Parallel Trade Policy in General Equilibrium*

RETO FÖLLMI†

University of St.Gallen

BJÖRN HARTMANN‡

University of St.Gallen

JOSEF ZWEIMÜLLER§

University of Zürich

December 2024

Abstract

We introduce a simple two-country, monopolistic competition model of parallel trade policy with potentially imperfect enforcement of parallel trade prohibition. Consumers have 0/1 preferences such that large per capita income differences can induce some Northern firms to refrain from exporting to the South in order to avoid international arbitrage. We show that parallel trade restrictions have a pro-competitive effect on prices and change the terms of trade in favour of the poor country. Hence, the South prefers to ban parallel trade while the North prefers to allow it. The global welfare-maximising policy is to forbid parallel trade. Moreover, trade liberalisation increases the terms of trade of the South if there is a sufficiently strong parallel trade prohibition.

Keywords: non-homothetic preferences, parallel imports, arbitrage, extensive margin, North-South trade

JEL classification: F10, F12, F13

*We thank David Hémous, Kiminori Matsuyama, Ralph Ossa, Karl Schulz, David Torun and seminar participants at LMU Munich, U Namur, Collegio Carlo Alberto Turin, U Zurich, U St.Gallen as well as conference participants at the SSES 2024 Meeting, BiGSEM, and the St. Petersburg IO and Spatial Economics conference for helpful comments and suggestions.

†Reto Föllmi, University of St.Gallen, CEPR, CESifo. reto.foellmi@unisg.ch

‡Björn Hartmann, University of St.Gallen. bjoern.hartmann@unisg.ch

§Josef Zweimüller, University of Zürich, CEPR, CESifo, and IZA. josef.zweimueller@econ.uzh.ch

1 Introduction

Recent years have seen a significant shift in the approach to international patent exhaustion. This shift involves the adoption or confirmation of policies of international patent exhaustion, whereby patents are considered exhausted after the first sale of a good, regardless of where the sale takes place. This policy change raises the question whether it represents a universally optimal policy or whether it primarily reflects the preferences of a particular group of countries.

We present a two-country general equilibrium model of international trade with non-homothetic preferences, and discuss the role of parallel trade policy restrictions. The model is based on the framework introduced by Föllmi, Hepenstrick and Zweimüller (2018, hereafter FHZ). We extend their model along a crucial dimension: We include an explicit choice of parallel trade policy. Thus, a government's trade policy is a joint decision on the optimal level of parallel trade and trade costs / tariffs. To the best of our knowledge, we are the first ones to study this jointly.

Our model reveals divergent preferences on parallel trade between the rich North and the poor South. The North prefers to allow parallel trade, because it can save resources by limiting exports to the South. Consequently, the South prefers to ban parallel trade in order to gain access to a wider range of goods. In case parallel trade is allowed, we also get divergent preferences regarding trade costs. The North has a standard preference for free trade, while the South wants to raise the barriers of trade. High trade costs reduce the arbitrage threat for Northern firms, thereby increasing exports to the benefit of Southern consumers. In contrast, if parallel trade is banned, both countries prefer free trade. The social planner would forbid parallel trade, because the welfare losses of the South caused by limited access to Northern goods strictly outweigh the gains for Northern consumers. Furthermore, we analyse the effects of a divergence or convergence in income between North and South. Unsurprisingly, an income convergence benefits both countries. The South can afford more goods from the North, while producing and exporting more varieties to the North, resulting in benefits for all consumers. However, only the North benefits from an income divergence. All additional resources of the North go into the production of new, non-traded varieties, which the South is unable to afford. Thus, our model predicts asymmetric welfare gains as countries grow richer.

1.1 Related Literature

Pricing-to-Market. Krugman (1986) was the first to study pricing-to-market.[1] He documented that European firms do not fully pass through movements in the exchange rate to US importers but rather maintain their export prices tied to local US prices. For more recent empirical evidence on pricing-to-market, see for example Alessandria and Kaboski (2011), Gopinath, Itskhoki and Rigobon (2010) or Fitzgerald and Haller (2014). On the theoretical side, Atkeson and Burstein (2008) can generate pricing-to-market behaviour of firms in a setting with Cournot competition, variable markups and marginal cost-push shocks. Auer, Chaney and Sauré (2018) explore how market-specific valuations for product quality affect relative prices. Corsetti and Dedola (2005) include nontradable inputs, such that the price elasticity of demand is country-specific and dependent on the exchange rate, causing international price discrimination. In our model, pricing-to-market is driven by differences in the willingness to pay across countries.

International Arbitrage. Pricing-to-market creates international arbitrage opportunities. In our representative agent case, firms face a trade-off between price and market size. If a firm charges a high price, it can only sell its products in the rich North and thus has a small market. This exclusion strategy leads to export zeros for this particular product. When a firm offers its product on the global market, it must consider the possibility of parallel imports.[2] Charging different prices in the two markets creates an opportunity for arbitrageurs to purchase the product cheaply in the South and ship it into the North where they can sell it at a slightly lower price than the official dealer and still make a profit. To avoid losing the Northern market to the arbitrageurs, the firm has to charge a reduced price in the North, constrained by the trade costs. Malueg and Schwartz (1994) present a partial equilibrium model with a single monopolistic producer per country. They find that banning parallel imports and allowing discriminatory pricing improves world welfare when demand dispersion across markets is large, because it allows to serve more markets. Our model predictions are consistent with their findings. Grossman and Lai (2008) discuss the effect of parallel imports on innovation in the context of government price controls applied to the pharmaceutical mar-

[1] Pricing-to-Market refers to the situation where producers choose varying markups, and hence varying prices, for the same good in different markets.

[2] Parallel trade occurs when a good protected by a patent, copyright, or trademark, which has been legally acquired in one country, is exported to another without the permission of the local intellectual property rights holder in the importing market (see Maskus (2000)). Hence, parallel imports are goods imported by unauthorized resellers.

ket. They show that incentives for product innovation in the North may be higher if the North allows parallel imports, but that the South will suffer welfare losses. Li and Maskus (2006) find that the distortions associated with parallel imports inhibit innovation. However, they point out that the global welfare implications depend on the exact interplay between parallel trade restrictions and trade costs. These two dynamic perspectives on the impact of parallel import policy can be seen as complementary to our static treatment focusing on inequality. Roy and Saggi (2012) model a North-South quality-differentiated duopoly competing in prices. They show that when between-country inequality is high, the North has an incentive to forbid parallel imports, since consumers in the North are harmed by the high price of the Northern duopolist, which refrains from exporting to the South due to the threat of arbitrage. In our model, the North always prefers to allow parallel imports as it benefits from an increased domestic variety when export exclusion occurs. The difference in our model's preferred parallel trade prediction is rooted in the possibility of firm entry. Allowing parallel trade induces some rich country firms to exclude the poor country market, which leaves more resources in the rich country and thus allows more firms to be established. We are the first to study parallel trade in general equilibrium and to point out this resource feedback mechanism.[3] In a companion paper, Föllmi, Hartmann and Zweimüller (2025) study the distributional impact of parallel trade policy when there is within-country inequality.

Inequality & Trade. Most of the literature focuses on the effect of trade on inequality.[4] While there is a consensus that trade can be detrimental to employment and wages at the local level, the distributional consequences of trade from a consumer perspective (at the country level) are much more controversial, with findings ranging from pro-poor to distributionally neutral or even pro-rich. Our work goes in the opposite direction, i.e. the effect of inequality on trade. Our model can replicate the Linder (1961) hypothesis, namely that higher per capita income of the poor country unambiguously increases the intensity of trade. Hallak (2010) finds support for the Linder hypothesis, both empirically and theoretically, using GDP per capita to capture demand similarity across countries. Fajgelbaum, Grossman and Helpman (2015) can replicate the Linder hypothesis in an FDI context, such that FDI

[3]There exists a narrow literature on *compulsory licensing*, the second major flexibility that WTO members may use in their patent policy next to the exhaustion regime. See Bond and Saggi (2014, 2018, 2020) for theoretical work on this issue.

[4]See, for example, Acemoglu et al. (2016); Adao et al. (2022); Autor et al. (2013); Behrens and Murata (2012); Borusyak and Jaravel (2021); Fajgelbaum and Khandelwal (2016); Galle et al. (2023); Waugh (2023).

is greater between countries with more similar per capita income levels.

