_F(\delta_{HH}, \frac{L_F\overline{B}}{f_{HF}w_H}),$$

and, thus, $\delta_H^F(\frac{L_F\overline{B}}{f_{HF}w_H}) > \delta_{HH}$. Besides, $\delta_H^F(\frac{L_F\overline{B}}{f_{FF}w_F}) = \delta_{FH} < \delta_{HH}$. Therefore, by the intermediate value theorem, there exists a unique solution $\delta_F^* \in (\frac{L_F\overline{B}}{f_{HF}w_H}, \frac{L_F\overline{B}}{f_{FF}w_F})$ s.t. $\delta_H^F(\delta_F^*) = \delta_{HH}$ and $\tilde{\Pi}_H(\delta_{HH}, \delta_F^*) = \mathcal{F}_H$. On the other hand, if $\delta_{FF} > \frac{L_F\overline{B}}{f_{HF}w_H}$, since $\tilde{\Pi}_H(\frac{L_H\overline{B}}{f_{FH}w_F}, \delta_{FF}) < \mathcal{F}_H$, $\tilde{\Pi}_H(\delta_{FH}, \frac{L_F\overline{B}}{f_{FF}w_F}) > \mathcal{F}_H$, and $\delta_H^{F\prime}(\delta_F) < 0$ for $\delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$, there exists a unique solution for $\delta_F^*$. For both cases, the discussions for $\delta_F = \delta_{FF}$ and $\forall\delta_F \in [\frac{L_F\overline{B}}{f_{FF}w_F}, +\infty)$ are similar to before.

***Case v)*** $\tilde{\Pi}_H(\frac{L_H\overline{B}}{f_{FH}w_F}, \delta_{FF}) < \mathcal{F}_H$ and $\frac{L_H\overline{B}}{f_{FH}w_F} > \frac{L_H\overline{B}}{f_{HH}w_H}$.

$\forall\delta_F \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$, $\delta_H^F(\delta_F)$ is continuous and differentiable, and $\delta_H^F(\delta_{FF}) = \frac{L_F\overline{B}}{f_{FH}w_F}$ and $\delta_H^F(\frac{L_F\overline{B}}{f_{FF}w_F}) = \delta_{FH}$. Since $\frac{L_H\overline{B}}{f_{FH}w_F} > \frac{L_H\overline{B}}{f_{HH}w_H} > \delta_{HH} > \delta_{FH}$, there exists a $\delta_F^- \in (\delta_{FF}, \frac{L_F\overline{B}}{f_{FF}w_F})$ s.t. $\delta_H^F(\delta_F^-) = \frac{L_H\overline{B}}{f_{HH}w_H}$. Then,

$$\forall\delta_F \in (\delta_F^-, \min\{\frac{L_F\overline{B}}{f_{FF}w_F}, \frac{L_F\overline{B}}{f_{HF}w_H}\}), \quad \frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F),\delta_F)}{\mathrm{d}\delta_F} > 0.$$

Then, one can prove uniqueness of the solution $\delta_F^* \in (\delta_F^-, \frac{L_F\overline{B}}{f_{FF}w_F})$ in a similar way as in Cases *iii*) and *iv*). Lastly, we need to check that there is no solution $\forall\delta_F \in [\delta_{FF}, \delta_F^-]$.

Within this interval, since $\delta_{HF} < \delta_{FF}$,

$$\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_{HF}) = \mathcal{F}_H > \tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F).$$

Hence, we show that the solution is unique.

Last, we consider the case of $f_{ij} = 0$, $\forall i, j = H, F$. Now $\tilde{\Pi}_{ij}(\delta_j)$ is continuously differentiable $\forall \delta_j \in (0, +\infty)$. As shown before, we have an implicit function $\delta_H^F(\delta_F)$ s.t. $\tilde{\Pi}_F(\delta_H^F(\delta_F), \delta_F) = \mathcal{F}_F$ holds, $\forall \delta_F \in (\delta_{FF}, +\infty)$. At the limit,

$$\lim_{\delta_F \to \delta_{FF}} \delta_F^H(\delta_F) = +\infty \text{ and } \lim_{\delta_F \to +\infty} \delta_F^H(\delta_F) = \delta_{FH}.$$

Given that $\delta_{FF} > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}$, we obtain

$$\lim_{\delta_F \to \delta_{FF}} \tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F) = \tilde{\Pi}_{HF}(\delta_{FF}) < \tilde{\Pi}_{HF}(\delta_{HF}) = \mathcal{F}_H.$$

and

$$\lim_{\delta_F \to +\infty} \tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F) = \tilde{\Pi}_{HH}(\delta_{FH}) > \tilde{\Pi}_{HH}(\delta_{HH}) = \mathcal{F}_H.$$

With Assumption 4, we have $\frac{\mathrm{d}\tilde{\Pi}_H(\delta_H^F(\delta_F), \delta_F)}{\mathrm{d}\delta_F} > 0$, then there exists a unique solution $\delta_F^*$ s.t. $\tilde{\Pi}_H(\delta_H^F(\delta_F^*), \delta_F^*) = \mathcal{F}_H$.

Overall, we prove that for all $w_F \in (\underline{w_F}, \overline{w_F})$, there exists a unique solution $(\delta_F^*(w_F), \delta_H^*(w_F))$ s.t. the ZEPCs in both countries hold.

**Step 2. The existence and uniqueness of $w_F^*$ s.t. the TBC.** In Step 1, we show $\forall w_F \in (\underline{w_F}, \overline{w_F})$, the solution $(\delta_F^*(w_F), \delta_H^*(w_F))$ is uniquely determined. $\forall \delta_F \in (\delta_{FF}, \frac{L_F \overline{B}}{f_{FF} w_F})$, there exists an implicit function $\delta_H^F(\delta_F, w_F)$ s.t.

$$\tilde{\Pi}_{FH}(\delta_H^F(\delta_F, w_F), w_F) + \tilde{\Pi}_{FF}(\delta_F, w_F) = \mathcal{F}_F \text{ and } \frac{\partial \delta_H^F(\delta_F, w_F)}{\partial \delta_F} < 0.$$

Since $\forall \delta_F \in [\frac{L_F \overline{B}}{f_{FF} w_F}, +\infty)$, $\tilde{\Pi}_{FF}(\delta_F, w_F) = 0$ and $\tilde{\Pi}_{FH}(\delta_{FH}(w_F), w_F) = \mathcal{F}_F$, we can refine $\delta_H^F(\delta_F, w_F) = \delta_{FH}(w_F)$ and, thus, $\frac{\partial \delta_H^F(\delta_F, w_F)}{\partial \delta_F} = 0$.

On the other hand, the ZEPC in country $H$ is independent of the endogenous wage, $w_F$. Therefore, similarly, $\forall \delta_F \in (\delta_{HF}, \frac{L_F \overline{B}}{f_{HF}})$, we obtain an implicit function $\delta_H^H(\delta_F)$ s.t. the corresponding ZEPC, and $\delta_H^{H\prime}(\delta_F) < 0$. We refine $\delta_H^H(\delta_F) = \delta_{HH}$, $\forall \delta_F \in [\frac{L_F \overline{B}}{f_{HF}}, +\infty)$ and, thus, $\delta_H^{H\prime}(\delta_F) = 0$.

