Theoretical Economics Letters, 2017, 7, 263-293
http://www.scirp.org/journal/tel
ISSN Online: 2162-2086
ISSN Print: 2162-2078

Ricardo Revisited: Benefits from Trade and the Role of Non-Convex Technologies

Jean-Paul Chavas, Taylor Hall

University of Wisconsin, Madison, WI, USA
Email: jchavas@wisc.edu

How to cite this paper: Chavas, J.-P. and Hall, T. (2017) Ricardo Revisited: Benefits from Trade and the Role of Non-Convex Technologies. *Theoretical Economics Letters*, **7**, 263-293.
https://doi.org/10.4236/tel.2017.72022

Received: September 19, 2016
Accepted: February 25, 2017
Published: February 28, 2017

Abstract

This paper explores the aggregate gains from trade with a focus on the role of non-convexity. After reviewing the example presented by Ricardo, we develop a general equilibrium model of trade under non-convex technologies and heterogeneous firms. The model is used to evaluate aggregate efficiency, with a focus on the case where trade restrictions are the only source of inefficiency. The analysis allows for non-linear pricing which becomes an integral part of efficiency under non-convex technologies. We establish bounds on the gains from trade. We show that the gains from trade are non-negative and that they tend to be small under convexity but can become large under non-convexity. This indicates that the search for larger gains from trade needs to be associated with non-convex technologies. Implications of our analysis for the benefits of globalization are discussed.

Keywords

Ricardo, Globalization, Gains from Trade, Non-Convexity

1. Introduction

Economic analysis has identified that trade opportunities generate efficiency gains. The argument was first presented by Adam Smith [1] and then refined by Ricardo [2] who demonstrated how countries can benefit from trade. The associated aggregate welfare gains have been used to support trade liberalization policies. Yet, while economists agree that there are efficiency gains from trade (e.g., Samuelson [3]), a question remains: How large are these gains? The empirical measurements of aggregate gains from trade have typically been relatively small. For example, Arkolakis *et al.* ([4], p. 95) estimated that the welfare gains from trade for the US have ranged from 0.7% to 1.4% of income. And Ossa [5] evaluated the aggregate welfare effects from trade liberalization to range from 0.5% and 2.4%. These small percentages are somewhat problematic for economists who argue in support of trade liberalization policies. This has stimulated the search for new trade models that could generate larger gains from trade. In particular, the roles of economies of scale, product differentiation, imperfect competition and firm heterogeneity have been examined (e.g., Krugman [6]; Melitz [7]; Bernard *et al.* [8]; Balistreri *et al.* [9]; Melitz and Trefler [10]; Caliendo and Rossi-Hansberg [11]; Melitz and Redding [12]; Edmond *et al.* [13]). Yet, Arkolakis *et al.* [4] argue that these new inquiries have not had much of an impact on the aggregate welfare gains from trade.[1] This raises the questions: Are there conditions under which the aggregate gains from globalization would be large? What are these conditions? The objective of this paper is to explore these issues in the context of a Ricardian model and to provide answers to these questions.

Ricardo [2] was the first to formalize the economic arguments in favor of free trade. Following Eaton and Kortum ([14] [15]), we motivate our analysis using Ricardo's example. We first examine the case of two countries and two goods and present empirical results from Ricardo's example. The empirical exercise relies on a flexible representation of technology using a constant elasticity of transformation (CET) specification. It allows for an evaluation of the separate effects of returns to scale and convexity on the gains from trade. The empirical results highlight three points: 1) in the two-country two-goods case, economies of scale have no effect on the gains from trade; 2) the benefits from trade tend to be small under convex technologies; and 3) non-convex technologies can generate large aggregate gains from trade. This last point suggests a need for a refined analysis of the economics of trade under non-convexity.

Introducing non-convexity in the welfare evaluation of trade is challenging as standard welfare theorems establishing linkages between competitive market equilibrium and Pareto efficiency apply only under convexity (e.g., Debreu [16]; Brown [17]). We address this challenge in the context of a general equilibrium model of an economy, allowing for non-convex technologies and heterogeneous firms. The model provides a basis to evaluate aggregate efficiency, with a focus on the case where trade restrictions are the only source of inefficiency.[2] The analysis of non-convex technologies builds on Chavas and Briec [18]. By allowing for firm entry and exit and heterogeneous technologies across firms, our general equilibrium model captures the effects of such factors on aggregate productivity (as argued by Melitz [7], Bernard *et al.* [8], Melitz and Trefler [10] and Melitz and Redding [12]). We develop a dual characterization of aggregate efficiency under non-convexity and trade restrictions. The analysis allows for non-linear pricing which becomes an integral part of efficiency under non-convex

[1]Notable exceptions include Melitz and Redding [12] and Edmond *et al.* [13]. Melitz and Redding [12] argue that, in the presence of firm heterogeneity, endogenous firm selection can contribute to increasing the aggregate welfare gain from trade. And Edmond *et al.* [13] identify scenarios where the aggregate gains from trade can be large when there are important initial inefficiencies due to imperfect competition.

[2]As such, this paper does not examine the role of oligopolistic competition and its effects on the gains from trade. Yet, we will examine the role of non-convexity and discuss its implications for pricing in Section 4.

technologies. This dual characterization is used to evaluate aggregate gains from trade. In this context, we establish bounds on the gains from trade. We show that the gains from trade are non-negative and that they tend to increase with any relaxation in trade restrictions. While these are well-known results under convexity, our analysis shows that they remain valid under non-convexity. We also show how the gains from trade are closely linked with the properties of price-dependent demand functions for imports. This extends the analysis presented by Arkolakis *et al.* [4] to situations of non-convexity. Most importantly, our analysis shows that the aggregate gains from trade tend to be small under convexity, but that they can become large under non-convexity. This indicates that the search for larger gains from globalization needs to be associated with non-convex technologies.

The paper is organized as follows. Section 2 motivates the analysis, relying on the example discussed by Ricardo [2]. Section 3 presents a general equilibrium model of trade. Section 4 discusses the effects of trade restrictions and their impacts on pricing under non-convexity. Section 5 investigates the implications for the aggregate benefit from trade. Finally, Section 6 concludes.

2. Motivations

Adam Smith [1] identified the role of division of labor and examined its effects on productivity. He illustrated his arguments using the making of pins as an example. First, he considered the case of a "workman not educated in this business" (Smith, [1], p. 4). He estimated that, when working alone, this workman could make up to 20 pins a day. Smith also considered a team of ten workers making pins, each one specializing in distinct successive tasks involved in the process of making pins (e.g., drawing the wire, straightening it, cutting it, etc.). Smith ([1], p. 4) claimed to have "seen a small factory of this kind." He assessed that these ten men could make up to 48,000 pins a day, or 4800 pins per person per day (Smith, 1776, p. 4). Compared to the first scenario (20 pins per day), this amounts to labor productivity being multiplied by a factor of 240, or a 23900 percent increase! Using this example, Smith argued that the division of labor can contribute to large increases in the productive power of labor.

Where do these productivity gains come from? Smith ([1], pp. 6-8) argues that there are three contributing factors: 1) "improved dexterity" associated with learning; 2) saving in time lost switching from one task to the next; and 3) the use of machines that can make human labor more productive. Note that the second factor amounts to introducing fixed costs in the analysis (since the time lost switching between tasks does not contribute to any output). To the extent that fixed costs imply non-convexity technologies, this is relevant in our analysis of the role of non-convexity. Finally, by emphasizing the importance of organization and learning, Smith ([1], p. 12) argued that productive differences among individuals are not so much the cause but the effect of the division of labor.

Smith ([1], p. 13-17) also argued that the division of labor is "limited by the extent of the market." He stressed by the benefits of specialization can be obtained only in the presence of exchange. This indicates that gains from trade and the benefits of specialization are closely linked (Stigler [19]). As noted by Caliendo and Rossi-Hansberg [11] and Becker and Murphy [20], the degree of specialization depends on the cost of coordination among workers/firms.

While Adam Smith suggested that specialization along with trade can improve economic efficiency, his arguments were not made analytically. Ricardo [2] provided the analytical arguments. His illustrative example was to consider the case of two countries: England (E) and Portugal (P); and two goods: cloth and wine. He assumed that the technology differs across countries, and that labor is the only input and is not traded. In this context, he investigated the gains from trade in outputs (cloth and wine) between England (E) and Portugal (P). Below, after reproducing Ricardo's example, we examine the factors affecting the gains from trade.

Assume that the production technology for cloth (C) and wine (W) in the i-th country is a Constant Elasticity of Transformation (CET) technology represented by (Powell and Gruen [21])

$$L_i \geq k_i \left[\alpha_i C_i^{1-\rho} + (1-\alpha_i) W_i^{1-\rho} \right]^{s/(1-\rho)} \tag{1}$$

where L_i is the amount of labor available in country i, k_i is a productivity parameter, α_i is a parameter between 0 and 1, $1/\rho$ is the Allen elasticity of transformation between cloth (C_i) and wine (W_i), and $1/s$ is the scale elasticity, $i \in \{E, P\}$. The specification (1) is flexible in the sense that it allows for an arbitrary Allen Elasticity of transformation $1/\rho$, and an arbitrary scale elasticity $1/s$. Note that a convex technology would restrict the elasticity of transformation to be non-positive (with $1/\rho \leq 0$) and returns to scale to be non-increasing (with a scale elasticity satisfying $0 < 1/s \leq 1$).

Let consumer preferences in the i-th country be represented by the utility function $u_i(c_i, w_i)$, where c_i is the consumption of cloth and w_i is the consumption of wine in country i, $i \in \{E, P\}$. Assuming that $u_i(c_i, w_i)$ is increasing and quasi-concave in (c_i, w_i), $i \in \{E, P\}$, consider the following allocation under free trade

$$\begin{aligned} \text{Max}_{c,w,C,W} \Big\{ & \sum_i u_i(c_i, w_i) : \sum_i c_i \leq \sum_i C_i, \sum_i w_i \leq \sum_i W_i; c_i \geq 0, w_i \geq 0, C_i \geq \varepsilon, W_i \geq \varepsilon, \\ & L_i = k_i \left[\alpha_i C_i^{1-\rho} + (1-\alpha) W_i^{1-\rho} \right]^{s/(1-\rho)}, i \in \{E, P\} \Big\}, \end{aligned} \tag{2a}$$

where $\varepsilon > 0$ is an arbitrarily small number.[3] The solution to (2a) gives a Pareto efficient allocation. And consider the corresponding allocation under autarky

$$\begin{aligned} \text{Max}_{c,w,C,W} \Big\{ & \sum_i u_i(c_i, w_i) : c_i \leq C_i, w_i \leq W_i; c_i \geq 0, w_i \geq 0, C_i \geq \varepsilon, W_i \geq \varepsilon, \\ & L_i = k_i \left[\alpha_i C_i^{1-\rho} + (1-\alpha) W_i^{1-\rho} \right]^{s/(1-\rho)}, i \in \{E, P\} \Big\}. \end{aligned} \tag{2b}$$

[3]We introduce $\varepsilon > 0$ in (2a) to guarantee the existence of a solution to (2a). This is particularly relevant under non-convexity where $\varepsilon > 0$ guarantees the compactness of the feasible set in (2a) when $\rho > 0$.

For illustration purpose, we assume that $u_i(c_i, w_i) = \beta_i \ln(c_i) + (1-\beta_i)\ln(w_i)$, where $\beta_i \in (0,1)$. Throughout the analysis, we hold consumer preferences constant. We simulated allocations obtained under free trade and autarky by solving Equations (2a) and (2b) under alternative scenarios. Our first scenario is the one considered by Ricardo ([2], p. 135). Ricardo implicitly assumed a linear technology under constant returns to scale. This corresponds to $\rho = 0$ and $s = 1$ in (1). Ricardo ([2], p. 135) considered each country facing different technologies and different factor endowments. He assumed that England has 220 workers, Portugal has 170 workers, and that it takes 100 workers (90 workers) to produce 1 unit of cloth in England (in Portugal) and 120 workers (80 workers) to produce one unit of wine in England (in Portugal). In the context of Equation (1), this corresponds to choosing $k_E = 220, k_P = 170, \alpha_E = 5/11$, and $\alpha_P = 9/17$. Finally, to replicate Ricardo's analysis, we assume that consumer preferences satisfy $\beta_E = 5/11$ and $\beta_P = 9/17$. Throughout our analysis, we hold the parameters (k, α, β) constant. However, we explore the implications of changing the parameters (ρ, s).

We start with the Ricardo scenario, where $\rho = 0$ and $s = 1$. Corresponding allocations are presented in the first row of **Table 1** and in **Figure 1**. This replicates Ricardo's analysis: under autarky, England and Portugal each produces 1 unit of wine and 1 unit of cloth; and under free trade, England specializes in the production of cloth (2.2 units, represented by point B in **Figure 1**) while Portugal specializes in the production of wine (2.125 units, represented by point D in **Figure 1**). At the aggregate, switching from autarky to free trade, total production increases from 2 to 2.2 units of cloth (+10 percent), and from 2 to 2.125 units of wine (+6.25 percent). It is represented by a move from point A to point T in **Figure 1**. This increased production is matched by an equivalent increase in aggregate consumption.[4] Ricardo's example illustrates that free trade does generate positive gains from specialization.

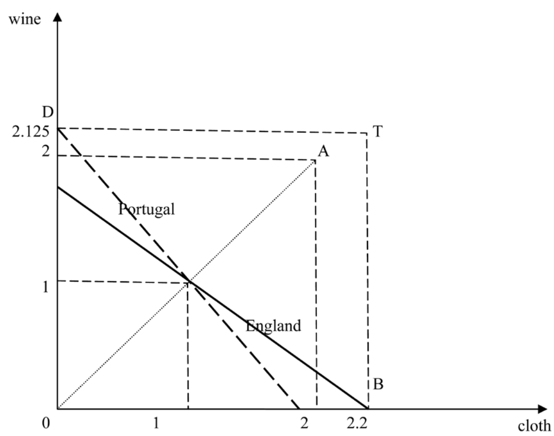

Figure 1. Gains from trade in the Ricardo's example ($\rho = 0$).