Finally, our perspective of allowing demand heterogeneity to shape the patterns of trade should be seen as complementary to the firm heterogeneity approach initiated by Melitz (2003) and subsequently explored by Arkolakis et al. (2012); Chaney (2008) and empirically confirmed by Bernard et al. (2007, 2009); Helpman et al. (2008); Hummels and Klenow (2005). In our model, all firms have identical Dixit-Stiglitz production technology. Thus, trade patterns are entirely driven by consumer heterogeneity, in contrast to the Melitz framework, where productivity cutoffs explain product-level trade zeros.

The remainder of the paper is structured as follows: Section 2 provides an overview of the current parallel trade policies and reviews the empirical estimates of parallel trade volumes. Section 3 introduces the model and describes the autarky equilibrium. Section 4 presents the two trade equilibria, while section 5 discusses the arbitrage equilibrium and its welfare implications. Section 6 concludes.

2 Parallel Trade in Practice

2.1 Institutional Context

In general, the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), does not provide for a uniform approach to parallel trade policy, leaving it to each member state to decide on a preferred policy.[5] At the same time, Article 6 of the Agreement states that nothing in the Agreement "shall be used to address the issue of the exhaustion of intellectual property rights" in dispute settlement, meaning that any position on parallel trade is valid in the eyes of the WTO and therefore cannot be challenged in court (WTO, 1994).[6]

Countries follow one of three approaches: national, regional, or international exhaustion of rights (Maskus, 2000). Under national exhaustion, the intellectual property right exhausts on first sale if the sale is made within the country. Under this rule, resales within a country

[5]When talking about parallel trade, it is important to note that the term applies only to 'genuine' goods, i.e. goods that were manufactured and placed on the market either by the property rights holder or with his explicit consent. Goods manufactured without the consent of the rights holder are considered infringing goods and are not covered by exhaustion of rights regulations. While IPR infringements are an integral part of any FTA negotiation (especially for FTAs between rich and poor countries), they are not the focus of this paper. We assume that all goods are produced by the property rights holder. See for example Bond and Saggi (2020) for an analysis of IPR infringements and imitations.

[6]The General Provisions and Basic Principles (including Article 6) of the TRIPS Agreement can be found here: https://www.wto.org/english/docs_e/legal_e/trips_e.htm#part1.

are legal, but parallel trade is not. This allows firms to engage in third-degree price discrimination and thus segment markets.[7] Consequently, this approach is the least friendly to international trade. Countries that follow national exhaustion are, for example, Brazil, Mexico, South Korea, Indonesia and several African economies, including Mozambique, Rwanda and Uganda (Calboli, 2022). Regional exhaustion means that patent protection is exhausted after an initial sale within a region, i.e. a particular group of countries such as the European Union. Therefore, parallel trade is allowed within that group of countries, but not if the sale was made outside of it. This system is applied in the EU[8] and the OAPI (the African Intellectual Property Organisation), which includes, among others, countries such as Cameroon, Ivory Coast, Mali, and Senegal. Concerning the EU, the European Court of Justice (ECJ) ruled as early as 1974 that national exhaustion of rights violated the free movement of goods postulated in Article 30 of the Treaty of Rome, meaning that the integrity of the single market trumped intellectual property rights.[9] This extends to arbitrage opportunities created by national price controls, as regularly applied to pharmaceuticals (Maskus, 2000).[10]

Finally, international exhaustion means that the intellectual property right expires after the first sale of the good, regardless of where it is sold. This approach allows for parallel imports from all over the world and is therefore the most friendly to international trade. It is practised by Canada, China, India, Japan, and the US, among others (Calboli, 2022). Until recently, the US followed the approach of national exhaustion. However, following a series of Supreme Court decisions, the US has adopted international exhaustion. First, in the area of copyright, in *Kirtsaeng v. John Wiley & Sons, Inc.* (2013), the Supreme Court held that foreign sales of copyrighted products exhaust the US copyright. In 2017, the international exhaustion of rights was extended to patent rights in the decision *Lexmark v. Impression Products*, overturning several lower court decisions and decades of US patent policy (Ivus and Lai, 2017). However, notable exceptions remain for the pharmaceutical sector and for contractual restrictions that explicitly prohibit parallel imports. China shifted from national to international exhaustion with its new patent law in 2008, while India already did so in 2002 (Calboli, 2022). See table 1 for an overview.

[7] The assumption of national exhaustion is the standard assumption in international trade theory and allows firms to engage in pricing-to-market.

[8] More specifically, in the European Economic Area (EEA), including Liechtenstein, Norway and Island. Switzerland and the United Kingdom (after Brexit) have unilaterally tied themselves to that approach, which means that parallel imports from the EEA *into* Switzerland and the UK are allowed, while the EEA does not recognise parallel imports originating from these two countries.

[9] See *Centrafarm B.V. and Adriaan de Peijper v. Sterling Drug Inc.*, 6 IIC 102 (1975) - *Negram III.*

[10] The ECJ rulings were codified by a European Council Directive in 2008 (Directive 2008/95/EC) and updated in 2015 (Directive 2015/2436/EC).

Table 1: Countries by Patent Exhaustion Regime

National	Regional	International	Switched to International
Australia		Canada	
Brazil	EU	Chile	China (2008)
Indonesia	OAPI	Japan	India (2002)
Mexiko		Singapore	United States (2017)
South Korea		Vietnam	

Note: Own visualisation based on Calboli (2022); Maskus (2000). The list is non-exhaustive. For countries that switched to international exhaustion, the date of the switch is indicated in the bracket.

2.2 Parallel Trade Volume

Empirically, the extent of parallel trade is a rather understudied topic, partly because parallel trade is difficult to observe. Since parallel trade is legal in many parts of the world, there is no reason for customs offices to keep a record of whether a particular good is being imported with or without the explicit consent of the patent holder. Nevertheless, there is convincing evidence that the possibility of parallel trade affects firms' pricing decisions in many markets.

The industries most affected are automobiles (Lutz, 2004; Yeung and Mok, 2013), consumer electronics (Burgess and Evans, 2005) and, most prominently, pharmaceuticals (Dubois and Sæthre, 2020; Dubois et al., 2022; Maskus, 2000; Goldberg, 2010). A report commissioned by the European Union estimated that parallel trade accounts for 10% - 20% of total trade volume within the EU (NERA, 1999), while Ganslandt and Maskus (2004) found that medicines subject to parallel imports experienced price reductions of between 12% - 19% compared to other medicines not subject to arbitrage.

Although these figures give the impression that parallel trade may be of secondary importance, this conclusion would be premature. The volume of parallel trade need not be large, or even exist, for it to have a significant impact on a firm's pricing and export decisions.[11] The mere possibility of parallel trade is enough for firms to adjust their pricing schedule or stop exporting altogether so that parallel trade does not occur.

Goldberg (2010) found that multinational pharmaceutical producers often refrain from exporting to India because the possibility of arbitrage and the practice of "global reference pricing" by several rich countries would lead to greater profit losses in high-price countries than profit gains from entering the Indian market. Lanjouw (2005) shows that pharmaceutical companies systematically delay entry into developing markets in order to enjoy monopoly

[11] In our model, parallel trade will *never* occur in equilibrium because firms adjust their behaviour accordingly.

profits unhindered by the threat of arbitrage. Danzon and Epstein (2008) find that a similar effect is at play between high- and low-price EU countries. Similarly, Ivus (2010) examines the effect of stronger intellectual property rights enforcement in response to the TRIPS agreement. She finds that stronger IPRs lead to a significant increase in export volumes to developing countries.

Regarding binding arbitrage price constraints, FHZ calibrate a multi-country trade model based on the framework of Simonovska (2015). They find that about 20% of all bilateral trade relations, representing 45% of total world trade, are subject to arbitrage threats. Furthermore, they find that arbitrage is particularly important for EU member states such as Germany, France and (at that time) the UK, due to the high degree of integration and low trade costs within Europe. This discussion suggests that the possibility of parallel trade has far-reaching implications for firms' pricing and export decisions, intending to effectively prevent parallel trade.