We further define the function measuring the difference between two implicit functions as follows:

$$\Delta(\delta_F, w_F) \equiv \delta_H^F(\delta_F, w_F) - \delta_H^H(\delta_F).$$

When $\frac{L_H\bar{B}}{f_{FH}w_F} \leqslant \delta_H^H(\delta_{FF}(w_F))$, $\delta_F^*(w_F) = \delta_{FF}(w_F)$, we refine $\Delta(\delta_{FF}(w_F), w_F) = 0$. When $\frac{L_H\bar{B}}{f_{FH}w_F} > \delta_H^H(\delta_{FF}(w_F))$, $\delta_F^*(w_F) > \delta_{FF}(w_F)$, we refine

$$\Delta(\delta_{FF}(w_F), w_F) = \frac{L_H\bar{B}}{f_{FH}w_F} - \delta_H^H(\delta_{FF}(w_F)) > 0.$$

When $f_{ij} = 0$, $\forall i, j = H, F$, $\delta_H^F(\delta_F, w_F)$ and $\delta_H^H(\delta_F)$ are continuously differentiable in $(\delta_{FF}(w_F), +\infty)$. We then refine

$$\Delta(\delta_{FF}(w_F), w_F) = \lim_{\delta_F \to \delta_{FF}(w_F)} \left[\delta_H^F(\delta_F, w_F) - \delta_H^H(\delta_F)\right] = +\infty$$

and

$$\lim_{\delta_F \to +\infty} \Delta(\delta_F, w_F) = \delta_{HH} - \delta_{FH}(w_F) > 0.$$

Recall that $\forall w_F \in (\underline{w}_F, \overline{w}_F)$, we have $\delta_{FF}(w_F) > \delta_{HF}$ and $\delta_{HH} > \delta_{FH}(w_F)$. According to our definition of $\Delta(\delta_F, w_F)$, with the existence and uniqueness of $(\delta_H^*(w_F), \delta_F^*(w_F))$, we obtain the following properties:

$$\Delta(\delta_F, w_F) \begin{cases} > 0, & \delta_F \in [\delta_{FF}(w_F), \delta_F^*(w_F)) \\ = 0, & \delta_F = \delta_F^*(w_F) \\ < 0, & \delta_F \in (\delta_F^*(w_F), +\infty). \end{cases}$$

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 $\frac{\partial\tilde{\Pi}_{FF}}{\partial w_F} < 0$,

$$\tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-) > \mathcal{F}_F.$$

Define the new implicit function as $\delta_H^F(\delta_F, w_F^-)$ s.t.

$$\tilde{\Pi}_{FH}(\delta_H^F(\delta_F, w_F^-), w_F^-) + \tilde{\Pi}_{FF}(\delta_F, w_F^-) = \mathcal{F}_F.$$

We evaluate the latter at $\delta_F = \delta_F^*(w_F)$ and obtain

$$\tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F^-), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-) = \mathcal{F}_F < \tilde{\Pi}_{FH}(\delta_H^F(\delta_F^*(w_F), w_F), w_F^-) + \tilde{\Pi}_{FF}(\delta_F^*(w_F), w_F^-),$$

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 $\delta_H^H(\delta_F)$ is independent of $w_F$, and we have:

$$\Delta(\delta_F^*(w_F), w_F^-) = \delta_H^F(\delta_F^*(w_F), w_F^-) - \delta_H^H(\delta_F^*(w_F)) > \delta_H^F(\delta_F^*(w_F), w_F) - \delta_H^H(\delta_F^*(w_F)) = 0.$$

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 $\delta_H^{H\prime}(\delta_F) < 0, \forall \delta_F \in (\delta_{HF}, \frac{L_F\overline{B}}{f_{HF}})$, and $\delta_H^{H\prime}(\delta_F) = 0, \forall \delta_F \in (\frac{L_F\overline{B}}{f_{HF}}, +\infty)$. If $\delta_F^*(w_F) < \frac{L_F\overline{B}}{f_{HF}}$,

$$\delta_H^*(w_F^-) = \delta_H^H(\delta_F^*(w_F^-)) < \delta_H^H(\delta_F^*(w_F)) = \delta_H^*(w_F)$$

and, thus, $\delta_H^{*\prime}(w_F) < 0$. If $\delta_F^*(w_F) \geqslant \frac{L_F\overline{B}}{f_{HF}}$,

$$\delta_H^*(w_F^-) = \delta_H^H(\delta_F^*(w_F^-)) = \delta_{HH} = \delta_H^H(\delta_F^*(w_F)) = \delta_H^*(w_F)$$

and, thus, $\delta_H^{*\prime}(w_F) = 0$. At the limit, when $w_F \to \underline{w_F}$,

$$\delta_{FH}(w_F) \to \delta_{HH} \text{ and } \lim_{w_F \to \underline{w_F}} \delta_H^*(w_F) = \delta_{HH} = \delta_{FH}(\underline{w_F}),$$

indicating that firms in both countries only sell to country $H$.

Similarly, one can show that, when $w_F$ increases to $w_F^+$, $\delta_F^*(w_F^+) < \delta_F^*(w_F)$. Besides, when $\delta_F^* > \frac{L_F\overline{B}}{f_{HF}w_H}$,

$$\delta_H^*(w_F^+) = \delta_H^H(\delta_F^*(w_F^+)) = \delta_{HH} = \delta_H^H(\delta_F^*(w_F)) = \delta_H^*(w_F).$$

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 $w_F \to \overline{w_F}$,

$$\delta_{FF}(w_F) \to \delta_{HF} \text{ and } \lim_{w_F \to \overline{w_F}} \delta_F^*(w_F) = \delta_{FF}(\overline{w_F}) = \delta_{HF},$$

showing that firms in both countries only sell to country $F$.

Now consider the trade-balance condition (TBC):

$$M_H(\delta_F^*(w_F), \delta_H^*(w_F))\mathcal{R}_{HF}(\delta_F^*(w_F))L_F = M_F(\delta_F^*(w_F), \delta_H^*(w_F), w_F)\mathcal{R}_{FH}(\delta_H^*(w_F)w_F)L_H,$$

where $\mathcal{R}_{ij}(\delta_j^*(w_F), w_i) = \int_{\varphi_{ij}^*(\delta_j^*(w_F))}^{+\infty} p_{ij}(\delta_j^*(w_F)w_i, \varphi)q_{ij}(\delta_j^*(w_F)w_i, \varphi)dG_i(\varphi)$. With the ZEPCs and the resource constraints (RCs), we can express the masses of entrants as follows:

$$M_i(\delta_F^*(w_F), \delta_H^*(w_F), w_i) = \frac{L_i}{\sum_j \tilde{\mathcal{R}}_{ij}(\delta_j^*(w_F), w_i)L_j}.$$

Therefore, the TBC can be re-written as

$$L_H \frac{\tilde{\mathcal{R}}_{HF}(\delta_F^*(w_F))L_F}{\sum_j \tilde{\mathcal{R}}_{Hj}(\delta_j^*(w_F))L_j} = L_F w_F \frac{\tilde{\mathcal{R}}_{FH}(\delta_H^*(w_F)w_F)L_H}{\sum_j \tilde{\mathcal{R}}_{Fj}(\delta_j^*(w_F)w_F)L_j}.$$

Recall that $\delta_F^*(w_F)$ strictly decreases with $w_F$, and $\delta_H^*(w_F)$ decreases with $w_F$. 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 to $L_H$, 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. Given the endogenous wage $w_F^*$, we can determine the solution $(\delta_F^*(w_F^*), \delta_H^*(w_F^*))$, as well as the corresponding quantities, cutoffs, and masses of entrants. □

## A.3 Proof of Proposition 2

*Proof.* For brevity, we refer to $\lambda_i^{opt}$ by $\lambda_i$ without ambiguity in this proof. 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 $i$:

$$u'(q_{ij}) = \frac{\lambda_i \tau_{ij} w_i}{\varphi}, \tag{24}$$

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 $\lambda_i$ 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 $\lambda_i$ 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 $\varphi < \frac{\lambda_i \tau_{ij} w_i}{u'(0)}$ should not produce from a social planner's view, so that their $q_{ij}(\lambda_i, \frac{\tau_{ij} w_i}{\varphi}) = 0$.