[4]Note that our simulation results are exact replications of Ricardo's analysis under autarky; and they are exact replications of production and aggregate consumption under free trade. However, note that the distribution of consumption under free trade depends on consumer preferences. Since Ricardo did not identify the nature of consumer preferences, our simulated distribution of consumption under free trade is not directly comparable to Ricardo's analysis.

Table 1. Evaluating the gains from trade: The case of Wine (W) and Cloth (C) Traded between England (E) and Portugal (P).[a]

Technology	Trade policy	Autarky: Prod	Autarky: Cons	Autarky: Wine Price[b]	Autarky: Total value[c] (VA)	Free Trade: Prod	Free Trade: Cons	Free Trade: Wine Price[b]	Free Trade: Total value[c] (VT)	Gains from Trade = VT/VA
Convex technology	$\rho = 0$ $s = 1$ (Ricardo's example)	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 2.2 W_P: 2.125 C_P: 0	w_E: 1.141 c_E: 1.016 w_P: 0.984 c_P: 1.184	E: 1.069 P: 1.069	4.472	1.093
Convex technology	$\rho = -0.1$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0.384 C_E: 1.690 W_P: 1.666 C_P: 0.365	w_E: 1.101 c_E: 0.949 w_P: 0.950 c_P: 1.106	E: 1.034 P: 1.034	4.177	1.021
Convex technology	$\rho = -0.3$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0.766 C_E: 1.261 W_P: 1.252 C_P: 0.758	w_E: 1.083 c_E: 0.933 w_P: 0.935 c_P: 1.086	E: 1.033 P: 1.033	4.104	1.003
Convex technology	$\rho = -0.5$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0.857 C_E: 1.158 W_P: 1.153 C_P: 0.853	w_E: 1.080 c_E: 0.929 w_P: 0.931 c_P: 1.082	E: 1.033 P: 1.033	4.089	1.0001
Non-convex technology	$\rho = 0.1$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 2.401 W_P: 2.310 C_P: 0	w_E: 1.240 c_E: 1.109 w_P: 1.070 c_P: 1.292	E: 1.073 P: 1.073	4.881	1.193
Non-convex technology	$\rho = 0.3$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 3.081 W_P: 2.932 C_P: 0	w_E: 1.574 c_E: 1.423 w_P: 1.358 c_P: 1.658	E: 1.085 P: 1.085	6.262	1.531
Non-convex technology	$\rho = 0.5$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 4.787 W_P: 4.468 C_P: 0	w_E: 2.399 c_E: 2.211 w_P: 2.069 c_P: 2.576	E: 1.106 P: 1.107	9.731	2.380
Non-convex technology	$\rho = 0.7$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 12.32 W_P: 11.02 C_P: 0	w_E: 5.914 c_E: 5.693 w_P: 5102 c_P: 6.630	E: 1.155 P: 1.155	25.05	6.125
Non-convex technology	$\rho = 0.9$ $s = 1$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 229.6 W_P: 176.1 C_P: 0	w_E: 94.52 c_E: 106.1 w_P: 81.54 c_P: 123.5	E: 1.347 P: 1.347	466.7	114.1
	Scale effects: $\rho = 0$, any $s > 0$	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.20 P: 0.889	4.089	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.069 P: 1.069	4.472	1.093

[a]The results are obtained as solutions of the optimization problems (2a) and (2b) under alternative scenarios. The analysis was done with $\varepsilon = 0.0001$. [b]The price of wine is measured as the consumer's marginal rate of substitution between wine and cloth, after normalizing the price of cloth to 1. [c]Note our total values differ from the results reported in Ricardo's analysis. The reason is that Ricardo (1717, p. 135) assumed that the prices of wine and cloth are both equal to 1 under free trade, while our simulated price of wine differs from 1. This affects consumption under free trade as well as the total value of goods.

The above discussion indicates that the gains from trade are somewhere between 6.25 and 10 percent. A more precise measure can be obtained by evaluating the total value of all goods (either produced or consumed). Treating cloth as the numeraire good (with a unit price), define the price of wine to be the consumer marginal rate of substitution between wine and cloth. This provides a basis to evaluate the total value of goods (cloth and wine) across scenarios. In the context of the Ricardo scenario, we find that the total value of goods under autarky is VA = 4.089 and the total value of goods under free trade is VT = 4.472. As reported in Table 1, VT/VA = 1.093. It means that, under the Ricardo scenario, the gains from trade and specialization correspond to a 9.3 percent increase in the total value of goods.

Comparing Smith's analysis with Ricardo's analysis creates a significant puzzle: the gains from trade in Ricardo's scenario (+9.3 percent) are much smaller than the numbers presented by Smith (who reported a 23,900 percent increase in productivity). Although these two numbers are not exactly comparable, the magnitude of the difference is huge. Why? Is the Ricardo's scenario generating benefits from trade that are "too small"? Or are the productivity gains from specialization reported by Smith "unrealistically high"? And is there a way to reconcile this large difference?

To help answer these questions, we proceed examining how the gains from trade can change under alternative scenarios. As noted above, the Ricardo scenario assumed $\rho = 0$ and $s = 1$, which is at the boundary of the region where the technology is convex (corresponding to $\rho \leq 0$ and $0 < 1/s \leq 1$). Below, we depart from the Ricardo scenario in three different ways: 1) when the technology becomes strictly convex in outputs (with $\rho < 0$); 2) when the technology becomes non-convex in outputs (with $\rho > 0$); and 3) when we allow returns to scale to vary (with $s \neq 1$).

First, Table 1 reports a set of scenarios where the possibilities of substitution between wine and cloth decline as ρ decreases from 0. Holding $s = 1$, this corresponds to technologies that are convex, but becoming "more convex" in the output space. Less substitution between wine and cloth implies a lower incentive for producers to specialize. While the autarky scenario remains unchanged, the free trade scenario exhibits less specialization and lower benefit from trade. This is illustrated in Figure 2 where, under free trade, England produces at point B', Portugal produces at point D', and aggregate production increases from A to T'. The gains from trade remain positive. But with less specialization, they decline compared to the Ricardo scenario (as point T' is lower and to the left of point T in Figure 2). The effects of reduced substitution possibilities between wine and cloth are shown in Table 1, where the ratio VT/VA declines from 1.093 under the Ricardo scenario $(\rho = 0)$ to 1.021 when $\rho = -0.1$, and to 1.003 when $\rho = -0.3$. It means that a small change in ρ from 0 to −0.1 implies a rapid decline in the gains from trade from +9.3 percent (under the Ricardo scenario) to +2.1 percent. And a further change in ρ to −0.3 would imply a decline in the gains from trade to just +0.3 percent. In other words, introducing strict convexity

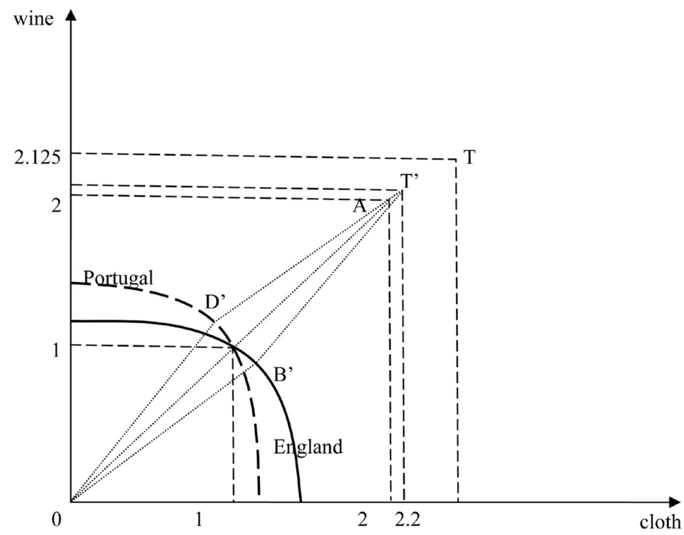

Figure 2. Gains from trade under a convex technology.

in outputs tends to generate small benefits from trade (as the incentive to specialize falls sharply). As strict convexity reflects diminishing marginal productivity, this means that models that mix trade liberalization with diminishing marginal productivity are unlikely to find large gains from trade. This is a sobering result in two ways. First, it suggests a fundamental inconsistency between two views commonly held by economists: 1) that diminishing marginal productivity is a basic characteristic of technology and resource scarcity; and 2) that the benefits from trade liberalization can be large. We reflect on these important issues below. Second, if diminishing marginal productivity implies small gains from trade, then we are no closer to offering an explanation for the large difference between Smith's and Ricardo's analyses(as noted above).

Second, Table 1 also reports a set of scenarios where the technology becomes non-convex in outputs as ρ increases from 0 (again holding $s = 1$). Here the non-convexity in outputs implies a stronger incentive for producers to specialize. While again the autarky scenario remains unchanged, the free trade scenario exhibits large benefit from specialization and trade. This is illustrated in Figure 3 where, under free trade, England produces at point B", Portugal produces at point D", and aggregate production increases from A to T". Now, the gains from trade can become very large. This is shown in Figure 3 where T" is higher and to the right of point T (corresponding to the Ricardo scenario). These effects are documented in Table 1, where the ratio VT/VA increases from 1.093 under the Ricardo scenario $(\rho = 0)$ to 1.193 when $\rho = 0.1$, to 1.531 when $\rho = 0.5$ and to 114.1 when $\rho = 0.9$. It shows that a small change in ρ from 0 to 0.1 implies a significant rise in the gains from trade from +9.3 percent (under the Ricardo scenario) to +19.3 percent. And this rise can be very rapid: it reaches +11,310 percent when $\rho = 0.9$. In other words, introducing non-convexity in outputs can generate very large increases in the benefits from trade and specialization. This is a key insight for our analysis: introducing a non-convex technology in Ricardo's analysis can generate gains from trade that are broadly consis-

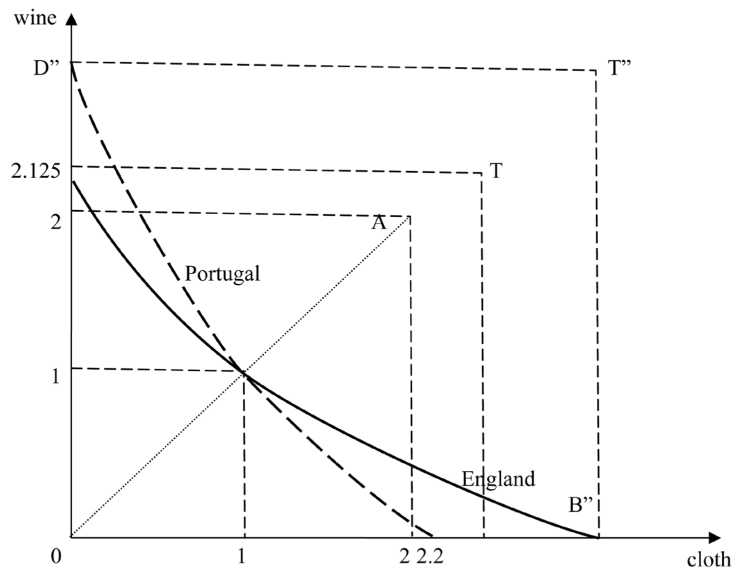

Figure 3. Gains from trade under a non-convex technology.

tent with Smith's analysis (reporting very large productivity gains from specialization). In other words, allowing for non-convex technology can help reconcile Smith's and Ricardo's analyses. But as discussed next, it is a particular type of non-convexity that is needed.

Third, we explore the implications of scale economies. The last row in Table 1 reports the effects of changing the scale elasticity s on the gains from trade (this time holding $\rho = 0$). This covers the case where the technology exhibits decreasing returns to scale (when $0 < 1/s < 1$) as well as increasing returns to scale (when $1/s > 1$). Table 1 reports that, for a given ρ, changing s has no effect on the gains from trade: VT/VA is the same for any $s > 0$. This result has an intuitive explanation. Applied at the country level, Ricardo's analysis (as well as our model in (2a)-(2b)) assumes that labor is fixed and non-traded. This effectively fixes the scale of operation in each country. It prevents us from observing any effect of scale on the incentives to specialize and thus on the benefits from trade. This result holds whether the technology exhibits decreasing returns to scale $(0 < 1/s < 1)$ or increasing returns to scale $(s > 1)$. In this latter case, note that the technology would be non-convex.

This generates two key insights. First, in general, assuming increasing returns is not sufficient to generate large gains from trade. Second, non-convexity can contribute to large gains from trade only to the extent that it applies to the production of traded goods.

While Table 1 compares free trade with autarky, it is also of interest to evaluate the effects of partial trade restrictions. Using the Ricardo example, we evaluate the effects of Portugal implementing a unilateral import ban. This is done by solving the optimization problem in (2a) subject to the additional restriction that imports into Portugal are non-positive. The simulated results are reported in Table 2. Table 2 shows the effects of an import ban imposed by Portugal compared to free trade. A comparison with Table 1 shows that, as expected, the

Table 2. Evaluating the effects of trade restrictions: The Case of Wine (W) and Cloth (C) Traded between England (E) and Portugal (P).