3 Autarky

3.1 Consumers and Labour Market

The economy is populated by $\mathcal{P}$ identical households. Each household is endowed with L units of labour, which it supplies inelastically to the labour market. We assume that the labour market is perfectly competitive within countries and that workers are paid their marginal product. Labour is immobile across countries. Households spend their income on a continuum of differentiated goods. Goods are indivisible and yield positive utility for the first unit and zero utility for each additional unit. Thus, the consumption decision for each good is a discrete choice between buying one unit of the good or not buying it at all.[12] Let $c(j)$ denote an indicator that takes the value 1 if good j is purchased and the value 0 if not. Then, preferences can be represented by the utility function:

$$U = \int_{\Omega} c(j)dj, \qquad \text{with } c(j) \in \{0, 1\}$$

[12]Such preferences have been used, among others, by Falkinger (1994) and Föllmi and Zweimüller (2006) to analyse inequality and growth, by Matsuyama (2000) to explore non-homotheticities in a Ricardian trade model, by Murphy et al. (1989) to study the process of industrialisation, and by Föllmi et al. (2018) to explain trade zeros caused by demand-side heterogeneity. The preferences used in Fajgelbaum, Grossman and Helpman (2011, 2015) or Mongey and Waugh (2024), which build on the theory of McFadden (1978), exhibit a similar discrete choice.

where Ω denotes the set of products available in a country. These preferences are additively separable, non-homothetic and have only an extensive consumption margin. The budget constraint is given by $\int_0^N p(j)c(j) \leq WL$, where W is the wage rate.[13] Utility maximisation gives the binary purchase decision rule:

$$c(j) = 1 \quad \text{if } 1 \geq \lambda p(j)$$
$$c(j) = 0 \quad \text{if } 1 < \lambda p(j)$$

where λ is the Lagrange multiplier and $1/\lambda$ can be interpreted as the willingness to pay. A consumer will purchase one unit of good j if the willingness to pay is at least equal to the good's price $p(j)$. Figure 1 illustrates the individual demand function that results from this utility function. Individual demand is a one-step function of the price, where the reservation price is the inverse of the consumer's marginal utility of income.[14]

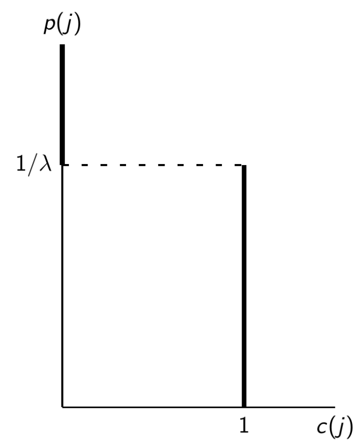


Figure 1: Individual demand function

3.2 Firms and Production

On the supply side, we assume a Krugman (1980) model of international trade with a mass of firms each producing a single, differentiated product. Firms are homogeneous in costs

[13]Note that the integral in the utility function is defined over the whole set of goods available in the economy Ω, while the budget constraint is defined from zero to N, which denotes the set of goods actually purchased by the consumer.

[14]Note that because of the symmetry between goods, there is ex-ante indeterminacy as to which products are consumed first. Introducing a "hierarchy of needs" would resolve this. However, this indeterminacy is unproblematic as long as we look only at equilibrium outcomes. A hierarchy may be operating in the background, but it does not need to be modelled explicitly.

and labour is the only production input, with $1/a$ denoting the marginal costs of production and F the fixed costs (in units of labour). Producing a quantity q of good j thus requires a total labour input of $L(j) = F + q(j)/a$. Given our preferences, market size is given by the population size $\mathcal{P}$, which means that the firm's decision problem boils down to choosing the optimal price $p(j)$. There is free entry into the market. Therefore, the zero-profit conditions endogenously determine the number of active firms in equilibrium.

3.3 Equilibrium

The equilibrium is symmetric since both households and firms are homogeneous. Profit-maximizing firms choose the price of goods to be equal to the consumer's willingness to pay, $p = 1/\lambda$. Firms sell one unit to each consumer, implying a total sales volume of $\mathcal{P}$. Without loss of generality, we use the wage rate as numéraire, $W = 1$. As the first equilibrium condition, the zero-profit condition $[p - W/a]\mathcal{P} = WF$ determines the equilibrium goods price at $p = (aF + \mathcal{P})/a\mathcal{P} = 1/\lambda$. The implied markup, the ratio of price over marginal costs, is then

$$\mu = \frac{aF + \mathcal{P}}{\mathcal{P}}$$

Note that markups are determined by technology parameters a and F and market size $\mathcal{P}$.[15] Intuitively, higher fixed costs F must go together with a higher equilibrium markup, such that the firms break even.

The second equilibrium condition is the resource constraint in the economy, which ensures full employment $[F + \mathcal{P}/a]N = \mathcal{P}L$. The resource constraint determines the equilibrium product variety:

$$N = \frac{a\mathcal{P}}{aF + \mathcal{P}}L$$

4 Trade between Rich and Poor

Now consider two countries that trade with each other. We refer to them as North and South, where the South is assumed to be the poorer of the two and its variables are denoted by an asterisk. For now, we retain the assumption of the representative agent within economies. The population sizes are $\mathcal{P}$ and $\mathcal{P}^*$, where we make no ex-ante assumption about the relative

[15] Variable markups, especially in the context of international trade and pricing-to-market, have received considerable attention in recent years. See for example Amiti, Itskhoki and Konings (2019); Demidova (2017); De Loecker and Eeckhout (2020); Edmond, Midrigan and Xu (2015); Feenstra (2018); Melitz and Ottaviano (2008).

sizes. The labour endowments are denoted by L and L^*, where we assume that $L > L^*$, which ensures that the North is the rich country and the South the poor one. We set the Northern wage as the numéraire and define the relative wage as $\omega = \frac{W^*}{W} = W^*$.

Trade is costly. We assume that trade costs are of the iceberg type: it costs τ unit of the good to get one unit of the good to arrive at its destination, which means that $\tau - 1$ units are lost during transport. We allow for the enforcement of parallel trade prohibition to be potentially imperfect. Before entering the market, a firm does not know whether parallel trade restrictions will be in place for its product. With probability π, price setting is unrestricted as international property rights are fully enforced, while with probability $1 - \pi$, parallel imports are allowed and firms' price setting is constrained by the threat of parallel imports. Thus, parameter π can be interpreted as a measure of the enforcement of international property rights.[16] We assume that parallel trade policy is symmetric, i.e. π is the same in the North and the South.[17]

An exporter from the North will charge a price p in the home market and a price p^* in the foreign market. The degree of price discrimination between markets reflects the difference in the willingness to pay between these markets. Naturally, we would expect that $p \geq p^*$, since consumers in the North have a higher willingness to pay. Optimally, producers would like to charge profit-maximising prices in both markets, unconstrained by any arbitrage threats.[18]

[16] In reality, firms know perfectly well whether parallel trade is allowed or not. However, as discussed in the previous section, the actual extent of parallel trade is uncertain and is likely to vary from time to time. Therefore, a more realistic interpretation of the parameter π would be that parallel trade is legal, but firms are ex-ante uncertain about the extent of parallel trade.

[17] Note that the assumption of symmetry in parallel trade policy between North and South is not strictly necessary. We could alternatively set up the model with an asymmetric parallel trade policy. With a (potential) asymmetry, the North and the South would each set their own π, call it π_R and π_P. Strictly speaking, parallel trade restrictions are parallel *import* restrictions, countries allow or do not allow imports into their territory. Thus, π_R (π_P) denotes the probability that parallel imports *into the North (South)* are not allowed, thus protecting firms from arbitrage threats. Since the price level in the North is (weakly) higher than in the South, the prices in the South are effectively the lowest in the world. Products cannot be bought anywhere for less than the price in the South. Therefore, no one will want to conduct parallel imports into the South because the current prices cannot be undercut. Consequently, the parallel trade policy of the South is irrelevant in equilibrium. Hence, it does not matter whether we assume a symmetric or asymmetric parallel trade policy. The model in Bond and Saggi (2020) exhibits the same property.

[18] Consider the following example of arbitrage threats: A German pharmaceuticals producer wants to sell its product both in Germany and in Poland. Ideally, it would charge a price in Germany equal to the German consumer's willingness to pay $p^{DE} = 1/\lambda^{DE}$ and a price in Poland equal to the Polish consumer's willingness to pay $p^{PL} = 1/\lambda^{PL}$. If the difference between $1/\lambda^{DE}$ and $1/\lambda^{PL}$ is greater than the trade costs between Germany and Poland, arbitrage opportunities arise. Arbitrageurs can purchase the good cheaply in Poland, transport it back to Germany and undercut the producer on the German market. The producer would therefore charge a maximum price difference of τ in order to prevent arbitrage opportunities from arising in the first place.