When condition (24) 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 $\lambda_i$. Therefore, we obtain

$$\frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \lambda_i} < 0 \ , \ \frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi})}{\partial \frac{\tau_{ij}w_i}{\varphi}} < 0. \tag{25}$$

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

$$\lim_{\frac{\tau_{ij}w_i}{\varphi} \to +\infty} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = \lim_{\lambda_i \to +\infty} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = -f_{ij}w_i. \tag{26}$$

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

$$\lim_{\frac{\tau_{ij}w_i}{\varphi} \to 0} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = \frac{L_j \overline{B}^s}{\lambda_i} - f_{ij}w_i \ , \ \lim_{\lambda_i \to 0} \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty. \tag{27}$$

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 (25), (26), and (27), 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. $\pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) = 0$ 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 (25), we can apply the implicit function theorem to obtain

$$\frac{\mathrm{d}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\mathrm{d}\lambda_i} = -\frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\partial \lambda_i} / \frac{\partial \pi_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi^*_{ij}})}{\partial(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})} < 0 \text{ and } \frac{d\varphi^*_{ij}}{d\lambda_i} > 0.$$

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:

$$\sum_j \int_{\varphi^*_{ij}(\lambda_i)}^{+\infty} \left\{ \left[ \frac{1}{\varepsilon_u(q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1 \right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\lambda_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i \right\} dG_i(\varphi) = \mathcal{F}_i w_i. \quad (28)$$

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

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

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 (28), rewrite it as

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

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

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

When there is no confusion, we simplify $\underline{j}^i$ as $\underline{j}$. Furthermore, we use $q_{i\underline{j}}(\lambda_i, \varphi)$ rather than $q_{i\underline{j}}(\lambda_i, \frac{\tau_{ij} w_i}{\varphi})$ interchangeably for simplicity.

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

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

However, the LHS of (28) does not converge to 0, since $\frac{L_{\bar{j}} \overline{B}^s}{f_{i\bar{j}} w_i} > \frac{L_{\underline{j}} \overline{B}^s}{f_{i\underline{j}} w_i}$, $\lim_{\lambda_i \to \frac{L_{\underline{j}} \overline{B}^s}{f_{i\underline{j}} w_i}} \varphi^*_{i\bar{j}}(\lambda_i) < +\infty$, and $\lim_{\lambda_i \to \frac{L_{\underline{j}} \overline{B}^s}{f_{i\underline{j}} w_i}} q_{i\bar{j}}(\lambda_i, \varphi) > 0$.

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

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

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

In sum, as $\lambda_i$ goes from 0 to $\frac{L_{\bar{j}} \overline{B}^s}{f_{i\bar{j}} w_i}$, the LHS of (28) decreases from $+\infty$ to 0. Therefore, $\forall \mathcal{F}_i w_i > 0$, there exists a unique $\lambda_i^{opt} > 0$ s.t. the ZESPC (28) 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 are then determined by the resource constraint.

Last, 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 $\lambda_i$ 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 $\lambda_i$. Consider the ZESPCs in (28), where the LHS decreases with $\lambda_i$. We have LHS $\to +\infty$ when $\lambda_i \to 0$, and LHS $\to 0$ when $\lambda_i \to +\infty$. Therefore, $\forall \mathcal{F}_i w_i > 0$, there exists a unique $\lambda_i^{opt}$ s.t. the ZESPC is satisfied, and the corresponding equilibrium outcomes can be obtained. □

## A.4 Proof of Proposition 3

*Proof.* For brevity, we refer to $\delta_i^{cmkt}$ by $\delta_i$ 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 planner 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) $\lim_{\frac{\tau_{ij}w_i}{\varphi} \to 0} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = \frac{L_j\overline{B}}{\delta_i} - f_{ij}w_i$, $\lim_{\delta_i \to 0} \pi_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) = +\infty$. 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, $\forall i = H, F$:

$$\sum_j \Pi_{ij}(\delta_i) = \sum_j \int_{\varphi_{ij}^*(\delta_i)}^{+\infty} \left\{ \left[\frac{1}{1 - r_u(q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}))} - 1\right] \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi}) L_j - f_{ij}w_i \right\} dG_i(\varphi) = \mathcal{F}_i w_i. \tag{29}$$

We can obtain $\frac{\partial \Pi_{ij}(\delta_i)}{\partial \delta_i} < 0$. Thus, the LHS of (29) is decreasing in $\delta_i$. We further explore the possible range of the LHS by rewriting the ZEPC (29) as:

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

When $\delta_i \to 0$, the cutoff productivity $\varphi_{ij}^*(\delta_i) \to 0$, the cutoff quantity $q_{ij}(\delta_i, \frac{\tau_{ij}w_i}{\varphi_{ij}^*}) \to \overline{q}$, 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

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

Again, we simplify $\underline{j}^i$ as $\underline{j}$ and $q_{i\underline{j}}(\delta_i, \frac{\tau_{ij} w_i}{\varphi})$ as $q_{i\underline{j}}(\delta_i, \varphi)$ when possible.

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

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

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

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

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

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

and the LHS of (29) $\to 0$.

Overall, as $\delta_i$ increases from 0 to $\frac{L_{\overline{j}} \overline{B}}{f_{i\overline{j}} w_i}$, the LHS strictly decreases from $+\infty$ to 0. Therefore, $\forall \mathcal{F}_i w_i > 0$, there exists a unique $\delta_i^{cmkt} > 0$ s.t. $\sum_j \Pi_{ij}(\delta_i^{cmkt}) = \mathcal{F}_i w_i$, 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 $\delta_i$ guarantee non-negative profits. From the FOCs, $[1 - r_u(0)]\, u'(0) = \frac{\delta_i \tau_{ij} w_i}{\varphi^*_{ij}}$. Hence, when $\delta_i$ increases, the cutoff $\varphi^*_{ij}(\delta_i)$ increases. Recall that the LHS of the ZEPCs (29) decreases with $\delta_i$, the LHS $\to +\infty$ when $\delta_i \to 0$, and the LHS $\to 0$ when $\delta_i \to +\infty$. Therefore, $\forall \mathcal{F}_i w_i > 0$, there exists a unique $\delta_i^{cmkt}$ s.t. the ZEPC, 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 $M_i$, $\forall i, j = H, F$. □