Technology	Trade policy	Restricted Trade (Import ban imposed by Portugal) Prod	Cons	Wine Price[b]	Total value (VR)	Free Trade Prod	Cons	Wine Price[b]	Total value (VT)	Benefit from Trade = VT/VR
Convex technology	$\rho = 0$ $s = 1$ (Ricardo's example)	W_E: 0.934 C_E: 1.080 W_P: 1.077 C_P: 0.931	w_E: 1.080 c_E: 1.080 w_P: 0.931 c_P: 0.931	E: 1.20 P: 0.889	4.089	W_E: 0 C_E: 2.2 W_P: 2.125 C_P: 0	w_E: 1.141 c_E: 1.016 w_P: 0.984 c_P: 1.184	E: 1.069 P: 1.069	4.472	1.093
	$\rho = -0.1$ $s = 1$	W_E: 0.945 C_E: 1.066 W_P: 1.065 C_P: 0.942	w_E: 1.079 c_E: 1.066 w_P: 0.931 c_P: 0.942	E: 1.186 P: 0.90	4.086	W_E: 0.384 C_E: 1.690 W_P: 1.666 C_P: 0.365	w_E: 1.101 c_E: 0.949 w_P: 0.950 c_P: 1.106	E: 1.034 P: 1.034	4.177	1.022
	$\rho = -0.3$ $s = 1$	W_E: 0.959 C_E: 1.049 W_P: 1.049 C_P: 0.956	w_E: 1.078 c_E: 1.049 w_P: 0.930 c_P: 0.956	E: 1.168 P: 0.914	4.083	W_E: 0.766 C_E: 1.261 W_P: 1.252 C_P: 0.758	w_E: 1.083 c_E: 0.933 w_P: 0.935 c_P: 1.086	E: 1.033 P: 1.033	4.104	1.005
	$\rho = -0.5$ $s = 1$	W_E: 0.967 C_E: 1.039 W_P: 1.039 C_P: 0.965	w_E: 1.077 c_E: 1.039 w_P: 0.929 c_P: 0.965	E: 1.158 P: 0.923	4.082	W_E: 0.857 C_E: 1.158 W_P: 1.153 C_P: 0.853	w_E: 1.080 c_E: 0.929 w_P: 0.931 c_P: 1.082	E: 1.033 P: 1.033	4.089	1.002
Non-convex technology	$\rho = 0.1$ $s = 1$	W_E: 0.917 C_E: 1.10 W_P: 1.096 C_P: 0.916	w_E: 1.081 c_E: 1.10 w_P: 0.932 c_P: 0.916	E: 1.222 P: 0.873	4.093	W_E: 0 C_E: 2.401 W_P: 2.310 C_P: 0	w_E: 1.240 c_E: 1.109 w_P: 1.070 c_P: 1.292	E: 1.073 P: 1.073	4.881	1.192
	$\rho = 0.3$ $s = 1$	W_E: 0.833 C_E: 1.212 W_P: 1.185 C_P: 0.844	w_E: 1.083 c_E: 1.212 w_P: 0.934 c_P: 0.844	E: 1.343 P: 0.803	4.126	W_E: 0 C_E: 3.081 W_P: 2.932 C_P: 0	w_E: 1.574 c_E: 1.423 w_P: 1.358 c_P: 1.658	E: 1.085 P: 1.085	6.262	1.518
	$\rho = 0.5$ $s = 1$	W_E: 0 C_E: 4.787 W_P: 1.952 C_P: 0.418	w_E: 1.048 c_E: 4.787 w_P: 0.904 c_P: 0.418	E: 5.481 P: 0.411	6.009	W_E: 0 C_E: 4.787 W_P: 4.468 C_P: 0	w_E: 2.399 c_E: 2.211 w_P: 2.069 c_P: 2.576	E: 1.106 P: 1.107	9.731	1.619
	$\rho = 0.7$ $s = 1$	W_E: 0 C_E: 12.32 W_P: 3.048 C_P: 0.234	w_E: 1.636 c_E: 12.32 w_P: 1.412 c_P: 0.234	E: 9.036 P: 0.148	13.01	W_E: 0 C_E: 12.32 W_P: 11.02 C_P: 0	w_E: 5.914 c_E: 5.693 w_P: 5102 c_P: 6.630	E: 1.155 P: 1.155	25.05	1.926
	$\rho = 0.9$ $s = 1$	W_E: 0 C_E: 229.6 W_P: 28.32 C_P: 0.013	w_E: 15.20 c_E: 229.6 w_P: 13.12 c_P: 0.013	E: 18.13 P: 0.001	229.6	W_E: 0 C_E: 229.6 W_P: 176.1 C_P: 0	w_E: 94.52 c_E: 106.1 w_P: 81.54 c_P: 123.5	E: 1.347 P: 1.347	466.7	2.032
	Scale effects: $\rho = 0$, any $s > 0$	W_E: 0.934 C_E: 1.080 W_P: 1.077 C_P: 0.931	w_E: 1.080 c_E: 1.080 w_P: 0.931 c_P: 0.931	E: 1.20 P: 0.889	4.089	W_E: 1 C_E: 1 W_P: 1 C_P: 1	w_E: 1 c_E: 1 w_P: 1 c_P: 1	E: 1.069 P: 1.069	4.472	1.093

[a]The results are obtained as solutions of the optimization problems (2a) and (2b) under alternative scenarios. [b]The price of wine is measured as the consumer's marginal rate of substitution between wine and cloth, after normalizing the price of cloth to 1.

effects of partial trade restrictions are less severe (since Portugal can still export to England) than complete autarky. But the same qualitative results apply. Again, we can measure the benefits from trade by the ratio VT/VR, where VT is the total value of goods under free trade and VR is the total value of goods under trade restriction. Table 2 shows that the benefits of trade are small under a convex technology: they never exceed 9.3 percent as $VT/VR \leq 1.093$ when $\rho \leq 0$. But the benefits from trade become much larger under non-convex technologies: VT/VR varies between 1.192 (+19.2 percent) to 2.032 (+103.2 percent) as ρ increases from 0.1 to 0.9. Again, this associates large benefits from trade with non-convex technologies.

The Ricardo example shows that non-convexity in wine and cloth production could give large benefits from trade when wine and cloth are traded. This suggests the need to consider the case of multiple traded goods (e.g., more than the two traded goods considered in Ricardo's example). But doing so introduces significant challenges to the analysis of trade benefits. Indeed, we know that introducing non-convexity in the analysis can invalidate the standard welfare theorems stating that competitive markets and marginal cost pricing can support a Pareto efficient allocation (e.g., Brown [17]). This raises the question: How to analyze the gains from trade under non-convexity in the context of a general equilibrium model involving many goods? This is the topic of the next sections.

3. A General Equilibrium Model

This section develops a general equilibrium model of a Ricardian economy. The model provides a basis to evaluate aggregate efficiency, with a focus on the case where trade restrictions are the only source of inefficiency. Thus, our analysis does not examine the role of oligopolistic competition and its effects on the gains from trade (as noted in footnote 2). Yet, we will examine the role of non-convexity and discuss its implications for pricing in section 4. And its implications for the benefits from trade are presented in Section 5.

Consider an economy involving a set $\boldsymbol{K} = \{1, \cdots, K\}$ of goods produced by a set $\boldsymbol{M} = \{1, \cdots, M\}$ of firms. Using the netput notation, the j-th firm produces $\boldsymbol{y}_j = (y_{1j}, \cdots, y_{Kj}) \in Y_j \subset \mathbb{R}^K$, where y_{kj} is the k-th output (k-th input if negative) of the j-th firm, and Y_j is the feasible set representing the technology available to the j-th firm, $j \in \boldsymbol{M}$. Considering the case of K goods allows for product differentiation.

We want to cover a broad range of technological possibilities. As such, we do not impose strong restrictions on each set Y_j. First, we allow each firm to face a different technology, with each Y_j possibly varying across firms, $j \in \boldsymbol{M}$. Second, we allow each firm to be either active $(\boldsymbol{y}_j \neq 0)$ or inactive $(\boldsymbol{y}_j = 0)$. Thus, our analysis can capture the role of firm's entry/exit decisions (depending on economic conditions). Third, following the discussion presented in section 2, we do not assume that each set Y_j is convex, *i.e.*, we allow for non-convex technologies. This includes non-convexity among outputs as well as the presence of increasing returns in the input-output space. Fourth, our analysis can address issues related to the organization of each firm. Indeed, each feasible set Y_j can capture the role of intra-firm organization and its effect on firm productivity (e.g., the specialization of tasks within the firm as well as the need for coordination across tasks; see Caliendo and Rossi-Hansberg [11] and Becker and Murphy [20]).

The K goods are consumed by a set $\boldsymbol{N} = \{1, \cdots, N\}$ of households. The i-th household has initial endowment $\boldsymbol{w}_i = (w_{1i}, \cdots, w_{Ki})$, consumes $\boldsymbol{x}_i = (x_{1i}, \cdots, x_{Ki}) \in \mathbb{R}_+^K$ and has preferences represented by the utility function $u_i(\boldsymbol{x}_i), i \in \boldsymbol{N}$. Let $\boldsymbol{x} \equiv (x_1, \cdots, x_N)$, $\boldsymbol{y} \equiv (y_1, \cdots, y_M)$, and $Y \equiv Y_1 \times \cdots \times Y_M$. An allocation $(\boldsymbol{x}, \boldsymbol{y})$ is feasible if it satisfies $\boldsymbol{x} \in \mathbb{R}_+^{NK}$, $\boldsymbol{y} \in Y$, and

$$\sum\nolimits_{i \in N} \boldsymbol{x}_i \leq \sum\nolimits_{i \in N} \boldsymbol{w}_i + \sum\nolimits_{j \in M} \boldsymbol{y}_j. \tag{3}$$

Equation (3) is the commodity balance equation, stating that aggregate consumption cannot exceed aggregate supply. Throughout the paper, we assume that the set Y is closed and bounded, and that the set $\left\{\sum\nolimits_{j \in M} \boldsymbol{Y}_j + \sum\nolimits_{i \in N} \boldsymbol{w}_i\right\} \cap \mathbb{R}_+^K$ has a non-empty interior. Also, following Luenberger [22], we assume that the utility function $u_i(\boldsymbol{x}_i)$ is continuous, strongly monotonic and quasi-concave on $\mathbb{R}_+^K, i \in \boldsymbol{N}$.

We want to analyze the economic and welfare implications of trade. Allowing for non-convexity in *Y*, this section extends the analysis presented by Chavas and Briec [18], with a focus on the effects of trade restrictions. We consider that the economy includes two regions, *A* and *B*, where region *A* has a protectionist trade policy that restricts import from region *B*. Note that this covers the scenario of trade restrictions evaluated in **Table 2** (with Portugal as region *A* and England as region *B*). Import restrictions into region *A* are represented by import quotas $\boldsymbol{q} = (q_1, \cdots, q_K) \in \mathbb{R}_+^K$, where q_k the import quota imposed by region *A* on the k-th product. Let $\boldsymbol{N} = (\boldsymbol{N}_A, \boldsymbol{N}_B)$ and $\boldsymbol{M} = (\boldsymbol{M}_A, \boldsymbol{M}_B)$ where $\boldsymbol{N}_r$ is the set of households in region r, and $\boldsymbol{M}_r$ is the set of firms in region $r, r = A, B$. The net imports into region *A* are:

$$\boldsymbol{m}_A \equiv \sum\nolimits_{i \in N_A} \boldsymbol{x}_i - \sum\nolimits_{i \in N_A} \boldsymbol{w}_i - \sum\nolimits_{j \in M_A} \boldsymbol{y}_j$$

Quota restrictions imposed on imports to region *A* amounts to

$$\sum\nolimits_{i \in N_A} \boldsymbol{x}_i - \sum\nolimits_{i \in N_A} \boldsymbol{w}_i - \sum\nolimits_{j \in M_A} \boldsymbol{y}_j \leq \boldsymbol{q}. \tag{4}$$

Below, we focus our attention on two scenarios in (4): $\boldsymbol{q} = \infty$ corresponding to free trade; and $\boldsymbol{q} = 0$ corresponding to non-positive imports into region *A*.[5] In particular, we want to investigate the welfare difference between these two scenarios.[6]

Our analysis of welfare relies on the benefit function. Letting $\boldsymbol{g} \in \mathbb{R}_+^K$ with

[5]In the Ricardo example, $\boldsymbol{q} = 0$ corresponds to the import quota ban imposed by Portugal (analyzed in **Table 2**). The case of autarky (analyzed in **Table 1**) would be obtained by adding an additional constraint restricting imports to be non-positive in region B.

[6]While Equation (4) focuses on trade quotas, our analysis could also be presented in terms of trade tariffs with their negative effects on trade. Our dual approach presented in section 4 will provide this alternative characterization, with tariffs being equivalent to the quota rents $Q(\cdot)$.

$\boldsymbol{g} \neq 0$, and following Luenberger ([22], [23]), define the benefit function as

$$b_i(\boldsymbol{x}_i, U_i) = \text{Max}_\beta \{\beta : u_i(\boldsymbol{x}_i - \beta \boldsymbol{g}) \geq U_i, (\boldsymbol{x}_i - \beta \boldsymbol{g}) \in \mathbb{R}_+^K\}$$
$$\text{if} \quad \Delta_i \neq \emptyset, = -\infty, \quad \text{otherwise}. \tag{5}$$

where $\Delta_i = \{\beta : u_i(\boldsymbol{x}_i - \beta \boldsymbol{g}) \geq U_i, (\boldsymbol{x}_i - \beta \boldsymbol{g}) \in \mathbb{R}_+^K\}, i \in \boldsymbol{N}$. The benefit function $b_i(\boldsymbol{x}_i, U_i)$ is a welfare measure giving the number of units of the reference bundle $\boldsymbol{g}$ the i-th consumer is willing to give up starting from utility U_i to reach point $\boldsymbol{x}_i$. A convenient choice is to select the reference bundle $\boldsymbol{g}$ such that 1 unit of $\boldsymbol{g}$ is worth $1, making the benefit function $b_i(\boldsymbol{x}_i, U_i)$ in (5) a monetary measure of willingness to pay.[7] Under the quasi-concavity of $u_i(\boldsymbol{x}_i)$ on $\mathbb{R}_+^K$, Luenberger ([22] and [23]) showed that $b_i(\boldsymbol{x}_i, U_i)$ is upper semi-continuous in $(\boldsymbol{x}_i, U_i)$, concave in $\boldsymbol{x}_i$, non-increasing in U_i, and it satisfies the translation property $b_i(\boldsymbol{x}_i + \alpha \boldsymbol{g}, U_i) = \alpha + b_i(\boldsymbol{x}_i, U_i)$ for any finite $\alpha \in \mathbb{R}$. Finally, we say that the reference bundle $\boldsymbol{g}$ is good for the i-th consumer if, for any $\boldsymbol{x}_i \in \mathbb{R}_+^K$, we have $u_i(\boldsymbol{x}_i + \alpha \boldsymbol{g}) > u_i(\boldsymbol{x}_i)$ for all $\alpha > 0$.