However, if arbitrage threats are present (the exact conditions for this will be presented later in the paper), the maximum price difference between North and South is equal to the trade costs τ.

When the income gap between North and South is sufficiently small, all goods are traded internationally and all firms can charge full price without the threat of arbitrage. We call this the *full trade equilibrium*. However, when the income gap is large, the price difference between markets is large enough to make arbitrage worthwhile. In such an *arbitrage equilibrium*, it will be the case that some firms in the North choose not to export their products to the South and sell only on the domestic market, at unconstrained prices. The remaining firms in the North will export their product to the South and charge a constrained price at home to prevent parallel imports into their home market.

4.1 Full Trade Equilibrium

In the full trade equilibrium, parallel trade will never take place due to small price differences between North and South. Free entry implies that firms will earn zero profits in equilibrium. For all firms, total revenue is given by $p\mathcal{P} + p^*\mathcal{P}^*$. Due to the assumption of 0-1 preferences, demand in each market is given by its respective population size. Total costs are $W^i\Big[F + (\mathcal{P}^i + \tau\mathcal{P}^{-i})/a\Big]$, which consist of the fixed costs for setting up the production facility, and the marginal costs of production, taking into account the iceberg trade costs for exported goods. To express costs in money terms, we multiply them with the respective wage rate. Combining the two zero-profit conditions delivers the relative wage rate:

$$\omega = \frac{W^*}{W} = \left(\frac{aF + \mathcal{P} + \tau\mathcal{P}^*}{aF + \mathcal{P}^* + \tau\mathcal{P}}\right)$$

Factor price equalisation can be achieved under one of two conditions: (1) there are no trade costs, $\tau = 1$, or (2) the two countries are equally populous ($\mathcal{P} = \mathcal{P}^*$).[19] Within both economies, the total amount of resources required in production must equal the total amount of resources available. These aggregate resource constraints are given by:

$$\mathcal{P}L = N\left(F + \frac{\mathcal{P} + \tau\mathcal{P}^*}{a}\right) \qquad \mathcal{P}^*L^* = N^*\left(F + \frac{\mathcal{P}^* + \tau\mathcal{P}}{a}\right)$$

[19]Note that it is very much possible that the South has a higher wage rate than the North, however, relative nominal per capita income $\omega L^*/L$ would still be smaller than 1, ensuring that the South is the poor country of the two.

where N is the number of varieties produced in the North, and N^* number of varieties produced in the South. Given that labour is the only factor of production, the aggregate resource constraints are effectively labour market clearing conditions. From the aggregate resource constraints, we get the equilibrium number of varieties produced in the North and South:

$$N = \frac{aL\mathcal{P}}{aF + \mathcal{P} + \tau\mathcal{P}^*} \qquad N^* = \frac{aL^*\mathcal{P}^*}{aF + \mathcal{P}^* + \tau\mathcal{P}} \tag{1}$$

Higher trade costs τ imply that fewer varieties can be produced because more resources are required to cover the costs of shipment. Similarly, higher fixed costs imply that fewer varieties can be produced.

We require that countries have a balanced trade account, meaning that the value of imports must equal the value of exports. This balanced trade condition can be expressed as $p^*\mathcal{P}^*N = p\mathcal{P}N^*$. The left-hand side is the value of imports (exports) from the point of view of the South (North). Plugging (1) into the balanced trade condition, we get an expression for relative prices:

$$\frac{p}{p^*} = \frac{L}{\omega L^*} \tag{2}$$

Relative prices are determined by relative labour income. For the full trade equilibrium to exist, we required that relative prices are smaller than the trade costs τ, such that arbitrage is never profitable. Together with the equation above, this implies that relative labour incomes must be smaller than the trade costs $\tau \geq L/\omega L^*$.

Finally, the zero-profit conditions together with the relative price, yield the absolute price level in the North and the South

$$p = \frac{L}{a}\left(\frac{aF + \mathcal{P} + \tau\mathcal{P}^*}{L\mathcal{P} + \omega L^*\mathcal{P}^*}\right) \qquad p^* = p\frac{\omega L^*}{L} = \frac{\omega L^*}{a}\left(\frac{aF + \mathcal{P} + \tau\mathcal{P}^*}{L\mathcal{P} + \omega L^*\mathcal{P}^*}\right) \tag{3}$$

Having solved for the full trade equilibrium, we can work out the conditions under which this equilibrium persists. The first condition is $\tau \geq \frac{L}{\omega L^*}$. The second condition requires that Northern firms are willing to export to the South, given the price level in the South. Specifically, we require that prices in the South at least cover the marginal costs of exported products: $p^* \geq \tau/a$.[20] This constraint can be solved for the level of trade costs required $\tau \leq \frac{\omega L^*}{L}\left(\frac{aF+\mathcal{P}}{\mathcal{P}}\right)$. In this full trade equilibrium, consumers in both countries purchase all

[20]Strictly speaking, we also require that the price level in the North covers the marginal costs of exported products from the South. However, given that prices in the North are strictly larger than in the South, this constraint will always be fulfilled if the other constraint is fulfilled as well.

available goods. Hence, welfare is identical in both countries and given by:

$$U = U^* = \frac{aL\mathcal{P}}{aF + \mathcal{P} + \tau\mathcal{P}^*} + \frac{a\omega L^*\mathcal{P}^*}{aF + \mathcal{P} + \tau\mathcal{P}^*}$$

Consumers in the North pay higher prices for the same products, due to their larger willingness to pay. Consequently, they cover the majority of the total production costs. However, in sum, the higher prices and the higher income in the North offset each other such that welfare is identical in the South. We can state the following proposition

Proposition 1. *Assume a full trade equilibrium. Then, (a) all goods are traded. (b) Relative prices reflect relative purchasing power. (c) Welfare is equalised across countries. (d) A trade liberalisation increases welfare in both countries.*

Proof. In text. ■

4.2 Arbitrage Equilibrium

Full trade ceases to exist when per capita income differences exceed the trade costs. Then, price differences grow too large and arbitrage becomes possible. Firms respond by adjusting their pricing behaviour, or their exporter status, accordingly.

In equilibrium, all firms from the South choose to export. However, in the North, some firms choose not to export their product and only sell in the domestic market, at unconstrained prices. Adopting this strategy implies a smaller market but allows the firm to charge higher prices due to the Northern consumers' higher willingness to pay. The other firms from the North will export their product to the South while charging a reduced price at home. To see why this has to be an equilibrium, consider the case in which all firms export. If all Northern firms were to export as well, then the consumers in the North would not fully exhaust their budget constraint and have an infinitely large willingness to pay for additional products, which cannot be an equilibrium. This is because the reduced price p_T lies below the willingness to pay of the consumers in the North. Consequently, there exist two different prices in the North, one for domestic products and one for traded products. The zero-profit conditions can be written as:

$$p\mathcal{P} = W\left(F + \frac{\mathcal{P}}{a}\right) \tag{4}$$

$$\underbrace{\pi p\mathcal{P} + (1-\pi)\tau p^*\mathcal{P}}_{\text{Expected Revenue from North}} + \underbrace{p^*\mathcal{P}^*}_{\substack{\text{Revenue}\\ \text{from South}}} = W^i\left(F + \frac{\mathcal{P}^i + \tau\mathcal{P}^{-i}}{a}\right) \tag{5}$$

(where $-i =$ South if $i =$ North and vice versa). The ZPC of a domestic Northern producer is (4), while (5) is for a Northern exporter and all Southern firms. Since parallel trade enforcement is uncertain, exporters can charge unconstrained prices with probability π, and with probability $(1-\pi)$, they have to charge the constrained price, to avoid arbitrage by third parties. The left-hand side is (expected) revenue, which is given by the number of consumers in each market multiplied by the sales price (remember that every consumer only ever buys one unit, thus sales volume is given by the population size). The right-hand side is total costs. From (4), the price of non-traded goods in the North is given by:

$$p = \left(\frac{aF + \mathcal{P}}{a\mathcal{P}}\right) \tag{6}$$

Note that the price of non-traded varieties is identical to the price under autarky.