## A.5 Proof of Proposition 4

*Proof.* Without loss of generality, we assume that the solution of the decentralized equilibrium is $(\delta_F^{dmkt}, \delta_H^{dmkt})$ where $\delta_H^{dmkt} > \delta_F^{dmkt}$. We employ the definition of $\Delta(\delta_F)$, $\delta_H^H(\delta_F)$, and $\delta_H^F(\delta_F)$ in the proof of Proposition 1. At the equilibrium wage of decentralized market $w_F$, the corresponding outcome satisfies $\delta_H^H(\delta_F^{dmkt}) = \delta_H^F(\delta_F^{dmkt}) = \delta_H^{dmkt}$, and the centralized market outcome 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 $(\frac{L_F \overline{B}}{f_{HF} w_H}, +\infty)$, with the uniqueness of $\delta_H^{cmkt}$, $\delta_H^{cmkt} > \delta_F^{dmkt}$ 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,

$$\delta_H^H(\delta_{FF}) = \delta_H^{dmkt} = \delta_{HH} > \delta_{FF} \text{ and } \Delta(\delta_{FF}) = 0.$$

Then, $\forall \delta_F > \delta_{FF}$, $\Delta(\delta_F) < 0$. 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 $(\frac{L_F \overline{B}}{f_{FF} w_F}, +\infty)$, with the uniqueness of $\delta_F^{cmkt}$, $\delta_F^{cmkt} < \delta_H^{cmkt}$ 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 $(\frac{L_F \overline{B}}{f_{FF} w_H}, +\infty)$, with the uniqueness of $\delta_F^{cmkt}$, $\delta_F^{cmkt} > \delta_F^{dmkt}$. Similarly, given that $\delta_H^H(\delta_F)$ is strictly decreasing in $(\delta_{HF}, \frac{L_F \overline{B}}{f_{HF} w_H})$ and constant in $(\frac{L_F \overline{B}}{f_{HF} w_H}, +\infty)$, $\delta_F^{dmkt} \in (\delta_{FF}, \frac{L_F \overline{B}}{f_{HF} w_H})$, we have

$$\delta_H^{dmkt} = \delta_H^H(\delta_F^{dmkt}) > \delta_H^H(\delta_H^{cmkt}) = \delta_H^{cmkt}.$$

Furthermore, since $\delta_F^{cmkt} > \delta_F^{dmkt}$,

$$\Delta(\delta_F^{cmkt}) = \delta_H^F(\delta_F^{cmkt}) - \delta_H^H(\delta_F^{cmkt}) < 0,$$

we obtain $\delta_F^{cmkt} < \delta_H^H(\delta_F^{cmkt})$ and then $\delta_H^{cmkt} > \delta_F^{cmkt}$. 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 $\delta_j^{dmkt}$ and $\delta_i^{cmkt}$ by $\delta$ in the following statement. Then, for both the decentralized and centralized equilibria, the cutoff behaves as $\varphi_{ij}^{*\prime}(\delta) > 0$ for $\delta \in (0, \frac{L_j\overline{B}}{f_{ij}w_i})$ and $\varphi_{ij}^{*\prime}(\delta) = 0$ for $\delta \in (\frac{L_j\overline{B}}{f_{ij}w_i}, +\infty)$, and the quantity behaves as $\frac{\partial q_{ij}(\delta,\varphi)}{\partial\delta} < 0$ for $\delta \in (0, \frac{L_j\overline{B}}{f_{ij}w_i})$ and $\frac{\partial q_{ij}(\delta,\varphi)}{\partial\delta} = 0$ 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. □

## A.6 Proof of Proposition 5

*Proof.* 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

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

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 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 $\varphi$ irrespective of the cutoff and purely rely on the FOCs w.r.t. quantity of two equilibria. $\forall i, j = H, F$ and $\forall\varphi > 0$:

$$\begin{cases}\left[u''(q_{ij}^{cmkt})q_{ij}^{cmkt} + u'(q_{ij}^{cmkt})\right] = \dfrac{\delta_i^{cmkt}\tau_{ij}w_i}{\varphi} \\ u'(q_{ij}^{opt}) = \dfrac{\lambda_i^{opt}\tau_{ij}w_i}{\varphi},\end{cases} \tag{30}$$

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 (30), 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 $\varphi$ 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 $q_{ij}^{cmkt}(\tilde{\varphi}_{ij}) = q_{ij}^{opt}(\tilde{\varphi}_{ij})$. $\forall \varphi > \tilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) > q_{ij}^{cmkt}(\tilde{\varphi}_{ij})$, $1 - r_u(q_{ij}^{cmkt}(\varphi)) < 1 - r_u(q_{ij}^{cmkt}(\tilde{\varphi}_{ij})) = \frac{\delta_i^{cmkt}}{\lambda_i^{opt}}$, and, hence, $\frac{u'(q_{ij}^{cmkt}(\varphi))}{u'(q_{ij}^{opt}(\varphi))} > 1$ and $q_{ij}^{cmkt}(\varphi) < q_{ij}^{opt}(\varphi)$. Similarly, $\forall \varphi < \tilde{\varphi}_{ij}$ 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 $q$-$\varphi$-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 $\forall q \in [0, \overline{q})$ and $\tau_{HF} > \tau_{HH}$, $\frac{\partial[1-r_u(q)]u'(q)}{\partial q} < 0$, 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. $q_{ij}^{cmkt}(\tilde{\varphi}_{ij}) = q_{ij}^{opt}(\tilde{\varphi}_{ij})$. $\forall \varphi > \tilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) > q_{ij}^{opt}(\varphi)$. $\forall \varphi < \tilde{\varphi}_{ij}$, $q_{ij}^{cmkt}(\varphi) < q_{ij}^{opt}(\varphi)$. 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 $\tilde{\varphi}_{HH} < \tilde{\varphi}_{HF}$. □

## A.7 Proof of Proposition 6

*Proof.* 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

$$\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) \equiv u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi}) - u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) = u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))\left[\varepsilon_u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) - 1\right]. \tag{31}$$

Then,

$$v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) = u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})) + \alpha\omega(q_{ij}(\frac{\tau_{ij}w_i}{\varphi})). \tag{32}$$

Consider the weighted-average Lagrangian

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

where $\beta_i(\alpha)$ is the Lagrange multiplier depending on weight $\alpha$. 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. $\alpha$, 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:

$$v_\alpha(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})) = \beta_i(\alpha)\left[\frac{\tau_{ij}w_i}{\varphi^*_{ij}}q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}) + \frac{f_{ij}w_i}{L_j}\right]. \tag{33}$$

Differentiate the LHS w.r.t. $\alpha$ 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 (31) and (32), we can simplify the numerator as

$$\begin{aligned}&u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}})\left[\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))dG_i(\varphi)\right]\\ &-u(q_{ij}(\frac{\tau_{ij}w_i}{\varphi^*_{ij}}))\left[\sum_j L_j\int_{\varphi^*_{ij}}^{+\infty}u'(q_{ij}(\frac{\tau_{ij}w_i}{\varphi}))q_{ij}(\frac{\tau_{ij}w_i}{\varphi})dG_i(\varphi)\right]\gtrless 0. \qquad (34)\end{aligned}$$