Let

$$\mathbb{U} = \left\{(U_1, \cdots, U_N) : u_i(\boldsymbol{x}_i) = U_i, \sum_{i \in N} \boldsymbol{x}_i \leq \sum_{i \in N} \boldsymbol{w}_i + \sum_{j \in M} \boldsymbol{y}_j : \boldsymbol{x}_i \in \mathbb{R}_+^K, i \in N; y \in Y\right\}$$

be the set of attainable utilities. For a given $\boldsymbol{U} \equiv (U_1, \cdots, U_N) \in \mathbb{U}$, consider the maximization problem

$$V(\boldsymbol{U}, \boldsymbol{q}) = \text{Max}_{x,y} \left\{ \sum_{i \in N} b_i(\boldsymbol{x}_i, U_i) : \sum_{i \in N} \boldsymbol{w}_i + \sum_{j \in M} \boldsymbol{y}_j - \sum_{i \in N} \boldsymbol{x}_i \geq 0, \right.$$
$$\left. \sum_{i \in N_A} \boldsymbol{w}_i + \boldsymbol{q} + \sum_{j \in M_A} \boldsymbol{y}_j - \sum_{i \in N_A} \boldsymbol{x}_i \geq 0, \boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y \right\}. \tag{6}$$

For a given $(\boldsymbol{U}, \boldsymbol{q})$, $V(\boldsymbol{U}, \boldsymbol{q})$ in (6) is the largest feasible aggregate benefit that can be obtained subject to the feasibility constraint (3) and the import quota restriction (4).[8]

First, consider the free trade scenario, where $\boldsymbol{q} = \infty$. Then, the quota restriction (4) is not binding in (6). Following Luenberger [22], when $\boldsymbol{q} = \infty$, define an allocation satisfying (6) as a maximal allocation. Then, $V(\boldsymbol{U}, \infty)$ in (6) is Allais' distributable surplus (Allais [24]). $V(\boldsymbol{U}, \infty)$ being non-increasing in $\boldsymbol{U}$ reflects that reaching higher utilities is possible only with a redistribution of the aggregate surplus V, *i.e.* a reduction in V. In addition, define a zero-maximal allocation as a maximal allocation where $\boldsymbol{U} \in \{\boldsymbol{U}' : V(\boldsymbol{U}', \infty) = 0\}$. These are allocations that maximize aggregate benefit, where the resulting surplus is entirely redistributed.

As showed by Luenberger ([22], [25]), under free trade (where $\boldsymbol{q} = \infty$), there are close linkages between zero-maximality and Pareto efficiency.[9]

[7]In this context, the price normalization rule used in Table 1 and Table 2 (setting the price of cloth to 1) is equivalent to choosing the bundle $\boldsymbol{g}$ to be 1 unit of cloth.

[8]Note that we cannot rule out the possibility of multiple solutions to (6). Note that, if such situations were to arise, the aggregate benefit $V(\boldsymbol{U}, \boldsymbol{q})$ in (6) would still be unique even in the presence of multiple equilibrium.

[9]A feasible allocation is Pareto efficient if no individual can be made better off without making anyone else worse off.

Lemma 1: (Luenberger, [22], p. 191)

Suppose that the reference bundle $\boldsymbol{g}$ is good for at least one consumer. Let $\boldsymbol{q} = \infty$. If a feasible allocation $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is zero-maximal, then it is Pareto efficient.

Lemma 2: (Luenberger, [25])[10]

Let $\boldsymbol{q} = \infty$. If an allocation $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is zero-maximal, then it is Pareto efficient compared to all feasible allocations where $\boldsymbol{x}_i$ is in the interior of $\mathbb{R}_+^K$ for each $i \in N$.

Lemma 1 and 2 establish the close relationship existing between zero-maximality and Pareto efficiency under free trade. They associate Pareto efficiency with two intuitive properties: first, the maximization of aggregate benefit; and second, the complete redistribution of surplus. Finally, the set $\{\boldsymbol{U} : W(\boldsymbol{U}, \infty) = 0\}$ defines the Pareto utility frontier, *i.e.* the set of consumer utilities that can be reached under efficient allocations. Importantly, these results apply without assuming that the production set Y is convex.

While Lemma 1 and 2 apply under free trade (when $\boldsymbol{q} = \infty$), the above discussion suggests that (6) will be useful as well when $\boldsymbol{q} = 0$ (the no-import scenario), or more generally when $\boldsymbol{q} < \infty$. In this context, when $0 \le \boldsymbol{q} < \infty$ and conditional on $\boldsymbol{q}$, define the allocations given in (6) as trade-restricted maximal allocations. Then, $V(\boldsymbol{U}, \boldsymbol{q})$ in (6) would be the distributable surplus under import quotas $\boldsymbol{q}$. In addition, conditional on $\boldsymbol{q}$, define a trade-restricted zero-maximal allocation as a trade-restricted maximal allocation where $\boldsymbol{U} \in \{\boldsymbol{U}' : V(\boldsymbol{U}', \boldsymbol{q}) = 0\}$. These are allocations that maximize aggregate benefit subject to the import quota restrictions (4), where the resulting surplus is entirely redistributed. And $\boldsymbol{U} \in \{\boldsymbol{U}' : V(\boldsymbol{U}', \boldsymbol{q}) = 0\}$ can be interpreted as the trade-restricted utility frontier under the quota restrictions in (4).

The above general equilibrium model provides a basis for evaluating the welfare effects of trade. First, the trade-distorted maximal allocation given in Equation (6) gives $V(\boldsymbol{U}, \boldsymbol{q})$ as a measure of aggregate benefit obtained under quotas $\boldsymbol{q}$. And $V(\boldsymbol{U}, \boldsymbol{q})$ is an aggregate willingness-to-pay measure when the reference bundle $\boldsymbol{g}$ is chosen such that 1 unit of $\boldsymbol{g}$ is worth \$1. Second, conditional on $\boldsymbol{q}$, choosing $\boldsymbol{U}$ to satisfy $V(\boldsymbol{U}, \boldsymbol{q}) = 0$, Equation (6) characterizes a trade-distorted zero-maximal allocation. As discussed above, $\{\boldsymbol{U} : V(\boldsymbol{U}, \infty) = 0\}$ defines the Pareto utility frontier under free trade. Alternatively, $\{\boldsymbol{U} : V(\boldsymbol{U}, 0) = 0\}$ defines the utility frontier under no-import for region A. And more generally, $\{\boldsymbol{U} : V(\boldsymbol{U}, \boldsymbol{q}) = 0\}$ defines the utility frontier under import quotas $\boldsymbol{q}$: it is the set of consumer utilities that can be reached under the trade-restrictions (4). Finally, for a given $\boldsymbol{U}$, the welfare effects of a change from $\boldsymbol{q}$ to $\boldsymbol{q}'$ can be measured by the change in aggregate benefit: $\Delta V \equiv V(\boldsymbol{U}, \boldsymbol{q}') - V(\boldsymbol{U}, \boldsymbol{q})$, with $\Delta V > 0$ identifying a potential Pareto improving move. We will make extensive use of this welfare measure in our analysis of the gains from trade in Section 5.

[10]Luenberger ([25], *p.* 231) presents a proof without production. As noted in Luenberger ([25], p. 245), extending the proof with production is straightforward.

4. The Effects of Trade Restrictions: A Dual Approach under Non-Convexity

We now examine a dual general equilibrium model of trade and use it to examine the effects of trade quotas $\boldsymbol{q}$ on welfare. The dual approach is well known under convexity (see Dixit and Norman [26]; Luenberger [27]; Chau and Fare [28]). In this section, we extend the dual approach presented by Chavas and Briec [18] under non-convexity of Y, with a focus on the effects of trade restrictions. The economic implications are examined in section 5 below.

Define F as the set of continuous and non-decreasing functions f from $\mathbb{R}^K$ to $\mathbb{R}$ that satisfy the translation property: $f(\boldsymbol{z}+\alpha\boldsymbol{g}) = \alpha + f(\boldsymbol{z})$ for any $z \in \mathbb{R}^K$ and any finite $\alpha \in \mathbb{R}$. Imposing the translation property will allow us to interpret functions in F as measuring values that are congruent to aggregate benefit (see below). For a given $\boldsymbol{U} \in \mathbb{U}$ and quotas $\boldsymbol{q}$, consider the generalized Lagrangian functional L:

$$L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}) = \sum_{i\in N} b_i(\boldsymbol{x}_i,U_i) + f\left(\sum_{j\in M}\boldsymbol{y}_j\right) - f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{j\in M_A}\boldsymbol{y}_j\right) - Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right), \quad (7)$$

where $\boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y, f \in F$ and $Q \in F$, and where

$$\left[f\left(\sum_{j\in M}\boldsymbol{y}_j\right) - f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right)\right]$$

and

$$\left[Q\left(\sum_{j\in M_A}\boldsymbol{y}_j\right) - Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right)\right]$$

are "penalty functions" associated with constraints (3) and (4), respectively. Consider

$$L^*(\boldsymbol{U},\boldsymbol{q}) = \mathrm{Inf}_{f,Q}\mathrm{Sup}_{x,y}\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}) : \boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y, f \in F, Q \in F\right\}, \quad (8)$$

and

$$L^\#(\boldsymbol{U},\boldsymbol{q}) = \mathrm{Sup}_{x,y}\mathrm{Inf}_{f,Q}\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}) : \boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y, f \in F, Q \in F\right\}. \quad (9)$$

A first step in our analysis is presented next (All proofs are in the **Appendix**).

Lemma 3: (Weak Duality). For $\boldsymbol{U} \in \mathbb{U}$,

$$L^*(\boldsymbol{U},\boldsymbol{q}) \geq L^\#(\boldsymbol{U},\boldsymbol{q}) \geq V(\boldsymbol{U},\boldsymbol{q}) \quad (10)$$

The inequalities in (10) show that $L^*(\boldsymbol{U},\boldsymbol{q})$ is an upper-bound of both $L^\#(\boldsymbol{U},\boldsymbol{q})$ and $V(\boldsymbol{U},\boldsymbol{q})$. Situations where this upper-bound is reached are of significant interest (e.g., Bertsekas [29]; Rubinov *et al.* [30]; Nedic and Ozdaglar [31]). Some key results are presented next.

Proposition 1: Assume that a trade-restricted maximal allocation exists for $\boldsymbol{U} \in \mathbb{U}$, and that $L^*(\boldsymbol{U},\boldsymbol{q}) = L^\#(\boldsymbol{U},\boldsymbol{q})$. Then, there is a saddle-point $(\boldsymbol{x}^*,\boldsymbol{y}^*,f^*,Q^*)$ of the Lagrangian (7) where $\boldsymbol{x}^* \in \mathbb{R}_+^{NK}, \boldsymbol{y}^* \in Y, f^* \in F$ and $Q^* \in F$ satisfy

a) $$L\left(\boldsymbol{U}, \boldsymbol{x}, \boldsymbol{y}, f^*, Q^*, \boldsymbol{q}\right) \le L\left(\boldsymbol{U}, \boldsymbol{x}^*, \boldsymbol{y}^*, f^*, Q^*, \boldsymbol{q}\right) \le L\left(\boldsymbol{U}, \boldsymbol{x}^*, \boldsymbol{y}^*, f, Q, \boldsymbol{q}\right), \text{ for all } \boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y, f \in F, Q \in F, \tag{11}$$

b) $$\sum_{i\in N} \boldsymbol{x}_i^* \le \sum_{j\in M} \boldsymbol{y}_j^* + \sum_{i\in N} \boldsymbol{w}_i, \tag{12a}$$

$$\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \sum_{j\in M_A} \boldsymbol{y}_j^* \le \boldsymbol{q}, \tag{12b}$$

c) $$f^*\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) = f^*\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right), \tag{13a}$$

$$Q^*\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) = Q^*\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right), \tag{13b}$$

d) $$\left(\boldsymbol{x}^*, \boldsymbol{y}^*\right) \in \arg\max_{x,y} \Big\{\sum_{i\in N} b_i\left(\boldsymbol{x}_i, U_i\right) : \sum_{i\in N} \boldsymbol{w}_i + \sum_{j\in M} \boldsymbol{y}_j - \sum_{i\in N} \boldsymbol{x}_i \ge 0, \sum_{i\in N_A} \boldsymbol{w}_i + \boldsymbol{q} + \sum_{j\in M_A} \boldsymbol{y}_j - \sum_{i\in N_A} \boldsymbol{x}_i \ge 0, \boldsymbol{x} \in \mathbb{R}_+^{NK}, \boldsymbol{y} \in Y\Big\}, \tag{14}$$

e) $$L^*\left(\boldsymbol{U}, \boldsymbol{q}\right) = V\left(\boldsymbol{U}, \boldsymbol{q}\right). \tag{15}$$

Proposition 1 establishes that, if the condition $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = L^\#\left(\boldsymbol{U},\boldsymbol{q}\right)$ holds, then there exists a saddle-point of the generalized Lagrangian, as given in (11). Equations (12a)-(12b) state that $\left(\boldsymbol{x}^*, \boldsymbol{y}^*\right)$ in the saddle-point problem (11) is always a feasible solution that satisfies the constraints (3) and (4). At that point, Equations (13a)-(13b) show that the penalty functions are always zero:

$$f^*\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^*\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) = 0$$

and

$$Q^*\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^*\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) = 0$$

Comparing it with (6), (14) implies that $\left(\boldsymbol{x}^*, \boldsymbol{y}^*\right)$ is a trade-restricted maximal allocation. Finally, (15) states that $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$.

Note that, $f(\cdot)$ being non-decreasing, Equation (13a) implies that $f^*(c)$ does not depend on $\boldsymbol{c}$ in the range $\left[\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i, \sum_{j\in M} \boldsymbol{y}_j^*\right]$. It implies the "complementary slackness" condition: the penalty functions $f^*(c)$ becomes "flat" in c_i when the i-th constraint in (3) is non-binding. Similarly, $Q(\cdot)$ being non-decreasing, Equation (13b) states that $Q^*(c)$ does not depend on c in the range $\left[\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}, \sum_{j\in M_A} \boldsymbol{y}_j^*\right]$. Again, the "complementary slackness" condition: the penalty function $Q^*(c)$ become "flat" in c_i when the i-th constraint in (4) is non-binding.

While $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) \ge V\left(\boldsymbol{U},\boldsymbol{q}\right)$ in general from (10), Equation (15) implies that $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$ when a saddle-point exists. The condition $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$ has been called a condition of "zero duality gap" (Bertsekas [29]; Rubinov *et al.* [30] Nedic and Ozdaglar [31]). The linkages between $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = L^\#\left(\boldsymbol{U},\boldsymbol{q}\right)$ and $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$ are presented next.