Within both economies, the total number of resources must equal the total number of resources available. These aggregate resource constraints are given by:

$$\mathcal{P}L = N_T\left(F + \frac{\mathcal{P} + \tau\mathcal{P}^*}{a}\right) + N_N\left(F + \frac{\mathcal{P}}{a}\right) \qquad \mathcal{P}^*L^* = N^*\left(F + \frac{\mathcal{P}^* + \tau\mathcal{P}}{a}\right)$$

where N_N N_T are the number of non-traded and traded varieties from the North and N^* is the total number of products from the South.

From the resource constraint in the South, we can determine the number of varieties in the South

$$N^* = \frac{a\mathcal{P}^*L^*}{aF + \mathcal{P}^* + \tau\mathcal{P}} \tag{7}$$

The price level in the South is calculated using (5) and (6):

$$p^* = \frac{(1-\pi)(aF + \mathcal{P}) + \tau\mathcal{P}^*}{a\big(\mathcal{P}^* + (1-\pi)\tau\mathcal{P}\big)} \tag{8}$$

Notice that the prices in the South depend on the enforcement of parallel trade restrictions π. Consequently, also the price of traded varieties in the North, given by $p_T = \tau p^*$, depends on π. Finally, we require a balanced trade condition, such that the value of exports equals the value of imports:

$$\mathcal{P}^* p^* N_T = \Big[(1-\pi)\tau p^* + \pi p\Big]\mathcal{P} N^* \tag{9}$$

The bracket on the right-hand side of the equation denotes the price level in the North, which is a weighted average of the non-traded goods price and the traded goods price.

Given the price levels in the North and South, we can determine the number of traded varieties from the North, N_T, using the balanced trade condition (9):

$$N_T \;=\; \frac{\mathcal{P}}{\mathcal{P}^*}\underbrace{\left[\Big(1-\pi\Big)\tau + \pi\frac{\Big(1+a\frac{F}{\mathcal{P}}\Big)(\mathcal{P}^* + (1-\pi)\tau\mathcal{P})}{(1-\pi)(aF+\mathcal{P}) + \tau\mathcal{P}^*}\right]}_{\text{Terms of Trade: } \mathbb{E}p/p^* \equiv \xi(\tau,\pi)} \;\; \underbrace{\frac{a\mathcal{P}^* L^*}{aF + \mathcal{P}^* + \tau\mathcal{P}}}_{N^*} \tag{10}$$

Finally, from the resource constraint in the North, we get the number of non-traded varieties:

$$N_N = \underbrace{\left(\frac{a\mathcal{P}}{aF+\mathcal{P}}\right)}_{1/p}\left[L - L^* \underbrace{\left(\frac{aF+\mathcal{P}+\tau\mathcal{P}^*}{aF+\mathcal{P}^*+\tau\mathcal{P}}\right)}_{\omega} \;\; \underbrace{\left(\Big(1-\pi\Big)\tau + \pi\frac{\Big(1+a\frac{F}{\mathcal{P}}\Big)(\mathcal{P}^* + (1-\pi)\tau\mathcal{P})}{(1-\pi)(aF+\mathcal{P}) + \tau\mathcal{P}^*}\right)}_{\text{Terms of Trade}}\right] \tag{11}$$

Note that the number of exported varieties, N_T, and the number of varieties not exported, N_N, both depend on the enforcement of parallel trade restrictions π. For non-exported varieties N_N to exist, we require that the relative labour endowment between North and South is sufficiently large. Specifically, we need that $L/\omega L^* > \xi(\tau, \pi)$. In the special case of $\pi = 0$, when parallel trade is fully allowed, this condition simplifies to $L/\omega L^* > \tau$. Additionally, we have the trade condition, i.e. a condition on the trade costs τ, such that arbitrage trade actually happens. For a firm to be willing to export, the export price must at least cover the marginal costs $p^* \geq \tau/a$. Using (8), the trade condition is determined by:

$$\tau \leq \sqrt{\frac{aF+\mathcal{P}}{\mathcal{P}}} \tag{12}$$

5 Welfare Analysis of Arbitrage Equilibrium

5.1 Parallel Trade Policy

Prices. The prices of non-traded goods in the North, are independent of parallel trade restrictions. Due to the existence of arbitrage constraints, traded goods prices no longer depend on labour endowments L and L^* as was the case in the full trade equilibrium, but rather on the enforcement of parallel trade restrictions π. Specifically, stronger IPR enforcement lowers prices in the South and traded goods prices in the North. Hence, there is a pro-competitive effect of parallel trade enforcement on p^*. This is because stricter parallel trade restrictions make exports to the South more attractive for Northern firms. There is increased competition in the Southern goods market, which drives down prices. This is consistent with Dubois, Gandhi and Vasserman (2022), who find that allowing parallel trade mostly leads to price convergence *to the top*, implying that firms will raise prices in the poor country rather than lower prices in the rich country.

Terms of Trade. Let us define the terms of trade as $\mathbf{E}p/p^* \equiv \xi(\tau, \pi) = \frac{\pi p + (1-\pi)\tau p^*}{p^*}$ (see (10). The terms of trade, as they are defined here, are from the point of view of the South. Thus, an increase in ξ captures a shift in the terms of trade in favour of the South.

Non-trade goods prices in the North are independent of π, and all traded goods prices fall in π, hence the terms of trade $\xi(\tau, \pi)$ are increasing in π because the price level in the South falls faster. Northern exports to the South are worth relatively less than Southern exports to the South, which benefits the South. Moreover, relative (and absolute) population sizes do not affect parallel trade policy considerations: a rich (poor) country does not have a different parallel trade policy when trading with a large or small poor (rich) country. More parallel trade restrictions always shift the terms of trade in favour of the South. The North would therefore prefer fewer restrictions on parallel trade.

Export Probability. In the South, the export probability is always one, as every producer is also an exporter. In the North, however, some firms do not export, so the export probability is less than one and defined as $\frac{N_T}{N_T + N_N}$. Intuitively, the number of traded goods increases in π. As the threat of parallel imports diminishes, more firms are willing to export to the South, thereby increasing product variety in the South. This effect is the driving force behind the pro-competitive effect of parallel trade restrictions on prices in the South, as established previously. At the same time, the number of non-traded varieties in the North decreases with stronger parallel trade restrictions. In sum, the reduction in non-traded varieties strictly dominates the increase in traded varieties. Due to the iceberg trade costs τ, a disproportionate amount of resources is required to produce these export goods, which in turn means that fewer resources are available to produce non-traded varieties. Therefore, with a strict parallel trade ban, the North produces fewer goods. Together with the fact that more goods are exported to the South, this implies that the export probability is strictly increasing in π. This is consistent with empirical findings on exports and export probabilities such as Goldberg (2010) or Ivus (2010).

Trade Intensity. Let us define the "trade intensity" ϕ as the ratio between the total value of world trade and world GDP. The total value of trade is given by the sum of all exports $p^*\mathcal{P}^*N_T + \Big[(1-\pi)\tau p^* + \pi p\Big]\mathcal{P}N^*$, while world GDP is $L\mathcal{P} + \omega L^*\mathcal{P}^*$. Using equations (6), (7) and (8), trade intensity is given by

$$\phi = \frac{2a\mathcal{P}\mathcal{P}^*L^*}{aF + \mathcal{P}^* + \tau\mathcal{P}} \frac{1}{L\mathcal{P} + \omega L^*\mathcal{P}^*} \left[\pi \frac{(aF + \mathcal{P})}{a\mathcal{P}} + (1-\pi)\tau \frac{(1-\pi)(aF + \mathcal{P}) + \tau\mathcal{P}^*}{a(\mathcal{P}^* + (1-\pi)\tau\mathcal{P})}\right]$$

Naturally, trade intensity depends on the two trade policy choices, τ and π, and on per capita income in the South L^*. Specifically, higher per capita income in the South increases the intensity of trade. This result goes back to Linder (1961). He postulated that more similar countries should trade more with each other.[21] Finally, tighter parallel trade restrictions increase the intensity of trade. Countries trade the most, both absolutely and in proportion to GDP, when parallel trade is prohibited, ie. when $\pi = 1$. Intuitively, this result follows from the fact that the export probability of the North increases with π.