With the definition of $\underline{j}$ and $\overline{j}$, we simplify $q_{ij}(\frac{\tau_{ij}w_i}{\varphi})$ as $q_{ij}(\varphi)$ and rewrite (34) as

$$\varepsilon_u(q_{ij}(\varphi^*_{ij}))\gtrless\frac{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}\varepsilon_u(q_{i\underline{j}}(\varphi))u(q_{i\underline{j}}(\varphi))dG_i(\varphi) + L_{\overline{j}}\int_{\varphi^*_{i\overline{j}}}^{+\infty}\varepsilon_u(q_{i\overline{j}}(\varphi))u(q_{i\overline{j}}(\varphi))dG_i(\varphi)}{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}u(q_{i\underline{j}}(\varphi))dG_i(\varphi) + L_{\overline{j}}\int_{\varphi^*_{i\overline{j}}}^{+\infty}u(q_{i\overline{j}}(\varphi))dG_i(\varphi)}. \qquad (35)$$

Combining (33) with the FOC $\left\{\alpha\left[1-r_u(q_{ij}(\varphi^*_{ij}))\right]+(1-\alpha)\right\}u'(q_{ij}(\varphi^*_{ij})) = \beta_i(\alpha)\frac{\tau_{ij}w_i}{\varphi^*_{ij}}$, we obtain

$$\beta_i(\alpha)f_{ij}w_i = L_j\left\{(1-\alpha)\left[u(q_{ij}(\varphi^*_{ij})) - u'(q_{ij}(\varphi^*_{ij}))q_{ij}(\varphi^*_{ij})\right] + \alpha\left[-u''(q_{ij}(\varphi^*_{ij}))(q_{ij}(\varphi^*_{ij}))^2\right]\right\}. \qquad (36)$$

Since $\left[u(q_{ij}(\varphi^*_{ij})) - u'(q_{ij}(\varphi^*_{ij}))q_{ij}(\varphi^*_{ij})\right]$ and $\left[-u''(q_{ij}(\varphi^*_{ij}))(q_{ij}(\varphi^*_{ij}))^2\right]$ both increase in $q_{ij}(\varphi^*_{ij})$, the RHS of (36) increases in $q_{ij}(\varphi^*_{ij})$. Combine the conditions w.r.t $\underline{j}$ and $\overline{j}$:

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

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

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

If $\varepsilon'_u(q)>0$, then $\varepsilon_u(q_{i\underline{j}}(\varphi^*_{i\underline{j}}))>\varepsilon_u(q_{i\bar{j}}(\varphi^*_{i\bar{j}}))$. We can rewrite (35) as

$$\frac{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}\varepsilon_u(q_{i\underline{j}}(\varphi))u(q_{i\underline{j}}(\varphi))dG_i(\varphi)+L_{\bar{j}}\int_{\varphi^*_{i\bar{j}}}^{+\infty}\varepsilon_u(q_{i\bar{j}}(\varphi))u(q_{i\bar{j}}(\varphi))dG_i(\varphi)}{L_{\underline{j}}\int_{\varphi^*_{i\underline{j}}}^{+\infty}u(q_{i\underline{j}}(\varphi))dG_i(\varphi)+L_{\bar{j}}\int_{\varphi^*_{i\bar{j}}}^{+\infty}u(q_{i\bar{j}}(\varphi))dG_i(\varphi)}$$

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

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 $(\varphi^*_{i\underline{j}})^{cmkt}\gtrless(\varphi^*_{i\underline{j}})^{opt}$. □

## A.8 Proof of Proposition 7 and Lemma 3

*Proof.* 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. □

## A.9 Specialized results for CES preferences

*Proof.* 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 $\forall i,j=H,F$,

$$\begin{cases}\rho^2(q_{ij}(\varphi))^{\rho-1}=\dfrac{\delta^{dmkt}_j\tau_{ij}w_i}{\varphi}\\ \rho^2(q_{ij}(\varphi^*_{ij}))^{\rho-1}=\dfrac{\delta^{dmkt}_j\tau_{ij}w_i}{\varphi^*_{ij}}\\ (\dfrac{1}{\rho}-1)\dfrac{\tau_{ij}}{\varphi^*_{ij}}q_{ij}(\varphi^*_{ij})L_j=f_{ij}.\end{cases}\tag{37}$$

We obtain explicit expressions for the quantity functions as follows:

$$\begin{cases} q_{ij}(\varphi) = \dfrac{\rho}{1-\rho}\dfrac{f_{ij}}{L_j\tau_{ij}}(\dfrac{1}{\varphi_{ij}^*})^{\frac{\rho}{1-\rho}}\varphi^{\frac{1}{1-\rho}} \\ \dfrac{\varphi_{jj}^*}{\varphi_{ij}^*} = (\dfrac{f_{jj}}{f_{ij}})^{\frac{1-\rho}{\rho}}(\dfrac{\tau_{jj}}{\tau_{ij}})(\dfrac{w_j}{w_i})^{\frac{1}{\rho}}. \end{cases} \tag{38}$$

Then the ZEPCs read:

$$\begin{aligned} \mathcal{F}_i &= \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left[(\frac{1}{\rho}-1)\frac{\tau_{ij}}{\varphi}q_{ij}(\varphi)L_j - f_{ij}\right] dG_i(\varphi) \\ &= \frac{\rho}{(1-\rho)\gamma-\rho}\sum_j f_{ij}(\frac{1}{\varphi_{ij}^*})^{\gamma}. \end{aligned} \tag{39}$$

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. Assumption 3 requires $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$ where $\mathcal{C}_{ij} = (\delta_{ij}w_i)^{-1}$ and $\delta_{ij}w_i$ is the solution to the following counterfactual equilibrium:

$$\begin{cases} \rho^2(q_{ij}(\varphi))^{\rho-1} = \dfrac{\delta_{ij}w_i\tau_{ij}}{\varphi} \\ (\dfrac{1}{\rho}-1)\dfrac{\tau_{ij}}{\varphi_{ij}^*}q_{ij}(\varphi_{ij}^*)L_j = f_{ij} \\ \dfrac{\rho}{(1-\rho)\gamma-\rho}f_{ij}(\dfrac{1}{\varphi_{ij}^*})^{\gamma} = \mathcal{F}_i. \end{cases} \Rightarrow (\delta_{ij}w_i)^{\frac{\gamma}{\rho}} = \frac{\rho^{\frac{(1+\rho)\gamma+\rho}{\rho}}(1-\rho)^{\frac{(1-\rho)\gamma}{\rho}}}{(1-\rho)\gamma-\rho}L_j^{\frac{(1-\rho)\gamma}{\rho}}\left[(\frac{1}{\mathcal{F}_i})(\frac{1}{f_{ij}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{ij}})^{\gamma}\right]$$

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

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

and Assumption 3 can be rewritten as:

$$w_F \in (\underline{w_F}, \overline{w_F}) = \left(\frac{\mathcal{C}_{HH}}{\mathcal{C}_{FH}}, \frac{\mathcal{C}_{HF}}{\mathcal{C}_{FF}}\right) = \left((\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{\rho}{\gamma}}(\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{HH}}{\tau_{FH}})^{\rho}, (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{\rho}{\gamma}}(\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{HF}}{\tau_{FF}})^{\rho}\right).$$

$$(\frac{f_{HH}}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{HH}}{\tau_{FH}})^{\rho} < (\frac{f_{HF}}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\gamma}}(\frac{\tau_{HF}}{\tau_{FF}})^{\rho},$$