Lemma 4: For $\boldsymbol{U} \in \mathbb{U}$, we have $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = L^\#\left(\boldsymbol{U},\boldsymbol{q}\right)$ if and only if $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$.

Combining Proposition1 with Lemma 4, it follows that a zero duality gap, $L^*\left(\boldsymbol{U},\boldsymbol{q}\right) = V\left(\boldsymbol{U},\boldsymbol{q}\right)$, is equivalent to the existence of a saddle-pointin (11). But when does a zero duality gap exist? To answer this question, define

$$W\left(\boldsymbol{U},\boldsymbol{q},\gamma\right)=\text{Sup}_{x,y}\left\{\sum_{i\in N}b_i\left(\boldsymbol{x}_i,U_i\right):\gamma+\sum_{i\in N}\boldsymbol{w}_i+\sum_{j\in M}\boldsymbol{y}_j-\sum_{i\in N}\boldsymbol{x}_i\geq 0,\right.$$
$$\left.\sum_{i\in N_A}\boldsymbol{w}_i+\boldsymbol{q}+\sum_{j\in M_A}\boldsymbol{y}_j-\sum_{i\in N_A}\boldsymbol{x}_i\geq 0,\boldsymbol{x}\in\mathbb{R}_+^{NK},\boldsymbol{y}\in Y\right\},\tag{16}$$

where $\gamma\in\mathbb{R}^K$. Comparing (6) and (16), it is clear that $W\left(\boldsymbol{U},\boldsymbol{q},0\right)=V\left(\boldsymbol{U},\boldsymbol{q}\right)$. As showed by Rubinov *et al.* [30] and Nedic and Ozdaglar [31], a zero-duality gap (with $L^*\left(\boldsymbol{U},\boldsymbol{q}\right)=V\left(\boldsymbol{U},\boldsymbol{q}\right)$) exists if and only if $W\left(\boldsymbol{U},\boldsymbol{q}',\gamma\right)$ is upper semi-continuous in $\left(\boldsymbol{q}',\gamma\right)$ at $\left(\boldsymbol{q},0\right)$. This is the "smoothness condition" required to support a dual representation of economic efficiency under non-convexity. Through the rest of the paper, we assume that this condition is satisfied and that a zero-duality gap holds.[11]

Proposition 2: A trade-restricted maximal allocation satisfies

$$E\left(f,Q,\boldsymbol{U},\boldsymbol{q}\right)=\text{Inf}_x\left\{f\left(\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i\right)+Q\left(\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)\right.$$
$$\left.:u_i\left(\boldsymbol{x}_i\right)\geq U_i,i\in N;\boldsymbol{x}\in\mathbb{R}_+^{NK}\right\}.\tag{17}$$

$$\pi\left(f,G\right)=\text{Sup}_y\left\{f\left(\sum_{j\in M}\boldsymbol{y}_j\right)+Q\left(\sum_{j\in M_A}\boldsymbol{y}_j\right):\boldsymbol{y}\in Y\right\},\tag{18}$$

$$V\left(\boldsymbol{U},\boldsymbol{q}\right)=\text{Inf}_{f,Q}\left\{\pi\left(f,G\right)-E\left(f,Q,\boldsymbol{U},\boldsymbol{q}\right):f\in F,Q\in F\right\}.\tag{19}$$

We can interpret $f\left(\cdot\right)$ and $Q\left(\cdot\right)$ as measuring aggregate values. In Equation (17), it follows that $f\left(\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i\right)$ measures aggregate consumer expenditures net of initial endowments. And $Q\left(\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)$ in (17) measures the aggregate "consumer quota rent", *i.e.* the aggregate cost to consumers of the trade quotas $\boldsymbol{q}$. In this context, $E\left(f,Q,\boldsymbol{U},\boldsymbol{q}\right)$ in (17) is an aggregate expenditure function measuring the smallest possible aggregate expenditure that can generate utilities $\boldsymbol{U}$, conditional on $\left(f,Q,\boldsymbol{q}\right)$.

Similarly, in Equation (18), $f\left(\sum_{j\in M}\boldsymbol{y}_j\right)$ measures the aggregate revenue received by firms, while $Q\left(\sum_{j\in M_A}\boldsymbol{y}_j\right)$ is the aggregate "producer quota rent", *i.e.* the aggregate firm revenue associated with the trade quotas $\boldsymbol{q}$. In this context, $\pi\left(f,G\right)$ in (18) is an aggregate profit function measuring the largest possible aggregate income (including producer quota rents) made by firms.

Thus, from Proposition 2, trade-restricted maximal allocations are consistent with aggregate profit maximization (as given in (18)) and aggregate expenditure minimization (as given in (17)).

Finally, Equation (19) states that the maximized aggregate benefit $V\left(\boldsymbol{U},\boldsymbol{q}\right)$ in (6) is also the aggregate profit $\pi\left(f,G\right)$ net of aggregate expenditure $E\left(f,Q,\boldsymbol{U},\boldsymbol{q}\right)$, provided that $\left(f,Q\right)$ solve the minimization problem in (19).

[11]Note that there may be multiple solutions for $f^*\left(\cdot\right)$ and $Q^*\left(\cdot\right)$ in the saddle-point of the Lagrangian (7). Note that under a zero duality gap, if such situations were to arise, the aggregate benefit $V(\boldsymbol{U},\boldsymbol{q})$ would still be unique. Exploring the issue of non-uniqueness for $f^*\left(\cdot\right)$ and $Q^*\left(\cdot\right)$ appears to be a good topic for further research.

This minimization problem identifies the optimal pricing scheme f^* and quota rent Q^* supporting a trade-restricted maximal allocation. In this context, a trade-restricted zero-maximal allocation (where $\boldsymbol{U}$ satisfies $V(\boldsymbol{U}, q) = 0$) is a trade-restricted maximal allocation where all profits are redistributed to consumers.

Note that Proposition 2 extends the dual approach to trade analysis to situations of non-convexity. Indeed, under convexity, the dual approach to trade is well developed in the literature (Anderson *et al.* [32]; Neary [33]; Feenstra [34]; Anderson and Neary [35] and [36]; Chau *et al.* [37]). What is new here is that our approach applies under non-convexity. As shown in (17), (18) and (19), expenditure minimization and profit maximization still apply at the aggregate under non-convexity. What is different is the nature of pricing: the aggregate values $f(\cdot)$ and $Q(\cdot)$ are now functions.

Under convexity, these functions can always be taken to be linear. Indeed, applying the separating hyperplane theorem under convexity, the functions $f(\cdot)$ and $Q(\cdot)$ define separating hyperplanes whose slopesare Lagrange multipliers measuring(shadow) prices. Then, expenditure minimization and profit maximization would take the classical form found under competitive markets. When $\boldsymbol{q} = \infty$, this is the context where the standard welfare theorems apply, establishing close relationships between profit maximization, competition and Pareto efficiency (e.g., Debreu [16]). Yet, non-convexity can destroy the validity of such linkages. A simple example is the case of a technology exhibiting increasing returns to scale (e.g., in the presence of fixed cost) where, under uniform pricing, profit-maximizing competitive firms cannot make a positive profit and therefore would fail to produce efficiently (e.g., Brown, 1991). In this case, a possible solution is to implement a two-part tariff involving a flat fee (used to cover fixed cost) in addition to a unit price set equal to marginal cost. More generally, the functions $f(\cdot)$ and $Q(\cdot)$ in the generalized Lagrangian (7) are non-linear and define separating hypersurfaces whose slopes still provide measurements of (shadow) prices. Indeed, when $\boldsymbol{q} = \infty$, the gradients of $f^*(\cdot)$ are non-linear prices supporting an efficient allocation under non-convexity (Chavas and Briec [18]). Equation (18) shows that the problem arising under non-convexity does not come from profit maximization; rather it comes from uniform pricing. Indeed, (17), (18) and (19) allow for non-linear pricing which becomes an integral part of efficiency under a non-convex technology. This argument applies to both the revenue function $f(\cdot)$ and the quota rent function $Q(\cdot)$. To see that it applies to the quota rent function $Q(\cdot)$ consider the case where $\boldsymbol{q} = 0$. Then, in the case where region A does not export, Equation (4) reduces to market equilibrium condition in region A, implying that the quota rent becomes the revenue function in region A and that a non-linear function $Q(\cdot)$ becomes equivalent to non-linear pricing in region A.

The expenditure function $E(f, Q, \boldsymbol{U}, \boldsymbol{q})$ in (17) and the profit function $\pi(f, G)$ in (18) apply at the aggregate level. When evaluated at f^* and Q^*, these functions can be written as

$$E\left(f^*,Q^*,\boldsymbol{U},\boldsymbol{q}\right)=\text{Inf}_{x_i}\left\{f^*\left(\sum_{i'\in\{N-i\}}\boldsymbol{x}_{i'}^*+\boldsymbol{x}_i-\sum_{i'\in N}\boldsymbol{w}_{i'}\right)\right.$$
$$\left.+Q^*\left(\sum\nolimits_{i'\in\{N_A-i\}}\boldsymbol{x}_{i'}^*+\delta_i\boldsymbol{x}_i-\sum\nolimits_{i'\in N_A}\boldsymbol{w}_{i'}-\boldsymbol{q}\right);u_i\left(\boldsymbol{x}_i\right)\geq U_i,\boldsymbol{x}_i\in\mathbb{R}_+^K\right\} \tag{17'}$$

for $i\in N$,

where $\delta_i=\begin{Bmatrix}1\\0\end{Bmatrix}$ when $i\begin{Bmatrix}\in\\\notin\end{Bmatrix}N_A$, and

$$\pi\left(f^*,G^*\right)=\text{Sup}_{y_j}\left\{f^*\left(\sum_{j'\in\{M-j\}}\boldsymbol{y}_{j'}^*+\boldsymbol{y}_j\right)+Q^*\left(\sum_{j'\in\{M_A-j\}}\boldsymbol{y}_{j'}+\delta_j\boldsymbol{y}_j\right):\boldsymbol{y}_j\in Y_j\right\} \tag{18'}$$

for $j\in M$,

where $\delta_j=\begin{Bmatrix}1\\0\end{Bmatrix}$ when $j\begin{Bmatrix}\in\\\notin\end{Bmatrix}M_A$. Equations (17') and (18') are a decentralized version of a trade-restricted maximal allocation under non-linear prices. When $f^*(\cdot)$ and $Q^*(\cdot)$ are linear (e.g., under convexity), (17') reduces to expenditure minimization for the i-th consumer and represents consumer rationality; and (18') reduces to decentralized profit maximization for the j-th firm. Alternatively, when $f^*(\cdot)$ and $Q^*(\cdot)$ are non-linear, the gradients of $f^*(\cdot)$ and $Q^*(\cdot)$ take the form of non-linear (shadow) prices supporting a trade-restricted maximal allocation. And when $\boldsymbol{q}=\infty$, (17') and (18') define a decentralized non-linear price equilibrium supporting an efficient allocation under non-convexity (Chavas and Briec [18]; Mordukhovich [38]).

As noted above, non-linear prices may be needed to support a decentralized market allocation (e.g., in the presence of fixed cost). In this case, non-linear prices play two roles: 1) they must clear the market; and 2) they must provide the proper incentives for firms to produce under non-convex technologies. The first role (market clearing) has been associated with a Walrasian auctioneer. It is the only role present in competitive markets under convexity and uniform pricing. The second role arises under non-convexity: it involves non-linear pricing and a price discrimination scheme that can support a decentralized allocation.[12] It means the presence of a discriminating Walrasian auctioneer assigning different prices to different bundles.[13]

The quota rent function $Q(\cdot)$ plays an important role in the economics of trade. We show next how the properties of the function $Q(\cdot)$ provide useful information on the gains from trade.

[12]Non-linear pricing and price discrimination are rather common and can take many forms (Wilson [39]). They can go from a two-part tariff (including a flat fee used to cover fixed cost) to perfect price discrimination extracting all the consumer benefit. As discussed in Sections 3 and 4, under zero maximal allocations where $\boldsymbol{U}\in\left\{\boldsymbol{U}':V\left(\boldsymbol{U}',\boldsymbol{q}\right)=0\right\}$, any rent extracted from price discrimination is entirely redistributed to consumers. The way these rents are redistributed affects the distribution of welfare among households.

[13]An example is given by Aliprantis *et al.* [40] who considers the case of a discriminating auctioneer in the context of an exchange economy. Note that the analysis presented by Aliprantis *et al.* [40] is developed under convexity, indicating that price discrimination schemes can be implemented with or without non-convexity.

5. The Benefit from Trade

This section analyzes the general welfare effects of a change in trade policy from $\boldsymbol{q}$ to $\boldsymbol{q}'$. Let $\left(\boldsymbol{x}^o, \boldsymbol{y}^o, f^o, Q^o\right)$ be a saddle-point of $L(\cdot)$ in (11) under quotas $\boldsymbol{q}$, and let $\left(\boldsymbol{x}', \boldsymbol{y}', f', Q'\right)$ be a saddle-point of $L(\cdot)$ in (11) under quotas $\boldsymbol{q}'$. It follows from Proposition 1 that $\left(\boldsymbol{x}^o, \boldsymbol{y}^o\right)$ is a trade-restricted maximal allocation under $\boldsymbol{q}$, and that $\left(\boldsymbol{x}', \boldsymbol{y}'\right)$ is a trade-restricted maximal allocation under $\boldsymbol{q}'$. Our analysis will rely on the following result.