Proposition 2. *Assume that per capita income differences are large enough to maintain an arbitrage equilibrium. Then, some firms in the North do not export. Stricter IPR enforcement leads to (a) lower prices in the South, (b) an improvement in the terms of trade of the South ξ, (c) increases the export probability from the North and (d) raises trade intensity ϕ.*

Proof. In Appendix A. ■

Welfare. Given that consumers only ever purchase one unit of every good, utility is given

[21] Similarity should be understood as similarity in per capita income or more generally as similarity in demand, which does not necessarily coincide due to non-homothetic demand.

by the sum of varieties consumed. Thus, we have:

$$U = N_N + N^* \left(1 + \xi(\tau, \pi) \frac{\mathcal{P}}{\mathcal{P}^*}\right) \qquad U^* = N^* + N_T = N^* \left(1 + \xi(\tau, \pi) \frac{\mathcal{P}}{\mathcal{P}^*}\right)$$

The effects of parallel trade restrictions on welfare in North and South are given by:

$$\frac{\partial U}{\partial \pi} = \frac{\partial \xi(\tau, \pi)}{\partial \pi} \left(\frac{\partial N_N}{\partial \xi(\tau, \pi)} + \frac{\mathcal{P} N^*}{\mathcal{P}^*} \right) < 0 \qquad \frac{\partial U^*}{\partial \pi} = \frac{\mathcal{P} N^*}{\mathcal{P}^*} \frac{\partial \xi(\tau, \pi)}{\partial \pi} > 0$$

The shift in the terms of trade has a direct impact on utility: although consumers in the North benefit from an increased variety of traded Northern products (N_T increases), the decrease in non-traded varieties is strictly greater. Therefore, consumers in the North are harmed by more parallel trade restrictions. Southern consumers strictly benefit from more parallel trade restrictions due to more imports from the North. The intuition behind this result is simple: the shift in the terms of trade requires the North to export more varieties in order to meet the balanced trade condition, since its export goods have become comparatively less valuable. Due to the iceberg trade costs τ, a disproportionate amount of resources are required to produce these export goods, resulting in the North producing fewer varieties overall.

As a result, there are divergent preferences between North and South on the optimal parallel trade policy. The rich North prefers to allow parallel trade, while the poor South prefers to ban it in order to gain access to more Northern products. This result contrasts with Roy and Saggi (2012), who find that the North also prefers to ban parallel trade when income differences become too large. However, their model does not allow for firm entry and thus misses a key benefit for Northern consumers.

Proposition 3. *In an arbitrage equilibrium, stronger IPR enforcement increases welfare in the South, but decreases it in the North.*

Proof. In Appendix B. ■

5.2 Trade Costs

The effect of the trade costs τ on prices, trade intensity, and ultimately welfare depend strongly on the level of parallel trade restrictions π. To build intuition and to simplify notation, the discussion focuses on the two corner solutions $\pi = 0$ and $\pi = 1$.

Prices. The price of non-traded goods in the North is independent of trade costs. If there are no restrictions on parallel trade $\pi = 0$, prices in the South are given by $p^* = (aF + \mathcal{P} + \tau\mathcal{P}^*)/(a\mathcal{P}^* + a\tau\mathcal{P})$. In this case, a trade liberalisation *increases* prices in the South.[22] Lower trade costs increase the risk of arbitrage for Northern exporters. As a result, fewer of them are willing to export to the South in the first place, which reduces competition in the Southern goods market and thus raises prices. The possibility of parallel trade thus causes trade liberalisation to generate *less* competition instead of more.

If parallel trade is prohibited, $\pi = 1$, the price level in the South will be driven down to just cover the marginal costs of exporters $p = \tau/a$. This is due to the increased competition from Northern firms when the arbitrage threat is absent. Trade liberalisation then produces the standard result of lower prices. In fact, there is a complete pass-through of the cost reduction to the benefit of the consumer. These two results imply that there exists some intermediate level of parallel trade restriction $\hat{\pi}$, such that trade liberalisation does not affect p^*.[23]

Terms of Trade. If there are no restrictions on parallel trade, $\pi = 0$, then a trade liberalisation decreases the terms of trade ξ. This is because trade liberalisation raises prices in the South, as previously established. Thus, exports from the North become more valuable relative to their Southern counterparts. Consequently, the South would prefer to restrict trade by imposing higher trade costs τ, while the North would prefer free trade.

In the opposite case of $\pi = 1$, trade liberalisation benefits the South. Since there is no arbitrage threat in this case, and many Northern firms are willing to export, trade liberalisation lowers prices in the South and shifts the terms of trade in their favour.

Trade Intensity. Given the results just established, it is clear that a trade liberalisation reduces the export probability of the North and consequently also the intensity of trade ϕ if parallel imports are allowed. Lower trade barriers increase the possibility of parallel trade, which discourages some Northern firms from exporting their goods to the South, thus avoiding the threat of arbitrage altogether. This result was established in Föllmi et al. (2018), and we can confirm their finding in our generalised setting.

Only when parallel trade restrictions are sufficiently strong does this turn into the more

[22]This is conditional on the two countries being similarly populous. If the South is significantly larger than the North, the reverse may be true and we have the standard result of a pro-competitive effect of trade liberalisation.

[23]Specifically, this intermediate level of π is determined by $\hat{\pi} = 1 - \frac{\mathcal{P}^*}{\sqrt{\mathcal{P}(aF+\mathcal{P})}}$.

familiar finding that trade liberalisation increases both the probability of exporting and the intensity of trade. Thus, one of the core results of trade theory rests on the assumption that parallel imports do not play an outsized role in trade policy.

Proposition 4. *Assume an arbitrage equilibrium. If IPR enforcement is weak, a trade liberalisation (a) raises prices in the South, (b) lower terms of trade of South and (c) decreases the trade intensity. The opposite holds in case of strong IPR enforcement.*

Proof. In Appendix C. ■

Welfare. First, consider the case with parallel trade, $\pi = 0$, which can be considered the more important case from a legal point of view, since most countries allow it. Welfare in both countries is again given by the sum of all varieties consumed:

$$U(\pi = 0, \tau) = \frac{a\mathcal{P}L}{aF + \mathcal{P}} + \frac{a\mathcal{P}^* L^*(aF + \mathcal{P} - \tau^2\mathcal{P})}{(aF + \mathcal{P})(aF + \mathcal{P}^* + \tau\mathcal{P})} \qquad U^*(\pi = 0, \tau) = \frac{aL^*(\mathcal{P}^* + \tau\mathcal{P})}{aF + \mathcal{P}^* + \tau\mathcal{P}}$$

It is straightforward to verify that $\partial U/\partial\tau < 0$ and $\partial U^*/\partial\tau > 0$. Consumers in the South prefer high trade costs because they get access to more products from the North. The high trade barriers reduce the threat of arbitrage, allowing Northern firms to charge higher prices in the South, making exports more lucrative, so more firms become exporters. This increase in the trade intensity benefits consumers in the South but is detrimental to welfare in the North. The additional exports require many resources for production, leaving fewer for non-traded goods production. In the left part of figures 2a and 2b, when parallel trade restrictions are low, higher trade costs strictly reduce welfare in the North while increasing it in the South.

The other extreme case of no parallel trade, $\pi = 1$, can be considered more relevant from a practical point of view, since the volume of parallel trade is far from being the majority of the total volume of trade and, as it will turn out, is the preferable policy from a global welfare point of view. Then, welfare is given by:

$$U(\pi = 1, \tau) = \frac{a\mathcal{P}L}{aF + \mathcal{P}} \qquad U^*(\pi = 1, \tau) = \frac{aL^*}{\tau}\frac{aF + \mathcal{P} + \tau\mathcal{P}^*}{aF + \mathcal{P}^* + \tau\mathcal{P}}$$

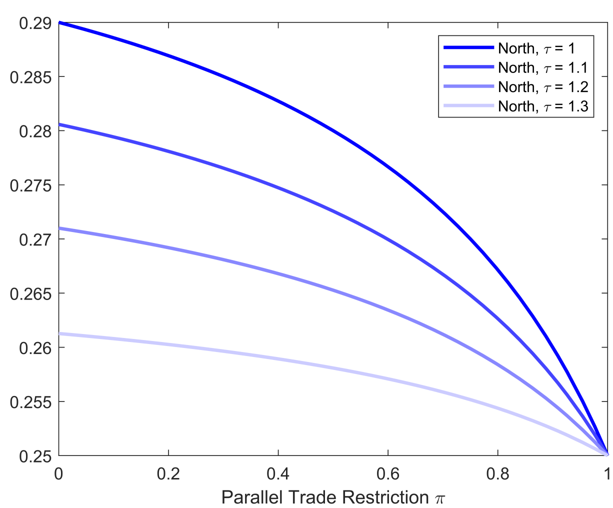
Clearly, the North is indifferent to the level of trade costs in this scenario, the welfare effects of more non-traded goods and fewer traded goods exactly cancel out. In the absence of parallel trade, all firms wishing to export their product already do so,[24] so any change in

[24]Even in this case, not all Northern firms export their goods, because the relative income difference is still too large to generate a full trade equilibrium.