As for Assumption 4, we first rewrite the ZEPC in terms of $\delta_H$ and $\delta_F$, $\forall i = H, F$:

$$L_H^{\frac{(1-\rho)\gamma}{\rho}}(\frac{1}{f_{iH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{iH}})^{\gamma}(\frac{1}{\delta_H})^{\frac{\gamma}{\rho}}+L_F^{\frac{(1-\rho)\gamma}{\rho}}(\frac{1}{f_{iF}})^{(1-\rho)\gamma-1}(\frac{1}{\tau_{iF}})^{\gamma}(\frac{1}{\delta_F})^{\frac{\gamma}{\rho}} = \frac{\mathcal{F}_i(w_i)^{\frac{\gamma}{\rho}}[(1-\rho)\gamma-\rho]}{\rho^{\frac{(1+\rho)\gamma+\rho}{\rho}}(1-\rho)^{\frac{(1-\rho)\gamma}{\rho}}}.$$

Then, the positivity of the corresponding Jacobian determinant can be expressed as:

$$\begin{vmatrix} (\frac{1}{f_{HH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{HH}})^{\gamma}, & (\frac{1}{f_{HF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{HF}})^{\gamma} \\ (\frac{1}{f_{FH}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{FH}})^{\gamma}, & (\frac{1}{f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{FF}})^{\gamma} \end{vmatrix} > 0.$$

We can simplify both Assumption 3 and 4 as:

$$(\frac{1}{f_{HH}f_{FF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{HH}\tau_{FF}})^{\gamma} > (\frac{1}{f_{FH}f_{HF}})^{\frac{(1-\rho)\gamma-\rho}{\rho}}(\frac{1}{\tau_{FH}\tau_{HF}})^{\gamma}. \tag{40}$$

Therefore, assuming (40) can sufficiently restrict the parameter space such that the decentralized market equilibrium is uniquely determined.

Now we can solve for the cutoff productivities with (38) and (39). 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:

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

and

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

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

$$M_i\left\{\frac{\gamma-\rho}{(1-\rho)\gamma-\rho}\sum_j f_{ij}(\frac{1}{\varphi_{ij}^*})^{\gamma}+\mathcal{F}_i\right\} = L_i. \tag{41}$$

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

$$M_i = \frac{L_i\rho}{\mathcal{F}_i\gamma}.$$

$\forall i \neq j$, we further rewrite the TBC as

$$\frac{L_i f_{ij}}{\mathcal{F}_i} w_i (\frac{1}{\varphi^*_{ij}})^{\gamma} = \frac{L_j f_{ji}}{\mathcal{F}_j} w_j (\frac{1}{\varphi^*_{ji}})^{\gamma},$$

and the implicit solution of the relative wage ratio is:

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

**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 $\forall i, j = H, F$,

$$\begin{cases} \rho^2 (q_{ij}(\varphi))^{\rho-1} = \dfrac{\delta_i^{cmkt} \tau_{ij} w_i}{\varphi} \\ \rho^2 (q_{ij}(\varphi^*_{ij}))^{\rho-1} = \dfrac{\delta_i^{cmkt} \tau_{ij} w_i}{\varphi^*_{ij}} \\ (\dfrac{1}{\rho} - 1) \dfrac{\tau_{ij}}{\varphi^*_{ij}} q_{ij}(\varphi^*_{ij}) L_j = f_{ij}. \end{cases} \quad (42)$$

We obtain explicit expressions for the quantity functions as follows:

$$\begin{cases} q_{ij}(\varphi) = \dfrac{\rho}{1-\rho} \dfrac{f_{ij}}{L_j \tau_{ij}} (\dfrac{1}{\varphi^*_{ij}})^{\frac{\rho}{1-\rho}} \varphi^{\frac{1}{1-\rho}} \\ \dfrac{\varphi^*_{ii}}{\varphi^*_{ij}} = \dfrac{\tau_{ii}}{\tau_{ij}} (\dfrac{f_{ii} L_j}{f_{ij} L_i})^{\frac{1-\rho}{\rho}}. \end{cases} \quad (43)$$

The ZEPCs for the centralized market can also be simplified to (39). Note that, in contrast to the decentralized market equilibrium, the cutoff productivity $\varphi^*_{ij}$ is proportional to $\varphi^*_{ii}$ rather than to $\varphi^*_{jj}$ in the centralized market equilibrium. Therefore, the ZEPC of country $i$ alone pins down the cutoffs $\varphi^*_{ii}$ and $\varphi^*_{ij}$. We can then obtain the explicit cutoffs as follows:

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

With the solutions of cutoffs and (41), we obtain the solution of masses of entrants as

$M_i = \frac{L_i \rho}{\mathcal{F}_i \gamma}$. □

## A.10 Specialized results for CARA preferences

*Proof.* 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 $\mathcal{W}$, which satisfies $z = \mathcal{W}(z)e^{\mathcal{W}(z)}$, and obtain explicit expressions for the quantity functions as follows:

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\left[1 - \mathcal{W}(e\dfrac{\varphi^*_{ij}}{\varphi})\right] \\ \quad \varphi^*_{ij} = \dfrac{\tau_{ij}w_i}{\tau_{jj}w_j}\varphi^*_{jj}. \end{cases} \tag{44}$$

Define $z_{ij} \equiv \mathcal{W}(e\frac{\varphi^*_{ij}}{\varphi})$, 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

$$d\varphi = \frac{-\varphi^*_{ij}(z_{ij}+1)}{z_{ij}^2 e^{z_{ij}-1}}dz_{ij},\ dG(\varphi) = -\gamma\frac{1}{(\varphi^*_{ij})^\gamma}\frac{z_{ij}+1}{z_{ij}}(z_{ij}e^{z_{ij}-1})^\gamma dz_{ij}. \tag{45}$$

Then, the ZEPCs can be rewritten as:

$$\mathcal{F}_i w_i = \sum_j \int_{\varphi^*_{ij}}^{+\infty}\left\{\left[\frac{1}{1 - r_u(q_{ij}(\varphi))} - 1\right]\frac{\tau_{ij}w_i}{\varphi}q_{ij}(\varphi)L_j - f_{ij}w_i\right\}dG(\varphi) = \frac{\gamma\kappa_1 w_i}{a}\sum_j \frac{L_j\tau_{ij}}{(\varphi^*_{ij})^{\gamma+1}}, \tag{46}$$

where $\kappa_1 = \int_0^1(\frac{1}{z} + z - 2)\frac{z+1}{z}(ze^{z-1})^{\gamma+1}dz$.[26]

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. In Assumption 3, we requires $\mathcal{C}_{HF}\mathcal{C}_{FH} > \mathcal{C}_{FF}\mathcal{C}_{HH}$ where $\mathcal{C}_{ij} = (\delta_{ij}w_i)^{-1}$.