Proposition 3: Then, for any $\boldsymbol{q} \geq 0$ and $\boldsymbol{q}' \geq 0$,

$$\begin{aligned} & Q'\left(\sum_{i\in N_A} \boldsymbol{x}_i^o - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) - Q'\left(\sum_{i\in N_A} \boldsymbol{x}_i^o - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}'\right) \\ & \leq V\left(\boldsymbol{U}, \boldsymbol{q}'\right) - V\left(\boldsymbol{U}, \boldsymbol{q}\right) \\ & \leq Q^o\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i' - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) - Q^o\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i' - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}'\right) \end{aligned} \tag{20a}$$

where

$$Q'\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i^o - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) - Q'\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i^o - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}'\right) \geq 0 \quad \text{when} \quad \boldsymbol{q}' \geq \boldsymbol{q}. \tag{20b}$$

Proposition 3 provides a general characterization of the aggregate effects of trade policy. It applies for any $q \geq 0$ and $q' \geq 0$. When $q' \geq q$ (corresponding to a relaxation of import quotas from q to q'), Equation (20a) and (20b) imply that $\left[V\left(U, \boldsymbol{q}'\right) - V\left(U, \boldsymbol{q}\right)\right] \geq 0$. Here, $\left[V\left(U, q'\right) - V\left(U, q\right)\right]$ measures the aggregate gains from a change in trade quotas. This gives our first key result: any relaxation in import quotas tends to increase the gains from trade. While this result is well known under convexity (e.g., Samuelson [3]), our analysis shows that it applies as well under non-convexity. Importantly, this result holds under general conditions: it allows for heterogeneous technologies across firms, firm entry and exit, and any change in trade quotas.

The role of non-convexity and its effects on gains from trade can be investigated using Equations (20a) or (20b). As evaluated in **Table 2**, consider a trade liberalization move from $\boldsymbol{q} = 0$ to $\boldsymbol{q}' = \infty$. Then, equations (20a)-(20b) imply that

$$\begin{aligned} 0 & \leq Q'\left(\sum_{i\in N_A} \boldsymbol{x}_i^o - \sum_{i\in N_A} \boldsymbol{w}_i\right) - Q'(-\infty) \\ & \leq V\left(\boldsymbol{U}, \infty\right) - V\left(\boldsymbol{U}, 0\right) \\ & \leq Q^o\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i' - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i\right) - Q^o(-\infty) \end{aligned} \tag{21}$$

Equation (21) establishes bounds on the gains from trade given by $\left[V\left(\boldsymbol{U}, \infty\right) - V\left(\boldsymbol{U}, 0\right)\right]$. The lower bound is $\left[Q'\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i^o - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i\right) - Q'(-\infty)\right] \geq 0$; and the upper bound is $\left[Q^o\left(\sum\nolimits_{i\in N_A} \boldsymbol{x}_i' - \sum\nolimits_{i\in N_A} \boldsymbol{w}_i\right) - Q^o(-\infty)\right] \geq 0$. The lower bound being non-negative implies that the aggregate gains from trade are necessarily non-negative. Again, our contribution is to show that this result, which is well known under convexity, also holds under non-convexity.

Equation (21) also provides another important result: it establishes an upper bound on the gains from trade. Evaluating this upper bound can help evaluate how large the gains from trade can be. Equation (21) implies that the gains from trade can be large only if $\left[Q^o\left(\sum_{i\in N_A} \boldsymbol{x}'_i - \sum_{i\in N_A} \boldsymbol{w}_i\right) - Q^o(-\infty)\right]$ is large. This requires the quota rent function $Q^o(\cdot)$ to have steep slopes.

Additional insights can be obtained in situations where the function $Q^*(c)$ is taken to be differentiable. In this context, let $DQ^*(c)$ denote the $(1\times K)$ vector of derivatives of $Q^*(c)$ with respect to c. Then, Proposition 3implies the following result.

Corollary 1: Assume that $Q^*(c)$ is continuously differentiable in c and that $x_i^*(\boldsymbol{q})$ is continuous in $\boldsymbol{q}, i\in N_A$. Then, for any $\boldsymbol{q}\geq 0$,

$$\mathrm{d}V(\boldsymbol{U},\boldsymbol{q}) = DQ^*\left[\sum_{i\in N_A} \boldsymbol{x}_i^*(\boldsymbol{q}) - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right]\mathrm{d}\boldsymbol{q}, \tag{22}$$

and

$$\mathrm{d}DQ^*\left[\sum_{i\in N_A} \boldsymbol{x}_i^*(\boldsymbol{q}) - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right]\mathrm{d}\boldsymbol{q} \leq 0 . \tag{23}$$

where $DQ^*(c)\geq 0$ for all c, and $\mathrm{d}\boldsymbol{q}$ is a $(K\times 1)$ vector representing a small change in $\boldsymbol{q}$.

Under some regularity conditions, Equation (22) provides a measure of the aggregate welfare effect of a small change in trade in region A. It states that the welfare gain due to a small change in $\boldsymbol{q}$ is given by the marginal quota rent $DQ^*\left[\sum_{i\in NA} \boldsymbol{x}_i^*(\boldsymbol{q}) - \sum_{i\in NA} \boldsymbol{w}_i - \boldsymbol{q}\right]$. This marginal quota rent can be interpreted as a price-dependent demand for import in region *A*. This indicates that this demand for import provides all the information necessary to evaluate the gains from trade. This the argument presented by Arkolakis *et al.* [4] in their analysis of gains from trade under convexity. Thus, Equation (22) generalizes this result under non-convexity.

Equation (23) states that the marginal quota rent $DQ^*(\cdot)$ tends to be decreasing in $\boldsymbol{q}$. This gives the intuitive result that the price-dependent demands for import are in general downward sloping. This is a well-known result under convexity (e.g., Neary [33]; Feenstra [34], Falvey [41]; Anderson and Neary [42]).[14] Equation (23) extends this result under non-convexity. Thus, our analysis shows that many of the qualitative results obtained under convexity continue to hold under non-convexity.

Equations (22) and (23) are local results (in the sense that they apply for a small change in $\boldsymbol{q}$). Yet, they can provide useful information in the context of global changes. To see that, consider a change between two extreme cases: from

[14]Indeed, under convexity, a separating hyperplane exists in (14) and the standard Lagrangian approach applies: the penalty function $Q(\boldsymbol{c})$ in (7) can be taken to be linear, and $DQ^*(\boldsymbol{c}) = \lambda(\boldsymbol{q})\geq 0$ become the Lagrange multipliers measuring the slopes of the separating hyperplane associated with the trade constraints (4). Then, (20a) reduces to: $\lambda(\boldsymbol{q}')[\boldsymbol{q}'-\boldsymbol{q}] \leq V(\boldsymbol{U},\boldsymbol{q}') - V(\boldsymbol{U},\boldsymbol{q}) \leq \lambda(\boldsymbol{q})[\boldsymbol{q}'-\boldsymbol{q}]$. This implies that $\left[\lambda(\boldsymbol{q}')-(\boldsymbol{q})\right][\boldsymbol{q}'-\boldsymbol{q}]\leq 0$. This is the standard result obtained under convexity: increasing trade quotas $\boldsymbol{q}$ tends to reduce the unit quota rents $\lambda(\boldsymbol{q})$ (e.g., Neary [33]; Feenstra [34], Falvey [41]; Anderson and Neary [42]).

$q = 0$ where imports to region A are restricted to be zero; to $q = \infty$ corresponding to free trade. Then, using the fundamental theorem of calculus, Equation (22) implies that the gains from trade can be written as

$$V(\boldsymbol{U},\infty) - V(\boldsymbol{U},0) = \int_0^{\infty} DQ^* \left[\sum_{i \in N_A} \boldsymbol{x}_i^*(\boldsymbol{q}) - \sum_{i \in N_A} \boldsymbol{w}_i - \boldsymbol{q} \right] \mathrm{d}\boldsymbol{q}. \quad (24)$$

Interpreting $DQ^*(\cdot)$ as price-dependent import demand functions, it follows from (24) that the gains from trade are entirely determined by the properties of the price-dependent import demand functions. As noted above, this is consistent with the arguments presented by Arkolakis *et al.* [4]. This result applies with or without convexity.

A key issue is: how can non-convexity affect the magnitude of gains from trade? From the simulation analysis reported in **Table 2**, we know that non-convexity can generate large gains from trade. Comparing the results from **Table 2** with Equation (24),[15] we obtain a key insight: non-convexity must have potentially large effects on the marginal quota rents $DQ^*(\cdot)$. This is illustrated in **Figure 4** under three scenarios: 1) the case of linear technology (Ricardo's example given in **Figure 1**); 2) the case of a convex technology (corresponding to **Figure 2**); and 3) the case of non-convex technology (corresponding to **Figure 3**). In all cases, the function $DQ^*(q)$ is non-negative and non-increasing in q.

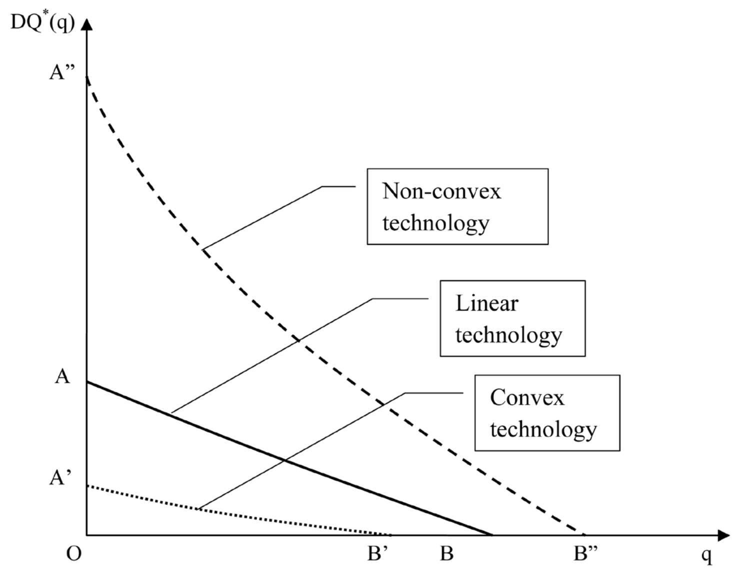

Figure 4. Gains from trade under alternative technologies.

But switching from a linear technology to a (strictly) convex technology means that $DQ^*(q)$ shifts down. Alternatively, introducing non-convexity means that

[15]Note that both Equation (24) and Table 2 evaluate the welfare effects of import quotas. But they do it in a slightly different way. Indeed, Equation (24) compares the maximized aggregate benefit with and without import quotas, holding utilities $\boldsymbol{U}$ constant. As such, Equation (24) provides a Hicksian aggregate willingness-to-pay measure of the effects of trade restrictions. In contrast, VR and VT in Table 2 compare the aggregate value of all consumer goods with and without import quotas, allowing utilities $\boldsymbol{U}$ to adjust. As such they are Marshallian measures of the effects of trade restrictions. From duality, we know that Hicksian and Marshallian measures are closely related; but that they are not equivalent in the presence of income effects. Since our analysis allows for income effects, our comparison of Equation (24) with VR and VT in Table 2 is limited to providing a broad characterization of the welfare effects of trade restrictions.

the price-dependent demand function $DQ^*(q)$ shifts up. From Equation (24), the gains from trade are given by the area OAB under linear technology, by the area OA'B' under convex technology, and by the area OA"B" under a non-convex technology. This illustrates that the gains from trade decrease with convexity and increase under non-convexity. As indicated in **Table 2**, these quantitative effects can be very large. These large effects are associated with steep slopes of the quota rent function $Q^*(\cdot)$. This occurs under non-convex technologies where stronger incentives to specialize lead to large marginal values $DQ^*(\cdot)$ and large gains from trade. This supports our argument: the key to obtain large gains from trade is the presence of technologies that are non-convex with respect to traded goods.

While the simulation results presented in **Table 1** and **Table 2** were motivated in the context of Ricardo's example (focusing on trade involving two countries and two outputs), they hold under general conditions. First, our general equilibrium model covers trade in outputs as well as inputs involving heterogeneous firms. This can capture trade benefits when the non-convexity comes from increasing returns to scale. (Recall that Ricardo's example could not capture such scale effects because it treats labor as a non-traded good). When increasing returns to scale come from fixed costs, this is consistent with the argument that fixed costs contribute to increasing the gains from trade (e.g., Krugman [6]; Melitz [10]).

Second, fixed cost can affect the productivity benefits of specialization in ways that are unrelated to returns to scale. For example, specialization can reduce the fixed resources used in the process of switching between one production activity and another (e.g., by saving in time lost switching from one task to another, as in Adam Smith's pin factory). Also, as discussed by Caliendo and Rossi-Hansberg [11], there are scenarios where more specialized organizations can reduce the cost of acquiring and processing information. Such productivity effects can be present within a firm. This stresses the need to examine the role of managerial abilities and coordination taking place within each firm (Caliendo and Rossi-Hansberg [11]; Becker and Murphy [20]). In this case, productivity and efficiency gains from specialization come from improved management and could exist irrespective of scale effects or market size.

More generally, non-convexity can contribute to productivity gains from specialization across firms (Baumol *et al.* [43]; Chavas and Kim [44]). This can apply to horizontal specialization of firms across products or locations as well as vertical specialization of firms within a marketing channel (e.g., the case of firms specializing in specific tasks associated with successive stages of a production process). Then, changes in firm specialization would imply changes in the horizontal, spatial and/or vertical organization of industries. Efficiency gains could be attained from greater firm specialization under non-convexity. Again, the non-convexity can come from fixed costs, with specialization gains obtained from saving fixed resources used in the production process (Baumol *et al.* [43], p. 75; Chavas and Kim [44]). This is consistent with Adam Smith's example, where specialization reduces the fixed amount of work time spent switching between production activities. This is also consistent with situations where specialization contributes to reducing the fixed costs of obtaining and processing information (as discussed in Caliendo and Rossi-Hansberg [11]). In these examples, the welfare gains from trade liberalization can be potentially large.

Finally, our analysis allows for heterogeneous technologies across firms, and firm entry and exit. As argued by Melitz [7], Bernard *et al.* [8], Melitz and Trefler [10] and Melitz and Redding [12], in the presence of firm heterogeneity, firm entry/exit can affect the types of firm that remain active, thus altering the structure of industries and their productivity. Our general equilibrium model does capture such effects. Yet, as noted by Arkolakis *et al.* [4], those effects may not be sufficient to yield large aggregate gains from trade. Our analysis identifies the presence of non-convex technologies with respect to traded goods as the crucial factor that can generate large gains from trade.