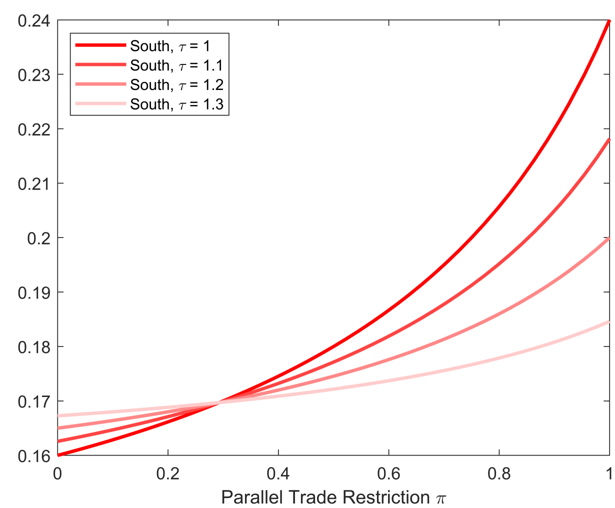
the trade costs does not induce any further reallocation in the export status of Northern firms. See in figure 2a, when parallel trade is ruled out, different levels of trade costs τ have no further effect on welfare. We conclude that the North (weakly) prefers free trade. The South is now also a free trader. When parallel trade is prohibited, prices in the South are determined by marginal costs of the exporters $p^* = \tau/a$ due to the high degree of competition in the market. A reduction in trade costs lowers marginal costs of exporting and thus product prices, which benefits the consumers. In contrast to the case of $\pi = 0$, Northern firms do not reduce exports to the South when trade costs are lower. In figure 2b on the right, the welfare gains of trade liberalisation are shown to be significant. As the South moves from being contra-free trade to pro-free trade as parallel trade restrictions increase, there must exist some intermediate level of parallel trade restriction at which the South is indifferent. In the figure, this turning point is found at about $\pi = 0.3$.

Proposition 5. *Assume an arbitrage equilibrium. If IPR enforcement is weak, a trade liberalisation raises welfare in the North, but lowers it in the South. If IPR enforcement is strong, a trade liberalisation leaves welfare in the North unaffected, while raising it in the South.*

Proof. In Appendix D. ■



(a) Welfare in North



(b) Welfare in South

Figure 2: Welfare Effects of Parallel Trade Restrictions

5.3 Optimal Trade Policy

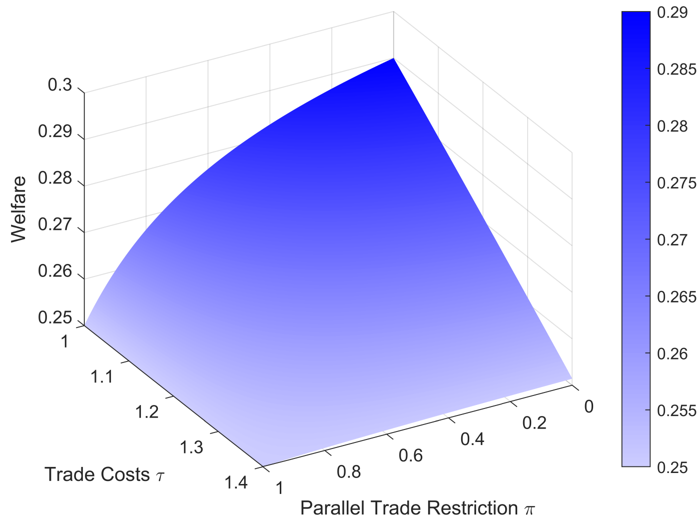
Having established the welfare implications of the two trade policy choices π and τ separately, the question naturally arises as to which mix is optimal from both a country and a global perspective. Figure 3 plots the welfare of the North and the South, given a joint choice of π and τ (note that the axes are inverted in the two figures). Welfare in the North is a well-behaved function of trade policy, with the maximum being at free trade ($\tau = 1$) and allowing parallel trade ($\pi = 0$). Welfare in the South shows the previously established conflictive preference for free trade, depending on the level of parallel trade restrictions. When parallel trade is allowed, raising trade barriers can mitigate the welfare losses to consumers. However, they are best off when parallel trade is banned and free trade prevails.

From a global welfare perspective, therefore, the key question concerns parallel trade policy. Suppose there exists a world government (or a world trade regulator), that wanted to maximise aggregate welfare of the North and the South. What level of parallel trade restriction would it choose? The world government chooses π to maximise aggregate welfare:

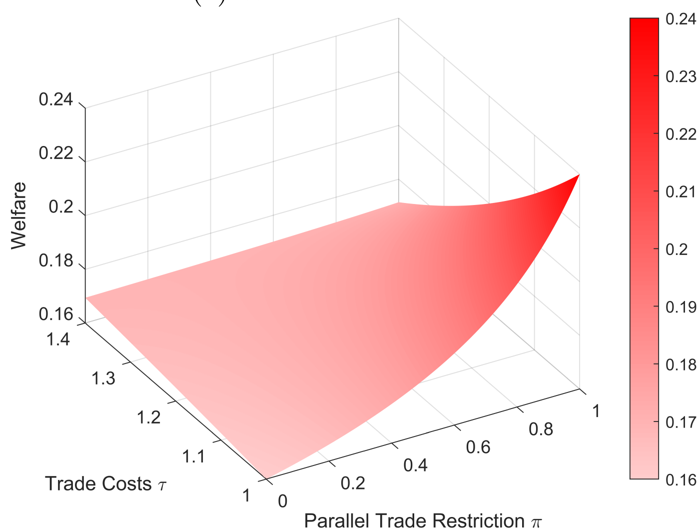
$$\frac{\partial}{\partial \pi} V(\pi) = \frac{\partial}{\partial \pi}\Big[\mathcal{P}U(\pi) + \mathcal{P}^* U^*(\pi)\Big] = \underbrace{\frac{\partial \xi}{\partial \pi}}_{>0} \left[\underbrace{\mathcal{P}\frac{\partial N_N}{\partial \xi}}_{<0} + \underbrace{\frac{\mathcal{P}(\mathcal{P}+\mathcal{P}^*)}{\mathcal{P}^*} N^*}_{>0}\right]$$

It can be shown that the world government would choose to completely restrict parallel trade, and consequently, to allow free trade (both countries have a preference for free trade when parallel trade is banned). Although the North would strictly prefer a different parallel trade policy, the welfare gains in the South outweigh the losses in the North. Note that the result is independent of the relative population sizes of North and South.

Thus, from a global welfare perspective, price discrimination by limiting parallel trade is preferable to a globally uniform price, because it allows more markets and consumers to be served. The proof is in Appendix E. We can confirm and generalise the result of Malueg and Schwartz (1994), who proved this result in a partial equilibrium setting with monopoly production. This result is in line with the literature on third-degree price discrimination and social welfare, e.g. in Schmalensee (1981), Varian (1985) and Schwartz (1990). In the case of both markets being served in the uniform price regime, Schmalensee (1981) showed that a necessary condition for welfare-enhancing price discrimination is an increase in total output. Varian (1985) proved that price discrimination raises welfare if it allows an otherwise unserved market to be served as well. This is the case in our setting: under price discrimination, the poor country is (partially) not served by producers from the rich country, causing export zeroes.[25]



(a) Welfare in North



(b) Welfare in South

Figure 3: Welfare Effects of Trade Policy between North and South.

[25]This result is not driven by the assumption of consumption indivisibility. Cowan (2016) proves this result holds for all demand functions of the class of logistic demand, which in all generality is given by: $q(p) = (1 + \exp(\frac{p-a}{b}))^{-1}$, where $b > 0$ is a scale parameter and $a > 0$ a location parameter. Varian's quasi-linear demand function and our 0/1-preferences both belong to this class. In the limit of $b \to 0^+$, the logistic demand becomes a step-function with a reservation price $p = a$. Hence, our result on optimal (parallel) trade policy is robust to more general specifications of utility, involving both intensive and extensive margins of consumption. The logistic demand function has an extensive margin because the reservation price is finite and an intensive margin because consumers can choose any value between 0 and 1.