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

According to the definition of the counterfactual equilibrium, we have:

$$\begin{cases} \delta_{ij} w_i = \dfrac{a\varphi_{ij}^*}{\tau_{ij}} \\ \dfrac{\gamma\kappa_1}{a}\dfrac{L_j \tau_{ij}}{(\varphi_{ij}^*)^{\gamma+1}} = \mathcal{F}_i \end{cases} \Rightarrow \delta_{ij} w_i = \left(\frac{a}{\tau_{ij}}\right)^{\frac{\gamma}{\gamma+1}} \left(\frac{\gamma\kappa_1 L_j}{\mathcal{F}_i}\right)^{\frac{1}{\gamma+1}}.$$

We then obtain the explicit expression of Assumption 3 and the admissible range of endogenous wage $w_F$:

$$\frac{\tau_{HH}}{\tau_{FH}} < \frac{\tau_{HF}}{\tau_{FF}}, \quad w_F \in (\underline{w_F}, \overline{w_F}) = \left(\frac{\mathcal{C}_{HH}}{\mathcal{C}_{FH}}, \frac{\mathcal{C}_{HF}}{\mathcal{C}_{FF}}\right) = \left( (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HH}}{\tau_{FH}})^{\frac{\gamma}{\gamma+1}}, (\frac{\mathcal{F}_H}{\mathcal{F}_F})^{\frac{1}{\gamma+1}} (\frac{\tau_{HF}}{\tau_{FF}})^{\frac{\gamma}{\gamma+1}} \right).$$

For Assumption 4, we can rewrite the ZEPCs in terms of $\delta_H$ and $\delta_F$ under decentralized market equilibrium as:

$$\begin{cases} \dfrac{L_H}{\tau_{HH}^{\gamma}}\dfrac{1}{\delta_H^{\gamma+1}} + \dfrac{L_F}{\tau_{HF}^{\gamma}}\dfrac{1}{\delta_F^{\gamma+1}} = \dfrac{\mathcal{F}_H (w_H)^{\gamma+1}}{\gamma\kappa_1 a^{\gamma}} \\ \dfrac{L_H}{\tau_{FH}^{\gamma}}\dfrac{1}{\delta_H^{\gamma+1}} + \dfrac{L_F}{\tau_{FF}^{\gamma}}\dfrac{1}{\delta_F^{\gamma+1}} = \dfrac{\mathcal{F}_F (w_F)^{\gamma+1}}{\gamma\kappa_1 a^{\gamma}}. \end{cases} \tag{47}$$

The assumption of the positive Jacobian determinant can then be simplified as:

$$\begin{vmatrix} \frac{L_H}{\tau_{HH}^{\gamma}}, & \frac{L_F}{\tau_{HF}^{\gamma}} \\ \frac{L_H}{\tau_{FH}^{\gamma}}, & \frac{L_F}{\tau_{FF}^{\gamma}} \end{vmatrix} > 0,$$

and thus, Assumption 4 requires:

$$(\frac{1}{\tau_{HH}\tau_{FF}})^{\gamma} - (\frac{1}{\tau_{FH}\tau_{HF}})^{\gamma} > 0. \tag{48}$$

Overall, condition (48) sufficiently restricts the exogenous parameter space such that the decentralized market equilibrium under CARA preferences is uniquely determined.

Now we can solve the cutoff productivities based on (44) and the ZEPCs (46). Note that (46) permits reducing the set of cutoff productivities to those for the domestic market in each country. However, (44) 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

$$(\varphi_{ii}^*)^{\gamma+1} = \frac{\gamma \kappa_1 L_i \tau_{ii} \left[ (\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1 \right]}{a \mathcal{F}_i \left[ (\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{ji} w_j}{\tau_{ii} w_i})^\gamma \right]}, \tag{49}$$

and for exporting sales of

$$(\varphi_{ij}^*)^{\gamma+1} = (\frac{\tau_{ij} w_i}{\tau_{jj} w_j} \varphi_{jj}^*)^{\gamma+1} = \frac{\gamma \kappa_1 L_j \tau_{ij} \left[ 1 - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma \right]}{a \mathcal{F}_i \left[ (\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma \right]} \tag{50}$$

The resource constraint can be simplified as:

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

where $\kappa_3 = \int_0^1 (z e^{z-1})^{\gamma+1} \frac{1-z^2}{z} dz$ 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 $M_i = \frac{L_i}{(\gamma+1)\mathcal{F}_i}$.

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

$$M_i \int_{\varphi_{ij}^*}^{+\infty} \frac{1}{1 - r_u(q_{ij}(\varphi))} \frac{\tau_{ij} w_i}{\varphi} q_{ij}(\varphi) L_j dG(\varphi) = M_j \int_{\varphi_{ji}^*}^{+\infty} \frac{1}{1 - r_u(q_{ji}(\varphi))} \frac{\tau_{ji} w_j}{\varphi} q_{ji}(\varphi) L_i dG(\varphi). \tag{51}$$

With (45) and solutions of $q_{ij}(\varphi)$ and $M_i$, we can rewrite (51) as

$$\frac{L_i w_i}{\mathcal{F}_i} \frac{\tau_{ij} L_j}{(\varphi_{ij}^*)^{\gamma+1}} = \frac{L_j w_j}{\mathcal{F}_j} \frac{\tau_{ji} L_i}{(\varphi_{ji}^*)^{\gamma+1}}.$$

With (50), we obtain the implicit solution for the relative wage ratio as

$$\frac{w_j}{w_i} = \frac{L_i}{L_j} \frac{(\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{jj} w_j}{\tau_{ij} w_i})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}{(\frac{\mathcal{F}_i w_i}{\mathcal{F}_j w_j})(\frac{\tau_{ii} w_i}{\tau_{ji} w_j})^\gamma - (\frac{\tau_{jj}\tau_{ii}}{\tau_{ji}\tau_{ij}})^\gamma}.$$

**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 $\mathcal{W}$ and obtain

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\left[1-\mathcal{W}(e\dfrac{\varphi_{ij}^*}{\varphi})\right] \\ \varphi_{ij}^* = \dfrac{\tau_{ij}}{\tau_{ii}}\varphi_{ii}^*. \end{cases} \tag{52}$$

Using identical definitions of $z_{ij}$ and $\kappa_1$ as in the decentralized equilibrium, we can simplify the ZEPC as

$$\mathcal{F}_i w_i = \sum_j \int_{\varphi_{ij}^*}^{+\infty} \left\{\left[\frac{1}{1-r_u(q_{ij}(\varphi))}-1\right]\frac{\tau_{ij}w_i}{\varphi}q_{ij}(\varphi)L_j - f_{ij}w_i\right\} dG(\varphi) = \sum_j \frac{L_j\gamma\tau_{ij}w_i\kappa_1}{a(\varphi_{ij}^*)^{\gamma+1}}. \tag{53}$$

In this case, since the demand shifter $\delta_i^{cmkt}$ is indexed by the origin, we can solve the cutoff productivity of origin $i$ based on the ZEPC (53) and the cutoff relation (52) in country $i$. Note that, in contrast to the decentralized market equilibrium, the cutoff productivity $\varphi_{ij}^*$ is proportional to $\varphi_{ii}^*$ rather than to $\varphi_{jj}^*$ in the centralized market equilibrium. Therefore, a country's own resource constraint alone pins down the cutoff productivity $\varphi_{ii}^*$. Then, $\forall i \neq j$, we can obtain explicit solutions for the productivity cutoffs in country $i$:

$$\begin{cases} (\varphi_{ii}^*)^{\gamma+1} = \dfrac{\tau_{ii}\gamma\kappa_1\left[L_i + L_j(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]}{a\mathcal{F}_i} \\ (\varphi_{ij}^*)^{\gamma+1} = \dfrac{\tau_{ij}\gamma\kappa_1\left[L_i(\frac{\tau_{ij}}{\tau_{ii}})^\gamma + L_j\right]}{a\mathcal{F}_i}. \end{cases} \tag{54}$$

With the resource constraint, we can similarly define $\kappa_3$ and obtain the solution for the masses of entrants as $M_i = \frac{L_i}{\mathcal{F}_i(\gamma+1)}$.