6. Concluding Remarks

This paper has investigated the gains from trade, with a focus on the role of non-convexity. Introducing non-convexity in the welfare evaluation of trade is challenging as it invalidates standard welfare theorems establishing linkages between market equilibrium and Pareto efficiency. To address this challenge, we developed a general equilibrium model that allows for non-convex technologies. The analysis allows for non-linear pricing which becomes an integral part of efficiency under non-convex technologies. The model is used to evaluate the aggregate welfare effects of globalization. We show that some standard results from trade theory remain valid under non-convexity. This includes the result that any relaxation in trade restrictions tends to generate aggregate efficiency gains. We also show how gains from trade are closely linked with the properties of price-dependent demand functions for exports. This extends the analysis presented by Arkolakis *et al.* [4] to situations of non-convexity. Most importantly, we show that the gains from trade tend to be small under convexity, but that they can become very large under non-convexity. This indicates that the search for larger gains from globalization needs to be associated with non-convex technologies.

Our analysis stresses the role of non-convex technologies and their effects on the magnitude of gains from trade. To the extent that the presence and nature of non-convexity can vary across firms and industries, the benefits from trade can also vary across industries. There is a need to evaluate how the nature of non-convexity in production matches trade liberalization policies, with implications for how the magnitude of gains from trade varies across industries and trade policies. Finally, we have noted that non-convex technologies often require the implementation of non-linear pricing. While the modeling of price discrimination schemes is well known in the analysis of trade policy, it is often presented in the context of inefficient allocations associated with rent-seeking behavior. Our analysis shows that price discrimination schemes can support an efficient allocation under non-convexity. There is a need to explore further how price discrimination schemes get implemented in the economic evaluation of trade policy. These appear to be good topics for future research.

Supported

This research was supported by a grant from the Graduate School, University of Wisconsin, Madison.

References

[1] Smith, A. (1776) The Wealth of Nations (London). Hackett Publishing Company, Indianapolis, 1993.

[2] Ricardo, D. (1817) On the Principles of Political Economy and Taxation (John Murray, London). In: Sraffa, P., Ed., *The Works and Correspondence of David Ricardo*, Vol. 1, Cambridge University Press, Cambridge, 1951.

[3] Samuelson, P. (1962) The Gains from International Trade Once Again. *Economic Journal*, **72**, 820-829. https://doi.org/10.2307/2228353

[4] Arkolakis, C., Costinot, A. and Rodriguez-Clare, A. (2012) New Trade Models, Same Old Gains? *American Economic Review*, **102**, 94-130. https://doi.org/10.1257/aer.102.1.94

[5] Ossa, R. (2014) Trade Wars and Trade Talks with Data. *American Economic Review*, **104**, 4104-4146. https://doi.org/10.1257/aer.104.12.4104

[6] Krugman, P. (1980) Scale Economies, Product Differentiation and the Patterns of Trade. *American Economic Review*, **70**, 950-959.

[7] Melitz, M.J. (2003) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71**, 1695-1725. https://doi.org/10.1111/1468-0262.00467

[8] Bernard, A.B., Eaton, J. Jensen, J.B. and Kortum,S. (2003) Plants and Productivity in International Trade. *American Economic Review*, **93**, 1268-1290. https://doi.org/10.1257/000282803769206296

[9] Balistreri, E.J., Hillberry, R.H. and Rutherford, T.F. (2011) Structural Estimation and Solution of International Trade Models with Heterogeneous Firms. *Journal of International Economics*, **83**, 95-108. https://doi.org/10.1016/j.jinteco.2011.01.001

[10] Melitz, M.J. and Trefler, D. (2012) Gains from Trade when Firms Matter. *Journal of Economic Perspectives*, **26**, 91-118. https://doi.org/10.1257/jep.26.2.91

[11] Caliendo, L. and Rossi-Hansberg, E. (2012) The Impact of Trade on Organization and Productivity. *Quarterly Journal of Economics*, **127**, 1393-1467. https://doi.org/10.1093/qje/qjs016

[12] Melitz, M.J. and Redding, S.J. (2014) Heterogeneous Firms and Trade. In: Gopinath, G., Helpman, E. and Rogoff, K., Eds., *Handbook of International Economics* 4, Elsevier, Amsterdam, 1-54. https://doi.org/10.1016/b978-0-444-54314-1.00001-x

[13] Edmond, C., Midrigan, V. and Yi Xu, D. (2015) Competition, Markups and the Gains from International Trade. *American Economic Review*, **105**, 3183-3221. https://doi.org/10.1257/aer.20120549

[14] Eaton, J. and Kortum, S. (2002) Technology, Geography, and Trade. *Econometrica*, **70**, 1741-1779. https://doi.org/10.1111/1468-0262.00352

[15] Eaton, J. and Kortum, S. (2012) Putting Ricardo to Work. *Journal of Economic Perspectives*, **26**, 65-90. https://doi.org/10.1257/jep.26.2.65

[16] Debreu, G. (1959) Theory of Value. Wiley, New York.

[17] Brown, D.J. (1991) Equilibrium Analysis with Non-Convex Technologies. In: Hildenbrand, W. and Sonnenschein, H., Eds., *Handbook of Mathematical Economics*, Volume 4, Elsevier Science Publishers, Amsterdam.

[18] Chavas, J.P. and Briec, W. (2012) On Economic Efficiency under Non-Convexity. *Economic Theory*, **50**, 671-701. https://doi.org/10.1007/s00199-010-0587-1

[19] Stigler, G.J. (1951) The Division of Labor Is Limited by the Extent of the Market. *Journal of Political Economy*, **59**, 185-193. https://doi.org/10.1086/257075

[20] Becker, G.S. and Murphy, K.M. (1992) The Division of Labor, Coordination Costs and Knowledge. *Quarterly Journal of Economics*, **107**, 1137-1160. https://doi.org/10.2307/2118383

[21] Powell, A.A. and Gruen, F.H.G. (1968) The Constant Elasticity of Transformation Production Frontier and Linear Supply System. *International Economic Review*, **9**, 315-328. https://doi.org/10.2307/2556228

[22] Luenberger, D.G. (1995) Microeconomic Theory. McGraw-Hill, Inc., New York.

[23] Luenberger, D.G. (1992) Benefit Functions and Duality. *Journal of Mathematical Economics*, **21**, 461-481. https://doi.org/10.1016/0304-4068(92)90035-6

[24] Allais, M. (1981) La Théorie Générale des Surplus. In: Economies et Sociétés, Institut des Sciences Mathématiques et Economiques Appliquées.

[25] Luenberger, D.G. (1992). New Optimality Principles for Economic Efficiency and Equilibrium. *Journal of Optimization Theory and Applications*, **75**, 221-264. https://doi.org/10.1007/BF00941466

[26] Dixit, A.K. and Norman, V. (1980) Theory of International Trade: A Dual General Equilibrium Approach. Cambridge University Press, London. https://doi.org/10.1017/CBO9780511628627

[27] Luenberger, D.G. (1994) Dual Pareto Efficiency. *Journal of Economic Theory*, **62**, 70-85. https://doi.org/10.1006/jeth.1994.1004

[28] Chau, N.H. and Fare, R. (2011) Shadow Pricing Market Access: A Trade Benefit Function Approach. *Journal of Economic Theory*, **146**, 1631-1663. https://doi.org/10.1016/j.jet.2011.05.011

[29] Bertsekas, D.P. (1995) Nonlinear Programming. Athena Scientific, Belmont, MA.

[30] Rubinov, A.M., Huang, X.X. and Yang, X.Q. (2002) The Zero Duality Gap Property and Lower Semicontinuity of the Perturbation Function. *Mathematics of Operations Research*, **17**, 775-791. https://doi.org/10.1287/moor.27.4.775.295

[31] Nedic, A. and Ozdaglar, A. (2008) A Geometric Framework for Nonconvex Optimization: Duality using Augmented Lagrangian Functions. *Journal of Global Optimization*, **40**, 545-573. https://doi.org/10.1007/s10898-006-9122-0

[32] Anderson, J.E., Bannister, G.J. and Neary, J.P. (1995) Domestic Distortions and International Trade. *International Economic Review*, **36**, 139-157. https://doi.org/10.2307/2527430

[33] Neary, J.P. (1995) Trade Liberalisation and Shadow Prices in the Presence of Tariffs and Quotas. *International Economic Review*, **36**, 531-554. https://doi.org/10.2307/2527359

[34] Feenstra, R.C. (1995) Estimating the Effects of Trade Policy. In: Grossman, G. and Rogoff, K., Eds., *Handbook of International Economics*, Vol. 3, Elsevier Science, Amsterdam, 1553-1595. https://doi.org/10.3386/w5051

[35] Anderson, J.E. and Neary, J.P. (1996) A New Approach to Evaluating Trade Policy. *Review of Economic Studies*, **63**, 107-125. https://doi.org/10.2307/2298117

[36] Anderson, J.E. and Neary, J.P. (2003) The Mercantilist Index of Trade Policy. *International Economic Review*, **44**, 627-649. https://doi.org/10.1111/1468-2354.t01-1-00083

[37] Chau, N.H., Fare, R., and Grosskopf, S. (2003) Trade-restrictiveness and Efficiency. *International Economic Review*, **44**, 1079-1095. https://doi.org/10.1111/1468-2354.t01-1-00101

[38] Mordukhovich, B.S. (2006) Decentralized Convex-Type Equilibrium in Nonconvex Models of Welfare Economics via Nonlinear Prices. *Mathematics Research Reports*, **32**, 1-16.

[39] Wilson, R.B. (1992) Nonlinear Pricing. Oxford University Press, Oxford.

[40] Aliprantis, C.D., Tourky, R. and Yannelis, N.C. (2001) A Theory of Value with Non-linear Prices. *Journal of Economic Theory*, **100**, 22-72. https://doi.org/10.1006/jeth.2000.2721

[41] Falvey, R.E. (1988) Tariffs, Quotas and Piecemeal Welfare Reform. *Journal of International Economics*, **25**, 167-183. https://doi.org/10.1016/0022-1996(88)90012-8

[42] Anderson, J.E. and Neary, J.P. (1992) Trade Reform with Quotas, Partial rent Retention and Tariffs. *Econometrica*, **60**, 57-76. https://doi.org/10.2307/2951676

[43] Baumol, W.J., Panzar, J.C. and Willig, R. (1982) Contestable Markets and the Theory of Industry Structure. Harcourt Brace Jovanovich, Inc., New York.

[44] Chavas, J.P. and Kim, K. (2007) Measurement and Sources of Economies of Scope. *Journal of Institutional and Theoretical Economics*, **163**, 411-427. https://doi.org/10.1628/093245607781871354

Appendix

Proof of Lemma 3: Note that

$$\mathrm{Sup}_{x,y}\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}):\boldsymbol{x}\in\mathbb{R}_{+}^{NK},\boldsymbol{y}\in Y\right\}\geq\mathrm{Inf}_{f,Q}\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}):f\in F,Q\in F\right\},$$

for all $\boldsymbol{x}\in\mathbb{R}_{+}^{NK},\boldsymbol{y}\in Y,f\in F,Q\in F$. Using (8) and (9), this proves the first inequality in (10).

Assume that $\sum_{j\in M}\boldsymbol{y}_j<\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i$. Consider a sequence $f^k(\boldsymbol{v})=a+k\sum_{i\in K}v_i$. For $k\geq 1$, we have

$$f^k\left(\sum_{j\in M}\boldsymbol{y}_j\right)-f^k\left(\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i\right)<k\left[\sum_{j\in M}\boldsymbol{y}_j-\sum_{i\in N}\boldsymbol{x}_i+\sum_{i\in N}\boldsymbol{w}_i\right]<0$$

Letting $k\to\infty$, we obtain

$$\lim_{k\to\infty}\left[f^k\left(\sum_{j\in M}\boldsymbol{y}_j\right)-f^k\left(\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i\right)\right]\to-\infty$$

Thus, $\sum_{j\in M}\boldsymbol{y}_j<\sum_{i\in N}\boldsymbol{x}_i-\sum_{i\in N}\boldsymbol{w}_i$ implies that $\mathrm{Inf}_f\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}):f\in F\right\}=-\infty$.

Similarly, assume that $\sum_{j\in M_A}\boldsymbol{y}_j<\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}$. Consider a sequence $Q^k(\boldsymbol{v})=b+k\sum_{i\in K}v_i$. For $k\geq 1$, we have

$$\begin{aligned}&Q^k\left(\sum_{j\in M_A}\boldsymbol{y}_j\right)-Q^k\left(\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)\\&<k\left[\sum_{j\in M_A}\boldsymbol{y}_j-\sum_{i\in N_A}\boldsymbol{x}_i+\sum_{i\in N_A}\boldsymbol{w}_i+\boldsymbol{q}\right]<0\end{aligned}$$

Letting $k\to\infty$, we obtain

$$\lim_{k\to\infty}\left[Q^k\left(\sum_{j\in M_A}\boldsymbol{y}_j\right)-Q^k\left(\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)\right]=-\infty$$

Thus, $\sum_{j\in M_A}\boldsymbol{y}_j<\sum_{i\in N_A}\boldsymbol{x}_i-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}$ implies that $\mathrm{Inf}_Q\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}):Q\in F\right\}=-\infty$. It follows that

$$\begin{aligned}L^{\#}(\boldsymbol{U},\boldsymbol{q})&=\mathrm{Sup}_{x,y}\mathrm{Inf}_{f,Q}\\&\left\{L(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f,Q,\boldsymbol{q}):\sum_{i\in N}\boldsymbol{w}_i+\sum_{j\in M}\boldsymbol{y}_j-\sum_{i\in N}\boldsymbol{x}\geq 0,\right.\\&\left.\sum_{i\in N_A}\boldsymbol{w}_i+\boldsymbol{q}+\sum_{j\in M_A}\boldsymbol{y}_j-\sum_{i\in N_A}\boldsymbol{x}_i\geq 0,\boldsymbol{x}\in\mathbb{R}_{+}^{NK},\boldsymbol{y}\in Y,f\in F,Q\in F\right\}\\&\geq\mathrm{Sup}_{x,y}\left\{\sum_{i\in N}b_i(\boldsymbol{x}_i,U_i):\sum_{i\in N}\boldsymbol{w}_i+\sum_{j\in M}\boldsymbol{y}_j-\sum_{i\in N}\boldsymbol{x}_i\geq 0,\right.\\&\left.\sum_{i\in N_A}\boldsymbol{w}_i+\boldsymbol{q}+\sum_{j\in M_A}\boldsymbol{y}_j-\sum_{i\in N_A}\boldsymbol{x}_i\geq 0,\boldsymbol{x}\in\mathbb{R}_{+}^{NK},\boldsymbol{y}\in Y\right\}\\&=V(\boldsymbol{U},\boldsymbol{q}),\end{aligned}$$

which proves the second inequality in (10). Q.E.D.