**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

$$\begin{cases} q_{ij}(\varphi) = \dfrac{1}{a}\ln(\dfrac{\varphi}{\varphi_{ij}^*}) & (55) \\ \varphi_{ij}^* = \dfrac{\tau_{ij}}{\tau_{ii}}\varphi_{ii}^*. & (56) \end{cases}$$

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

$$\sum_j L_j \int_{\varphi_{ij}^*}^{+\infty} u(q_{ij}(\varphi))dG(\varphi) = \lambda_i \left\{ \sum_j \left[ \int_{\varphi_{ij}^*}^{+\infty} \frac{\tau_{ij}w_i}{\varphi} q_{ij}(\varphi)L_j dG(\varphi) \right] + \mathcal{F}_i w_i \right\}. \tag{57}$$

With (55), we can rewrite the LHS of (57) 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 (55), the RHS of (57) becomes:

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

With (56), equating the LHS to the RHS obtains the solutions for the productivity cutoffs:

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

The resource constraints yield:

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

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

## A.11 Proof of Lemma 4

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

$$\left[\frac{(\varphi_{ii}^*)^{dmkt}}{(\varphi_{ii}^*)^{cmkt}}\right]^{\gamma+1} = \frac{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1\right]}{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{ji} w_j}{\tau_{ii} w_i})^\gamma\right]\left[1 + \frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right]} \gtrless 1,$$

which can be rewritten as:

$$\frac{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - 1\right]}{\left[(\frac{\tau_{ij}\tau_{ji}}{\tau_{ii}\tau_{jj}})^\gamma - (\frac{\mathcal{F}_j w_j}{\mathcal{F}_i w_i})(\frac{\tau_{ji} w_j}{\tau_{ii} w_i})^\gamma\right]} - \left[1 + \frac{L_j}{L_i}(\frac{\tau_{ii}}{\tau_{ij}})^\gamma\right] \gtrless 0.$$

The comparison can be further simplified to

$$(\frac{L_j}{L_i})(\frac{\tau_{ii}}{\tau_{ij}})^\gamma \left\{\frac{L_i}{L_j}\frac{\left[(\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ii} w_i})^\gamma - (\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{ji} w_j})^\gamma\right]}{\left[(\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{jj} w_j})^\gamma - (\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ij} w_i})^\gamma\right]} - 1\right\} \gtrless 0$$

$$\Leftrightarrow L_i\left[(\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ii} w_i})^\gamma - (\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{ji} w_j})^\gamma\right] \gtrless L_j\left[(\frac{1}{\mathcal{F}_j w_j})(\frac{1}{\tau_{jj} w_j})^\gamma - (\frac{1}{\mathcal{F}_i w_i})(\frac{1}{\tau_{ij} w_i})^\gamma\right]. \quad (58)$$

Therefore, we show that, if the LHS of (58) 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. □

## A.12 Proof of Lemma 5

*Proof.* 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 $A = \int_0^1 z^{\gamma+1}\left(e^{z-1}\right)^{\gamma+1} dz$. 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 $\left[\frac{(\varphi_{ii}^*)^{cmkt}}{(\varphi_{ii}^*)^{opt}}\right]^{\gamma+1} = \left[\frac{(\varphi_{ij}^*)^{cmkt}}{(\varphi_{ij}^*)^{opt}}\right]^{\gamma+1} < 1.$ □

## A.13 Proof of Lemma 6

*Proof.* 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 $\mathcal{W}$, we can rewrite $q_{ij}^{cmkt}(\varphi) = \frac{1}{a}\ln\left[\frac{\varphi}{(\varphi_{ij}^*)^{cmkt}}\mathcal{W}(e\frac{(\varphi_{ij}^*)^{cmkt}}{\varphi})\right]$.

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

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

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

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. □

# B Online Appendix

## B.1 Discussions on Sufficiency of the Optimal Conditions in the Planner's Problem

*Proof.* We will be focusing on the case where the equilibrium in the centralized market's and the planner's problems is sufficiently characterized by the unique solutions presented in Propositions 2 and 3. To illustrate this, we will use the social planner's problem in Proposition 2, while the discussions concerning Proposition 3 will be similar.

Without explicit constraints on exogenous parameters, we consider the setup exhibiting (i) preferences $u(q)$ that possibly exhibit an upper bound of production, $\overline{q} \equiv \min\{q \geqslant 0 \text{ s.t. } r_u(q) = 1\}$, (ii) the utility aggregator does not satisfy the Inada conditions. Specifically, the aggregate utility gains for the representative consumer in destination $j$ are the sum of their utility gains from consuming domestic and imported goods, $U_j = U_{Hj} + U_{Fj}$. With this setup, our model can cover a large class of demand structures and incorporate the autarky case, which is absent in traditional trade models, in the decentralized market equilibrium.

However, this general setup comes at a cost: we can not rule out the possibility of corner solutions to the planner's problem. First, the unique solution we show in Proposition 2 could generate a corner solution for some outcome variables. That is, if the unique solution $\lambda_i^{opt} \in [\frac{L_{\underline{j}}\overline{B}^s}{f_{i\underline{j}}w_i}, \frac{L_{\overline{j}}\overline{B}^s}{f_{i\overline{j}}w_i})$, the corresponding $q_{i\underline{j}}(\varphi)$=0 and $\varphi^*_{i\underline{j}} = +\infty$ and, thus, not all FOCs are satisfied. Second, we cannot rule out the possibility that the utility outcomes at the corners are greater than those at the unique solutions in Proposition 2. These cases are documented in Kokovin et al. (2022), who show that a small degree of trade openness for two symmetric countries can be harmful, and a social planner would increase welfare by prohibiting trade.

There are several potential methods to deal with this issue. With the specification of demand structure and productivity distribution, the most straightforward and harmless way is to numerically/analytically compare the outcomes under the unique (partial) interior solution with those at the corner. Another method is to employ a similar general aggregator as conditions (2) in Behrens et al. (2020) satisfying the Inada conditions. More specifically, one can assume that $U_j(U_{Hj}, U_{Fj})$ is satisfying $\frac{\partial U_j}{\partial U_{ij}} = \gamma_{ij} h_{ij}(U_{ij})\, h(U_j)$, where $\lim_{U_{ij}\to 0} h_{ij}(U_{ij}) = +\infty$, $\lim_{U_{ij}\to\infty} h_{ij}(U_{ij}) = 0$, $\forall i, j = H, F$. These conditions ensure countries to have positive trade flows and restrict the planner's allocations from any corner solutions. □