Proof of Proposition 1: Equation (11) follows from (8) and (9) when $L^*(\boldsymbol{U},\boldsymbol{q})=L^{\#}(\boldsymbol{U},\boldsymbol{q})$. The second inequality in (11) implies that

$$\begin{aligned}&f^*\left(\sum_{j\in M}\boldsymbol{y}_j^*\right)-f^*\left(\sum_{i\in N}\boldsymbol{x}_i^*-\sum_{i\in N}\boldsymbol{w}_i\right)+Q^*\left(\sum_{j\in M_A}\boldsymbol{y}_j^*\right)-Q^*\left(\sum_{i\in N_A}\boldsymbol{x}_i^*-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)\\&\leq f\left(\sum_{j\in M}\boldsymbol{y}_j^*\right)-f\left(\sum_{i\in N}\boldsymbol{x}_i^*-\sum_{i\in N}\boldsymbol{w}_i\right)+Q\left(\sum_{j\in M_A}\boldsymbol{y}_j^*\right)-Q\left(\sum_{i\in N_A}\boldsymbol{x}_i^*-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)\end{aligned}\tag{A1}$$

for all $f \in F$ and all $Q \in F$. Assume that $\sum_{i\in N} \boldsymbol{x}_*^* > \sum_{j\in M} \boldsymbol{y}_j^* - \sum_{i\in N} \boldsymbol{w}_i$. Then, there exists a strictly increasing function $f^a(\boldsymbol{v}) = \alpha_0 + \alpha_1 \cdot \boldsymbol{v}$, where $\alpha_1 \in \mathbb{R}_{++}^K$ and satisfying $f^a\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^a\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) < 0$. Letting $f^b(\boldsymbol{v}) = \alpha_0 + \alpha_2 \cdot \boldsymbol{v}$ where $\alpha_2 > \alpha_1$, we have

$$f^b\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^b\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) < f^a\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^a\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) < 0$$

This implies that $\left[f\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right)\right]$ does not have a lower bound, which contradicts (A1). This gives (12a).

Similarly, assume that $\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \sum_{j\in M_A} \boldsymbol{y}_j^* > \boldsymbol{q}$. Then, there exists a strictly increasing function $Q^a(\boldsymbol{v}) = \beta_0 + \beta_1 \boldsymbol{v}$, where $\beta_1 \in \mathbb{R}_{++}^K$ and satisfying $Q^a\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^a\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) < 0$. Letting $Q^b(\boldsymbol{v}) = \beta_0 + \beta_2 \cdot \boldsymbol{v}$ where $\beta_2 > \beta_1$, we have

$$\begin{aligned} &Q^b\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^b\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) \\ &< Q^a\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^a\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) < 0 \end{aligned}$$

. This implies that $\left[Q\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right)\right]$ does not have a lower bound, which contradicts (A1). This gives (12b).

Note that

$f^*\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) = f^*\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right)$ when $\sum_{j\in M} \boldsymbol{y}_j^* = \sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i$

Choosing $Q^c(\boldsymbol{v}) = Q^*(\boldsymbol{v})$ and $f^c \in F$ such that $f^c(\boldsymbol{v})$ do not depend on $\boldsymbol{v}$, (A1) implies that $f^*\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^*\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) \le 0$. Next, consider the case where

$$\sum_{j\in M} \boldsymbol{y}_j^* \ge \sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i \quad \text{and} \quad \sum_{j\in M} \boldsymbol{y}_j^* \ne \sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i$$

The function f being non-decreasing, it follows that $f^*\left(\sum_{j\in M} \boldsymbol{y}_j^*\right) - f^*\left(\sum_{i\in N} \boldsymbol{x}_i^* - \sum_{i\in N} \boldsymbol{w}_i\right) \ge 0$. Combining these results yields (13a).

This implies that $Q^*\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) = Q^*\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right)$ when $\sum_{j\in M_A} \boldsymbol{y}_j^* = \sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}$. Choosing $f^c(\boldsymbol{v}) = f^*(\boldsymbol{v})$ and $Q^c \in F$ such that $Q^c(\boldsymbol{v})$ do not depend on $\boldsymbol{v}$, (A1) implies that

$$Q^*\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^*\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) \le 0$$

Next, consider the case where $\sum_{j\in M_A} \boldsymbol{y}_j^* \ge \sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}$ and $\sum_{j\in M_A} \boldsymbol{y}_j^* \ne \sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}$. The function Q being non-decreasing yields $Q^*\left(\sum_{j\in M_A} \boldsymbol{y}_j^*\right) - Q^*\left(\sum_{i\in N_A} \boldsymbol{x}_i^* - \sum_{i\in N_A} \boldsymbol{w}_i - \boldsymbol{q}\right) \ge 0$. Combining these results gives (13b).

Assuming that a maximal allocation exists, using (12), (13a) and (13b), the first inequality in (11) implies (14). Finally, using (6), (8) and (11), equations (13a), (13b) and (14) imply (15). Q.E.D.

Proof of Lemma 4: The statement that $L^*(\boldsymbol{U},\boldsymbol{q}) = L^\#(\boldsymbol{U},\boldsymbol{q})$ implies $L^*(\boldsymbol{U},\boldsymbol{q}) = V(\boldsymbol{U},\boldsymbol{q})$ was shown in Proposition 1. The converse follows from (10). Q.E.D.

Proof of Proposition 2: The optimization with respect to $\boldsymbol{x}$ in (8) implies

$$E(f,Q,\boldsymbol{U},\boldsymbol{q}) \equiv \operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) - \sum_{i\in N} b_i(\boldsymbol{x}_i,\boldsymbol{U}_i) : \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\}. \tag{A2}$$

Note that

$$\begin{aligned}
&\operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) : u_i(\boldsymbol{x}_i) \ge \boldsymbol{U}_i, i\in N; \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\} \\
&\ge \operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) \right. \\
&\qquad \left. - \sum_{i\in N} b_i(\boldsymbol{x}_i,\boldsymbol{U}_i) : u_i(\boldsymbol{x}_i) \ge \boldsymbol{U}_i, i\in N; \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\} \\
&\ge \operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) \right. \\
&\qquad \left. - \sum_{i\in N} b_i(\boldsymbol{x}_i,\boldsymbol{U}_i) : \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\} \\
&= E(f,Q,\boldsymbol{U},\boldsymbol{q}).
\end{aligned} \tag{A3}$$

We now show that the reverse inequality also holds:

$$\begin{aligned}
&E(f,Q,\boldsymbol{U},\boldsymbol{q}) \\
&\ge \operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) : u_i(\boldsymbol{x}_i) \ge U_i, i\in N; \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\}.
\end{aligned} \tag{A4}$$

From (A2), this inequality clearly holds if $\sum_{i\in N} b_i(\boldsymbol{x}_i,U_i) = -\infty$. Next, consider the case where $\sum_{i\in N} b_i(\boldsymbol{x}_i,U_i) > -\infty$. Letting $\boldsymbol{x}_i' = \boldsymbol{x}_i - b_i(\boldsymbol{x}_i,\boldsymbol{U}_i)\boldsymbol{g}$, we have $\boldsymbol{x}' = (\boldsymbol{x}_1',\cdots,\boldsymbol{x}_K') \in \mathbb{R}_+^{NK}$, $u_i(\boldsymbol{x}_i') \ge U_i$ and $b_i(\boldsymbol{x}_i',U_i) = 0$ for all $i\in N$. Using the translation property of b, f and Q, we obtain

$$\begin{aligned}
E(f,Q,\boldsymbol{U},\boldsymbol{q}) &= \operatorname{Inf}_x \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) - \sum_{i\in N} b_i(\boldsymbol{x}_i,U_i) : \boldsymbol{x}\in\mathbb{R}_+^{NK} \right\} \\
&= \operatorname{Inf}_{x,x'} \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i' - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i' - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) \right. \\
&\qquad \left. - \sum_{i\in N} b_i(\boldsymbol{x}_i',U_i) : \boldsymbol{x}\in\mathbb{R}_+^{NK}, \boldsymbol{x}'\mathbb{R}_+^{NK}, \boldsymbol{x}_i' = \boldsymbol{x}_i - b_i(\boldsymbol{x}_i,U_i)\boldsymbol{g}, i\in N \right\} \\
&= \operatorname{Inf}_{x,x'} \left\{ f\left(\sum_{i\in N}\boldsymbol{x}_i' - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i' - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) : \boldsymbol{x}\in\mathbb{R}_+^{NK}, \boldsymbol{x}'\mathbb{R}_+^{NK}, \right. \\
&\qquad \left. \boldsymbol{x}_i' = \boldsymbol{x}_i - b_i(\boldsymbol{x}_i,U_i)\boldsymbol{g}, u_i(\boldsymbol{x}_i') \ge U_i, i\in N \right\} \\
&\ge \operatorname{Inf}_{x'} \left\{ \left(f\sum_{i\in N}\boldsymbol{x}_i' - \sum_{i\in N}\boldsymbol{w}_i\right) + Q\left(\sum_{i\in N_A}\boldsymbol{x}_i' - \sum_{i\in N_A}\boldsymbol{w}_i - \boldsymbol{q}\right) : u_i(\boldsymbol{x}_i') \ge U_i, i\in N, \boldsymbol{x}'\in\mathbb{R}_+^{NK} \right\}
\end{aligned}$$

which proves (A4). Combining (A3) and (A4), (A2) yields (17).

The optimization with respect to $\boldsymbol{y}$ in (8) implies (18). Finally, using (17) and (18), the optimization with respect to f and Q in (8) implies (19).

Q.E.D.

Proof of Proposition 3: From Proposition 1, the first inequality in (11) implies

$L\left(\boldsymbol{U},\boldsymbol{x},\boldsymbol{y},f^{o},Q^{o},\boldsymbol{q}\right)\le V\left(\boldsymbol{U},\boldsymbol{q}\right)$ for all $\boldsymbol{x}\in\mathbb{R}_{+}^{NK},\boldsymbol{y}\in Y$.

Evaluated at $\left(\boldsymbol{x}',\boldsymbol{y}'\right)$, this gives

$$L\left(\boldsymbol{U},\boldsymbol{x}',\boldsymbol{y}',f^{o},Q^{o},\boldsymbol{q}\right)\le V\left(\boldsymbol{U},\boldsymbol{q}\right). \tag{A5}$$

And the second inequality in (11) implies

$V\left(\boldsymbol{U},\boldsymbol{q}'\right)\le L\left(\boldsymbol{U},\boldsymbol{x}',\boldsymbol{y}',f,Q,\boldsymbol{q}'\right)$ for all $f\in F,Q\in F$.

Evaluated at $\left(f^{o},Q^{o}\right)$, this gives

$$V\left(\boldsymbol{U},\boldsymbol{q}'\right)\le L\left(\boldsymbol{U},\boldsymbol{x}',\boldsymbol{y}',f^{o},Q^{o},\boldsymbol{q}'\right) \tag{A6}$$

Adding (A5) and (A6), we obtain

$$\begin{aligned}V\left(\boldsymbol{U},\boldsymbol{q}'\right)-V\left(\boldsymbol{U},\boldsymbol{q}\right)&\le L\left(\boldsymbol{U},\boldsymbol{x}',\boldsymbol{y}',f^{o},Q^{o},\boldsymbol{q}'\right)-L\left(\boldsymbol{U},\boldsymbol{x}',\boldsymbol{y}',f^{o},Q^{o},\boldsymbol{q}\right),\\&=Q^{o}\left(\sum_{i\in N_A}\boldsymbol{x}_i'-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}\right)-Q^{o}\left(\sum_{i\in N_A}\boldsymbol{x}_i'-\sum_{i\in N_A}\boldsymbol{w}_i-\boldsymbol{q}'\right)\end{aligned} \tag{A7}$$

This gives the second inequality in (20a). To obtain the first inequality in (20a), switch $\boldsymbol{q}$ and $\boldsymbol{q}'$ in (A7) and multiply by (−1). Finally, (20b) follows from Q'(**c**) being non-decreasing. Q.E.D.

Proof of Corollary 1: Assume that $Q^{*}\left(c\right)$ is differentiable. Then, applying the mean value theorem to Proposition 3, it follows that for any $\boldsymbol{q}\ge 0$ and $\boldsymbol{q}'\ge 0$, there exist scalars $k_1\in\left[0,1\right]$ and $k_2\in\left[0,1\right]$ such that

$$\begin{aligned}&DQ'\left(\sum_{i\in N_A}\boldsymbol{x}_i^{o}-\sum_{i\in N_A}\boldsymbol{w}_i-\left(k_1\boldsymbol{q}'+\left(1-k_1\right)\boldsymbol{q}\right)\right)\left[\boldsymbol{q}'-\boldsymbol{q}\right]\\&\le V\left(\boldsymbol{U},\boldsymbol{q}'\right)-V\left(\boldsymbol{U},\boldsymbol{q}\right)\\&\le DQ^{o}\left(\sum_{i\in N_A}\boldsymbol{x}_i'-\sum_{i\in N_A}\boldsymbol{w}_i-\left(k_2\boldsymbol{q}'+\left(1-k_2\right)\boldsymbol{q}\right)\right)\left[\boldsymbol{q}'-\boldsymbol{q}\right],\end{aligned} \tag{A8}$$

where $DQ^{*}\left(\boldsymbol{c}\right)\ge 0$ for all $\boldsymbol{c}$. Let $d\boldsymbol{q}=\boldsymbol{q}'-\boldsymbol{q}$ and $\boldsymbol{q}'\to\boldsymbol{q}$. Assume that $DQ^{*}\left(\boldsymbol{c}\right)$ is continuous in $\boldsymbol{c}$ and $\boldsymbol{x}_i^{*}\left(\boldsymbol{q}\right)$ is continuous in $\boldsymbol{q},i\in N_A$. Then, (A8) implies Equations (22) and (23). Q.E.D.