5.4 Further Analysis

Big or Small Trade Partners? In case the South becomes relatively more populous, its terms of trade are better compared to a less populous poor country. This is due to the price effect. A larger population means that the fixed costs of production can be spread over more people, so goods can be sold at a lower price. Northern exporters have to match this price, so they also lower their prices. Since Southern exporters can still sell their goods in the North at the old price, this means that the terms of trade have shifted in favour of the South. Note that, if $\pi = 1$, this effect disappears. This is because, if parallel trade is fully restricted, the price in the South p^* has already been driven down to the marginal costs of exporting $p^* = \tau/a$, due to the effect of increased competition from Northern firms in the South. At this point, no further reduction in p^* is possible, given Northern firms exporting incentives.

Note that the income per capita difference has not changed. The comparison is between a large and a small, equally poor, Southern country. We have already shown that $\frac{\partial(N_T+N_N)}{\partial\xi} \leq 0$, i.e., the number of varieties produced in the North decreases in the terms of trade. Consequently, the North would strictly prefer many, less populous trading partners to a few, but very populous ones.

Income Convergence. What happens if we observe convergence between rich and poor, implying that inequality between North and South is decreasing? It is straightforward to verify that the terms of trade are independent of labour endowment $\frac{\partial\xi}{\partial L^*} = 0$. Poorer countries do not *per se* have worse terms of trade than richer ones. This is because prices, both in the North and South, are independent of labour endowments in an arbitrage equilibrium. So there is no change in the relative price, i.e. the terms of trade.

However, as the South becomes richer, the more varieties it produces. The absence of an intensive margin implies that all extra income is spent on additional varieties, thereby increasing N^*. Both the North and the South benefit from this, *ceteris paribus.* A richer South also implies that the North exports more varieties. This follows from the balanced trade condition: If the South exports more varieties, then the North must also export more varieties (holding prices fixed), so that the value of imports equals the value of exports. These two facts together imply that welfare in the South U^* is strictly increasing in its labour endowment. As for the North, we have already established that an increase in traded varieties necessarily implies a decrease in non-traded varieties. The conversion rate is greater than one, meaning that the decrease in non-traded varieties is greater than the increase in traded varieties, implying that the total number of varieties produced in the North is now smaller. However, taking into account the increased variety from the South, the North also benefits from having a richer trading partner.[26] This result is independent of the degree of parallel trade restriction π.

The Divergence. As a complement to the previous section, we examine the welfare implications of the North's labour endowment L. Under CES preferences, we would expect the welfare implications to be symmetric, both countries should benefit if one of the two becomes richer (in terms of higher labour endowment). This is, after all, what neoclassical trade theory is all about. But it turns out that this is not the case in our model.

Clearly, the number of varieties produced in the South, the number of traded varieties from the North, and the terms of trade are independent of L. Thus, welfare in the South does not change in response to a richer North. All additional resources from the North go into the production of new, non-traded varieties N_N. Note that the increase in non-traded varieties is inversely related to the price of these goods. Intuitively, if goods are more expensive, the same amount of resources will buy fewer goods.

Consequently, welfare in the North rises as a result of an increase in labour endowment through an increased variety of non-traded goods. Consumers in the South are equally well off as before because the number of traded varieties has not changed. Our model therefore predicts asymmetric welfare gains: If the South gets richer (increase in L^*), both countries benefit. However, if the North gets richer (increase in L), only the North benefits. This is reminiscent of the Prebisch-Singer hypothesis, according to which the South faces ever-worsening terms of trade because the relative price of commodities to manufactured goods falls over time due to different price elasticities of demand.[27]

Proposition 6. *Assume an arbitrage equilibrium. (a) The rich North prefers to have many, but small trading partners. (b) Income convergence (i.e. richer South) benefits both countries, but the South more so. (c) Income divergence (i.e. richer North) benefits only the North.*

Proof. In Appendix F. ■

[26]This result is consistent with empirical estimates of the welfare effects of the China shock. See for example Arkolakis, Costinot and Rodríguez-Clare (2012), Caliendo, Dvorkin and Parro (2019) or Galle, Rodríguez-Clare and Yi (2023).

[27]See Harvey et al. (2010) or Singer (1950). Of course, in our simple framework, the mechanism is not the same as in the Prebisch-Singer hypothesis. The terms of trade are unaffected by economic divergence, the effect works through the "new gains from trade", namely increasing product variety in the North.

6 Conclusion

This article develops a model of international trade in which an importer's per capita income plays a crucial role in determining export zeros and the prices of exported goods. This phenomenon arises from a demand-side effect: consumers in lower-income countries are less willing to pay for differentiated products than consumers in higher-income countries. As a result, firms in wealthier regions have less incentive to export to destinations in lower-income regions. Our model attributes the occurrence of export zeros to non-homothetic preferences, which differs from the conventional approach that emphasises firm heterogeneity.

The presence of arbitrage threats reduces the probability of exports. Firms from the rich North face a trade-off between market size and pricing power: (1) refrain from exporting to the South and charge higher prices in the North, or (2) sell in all markets but with limited price discrimination. In this case, the South prefers high barriers to trade. High trade costs reduce the arbitrage threat for Northern firms (because making reimports more expensive), thereby increasing exports to the benefit of Southern consumers. This creates a paradoxical situation where higher trade costs actually lead to higher trade volumes and provides a rationale as to why rich and poor countries might have opposing stances on trade liberalisation.

The preferences of the North and South regarding parallel trade policy diverge. The South favours restrictions on parallel trade in order to gain access to more Northern goods. Conversely, the North prefers to allow parallel trade because it lowers prices and expands the range of available products. From a global welfare perspective, banning parallel trade and eliminating trade costs emerges as the optimal trade policy mix.

References

Acemoglu, Daron, David Autor, David Dorn, Gordon H Hanson, and Brendan Price, "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics*, 2016, *34* (1).

Adao, Rodrigo, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz, "Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence," *The Quarterly Journal of Economics*, 2022, *137* (3), 1553–1614.

Alessandria, George and Joseph P Kaboski, "Pricing-to-Market and the Failure of Absolute PPP," *American Economic Journal: Macroeconomics*, 2011, *3* (1), 91–127.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings, "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*, 2019, *86* (6).

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 2008, *98* (5), 1998–2031.

Auer, Raphael A, Thomas Chaney, and Philip Sauré, "Quality Pricing-to-Market," *Journal of International Economics*, 2018, *110.*

Autor, David, David Dorn, and Gordon H Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 2013, *103* (6).

Behrens, Kristian and Yasusada Murata, "Globalization and Individual Gains from Trade," *Journal of Monetary Economics*, 2012, *59* (8).

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott, "Firms in International Trade," *Journal of Economic Perspectives*, 2007, *21* (3).

_ , _ , _ , and _ , "The Margins of US Trade," *American Economic Review*, 2009, *99* (2).

Bond, Eric W and Kamal Saggi, "Compulsory Licensing, Price Controls, and Access to Patented Foreign Products," *Journal of Development Economics*, 2014, *109.*

_ and _ , "Compulsory Licensing and Patent Protection: a North-South Perspective," *The Economic Journal*, 2018, *128* (610).

_ and _ , "Patent Protection in Developing Countries and Global Welfare: WTO Obligations versus Flexibilities," *Journal of International Economics*, 2020, *122.*

Borusyak, Kirill and Xavier Jaravel, "The Distributional Effects of Trade: Theory and Evidence from the United States," *NBER Working Paper No. 28957*, 2021.

Burgess, Matthew and Lewis Evans, "Parallel Importation and Service Quality: An Empirical Investigation of Competition between DVDs and Cinemas in New Zealand," *Journal of Competition Law and Economics*, 2005, *1* (4), 747–770.

Calboli, Irene, "Intellectual Property Exhaustion and Parallel Imports of Pharmaceuticals: A Comparative and Critical Review," *Access to Medicines and Vaccines*, 2022, p. 31.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4).

Corsetti, Giancarlo and Luca Dedola, "A Macroeconomic Model of International Price Discrimination," *Journal of International Economics*, 2005, *67* (1).

Cowan, Simon, "Welfare-Increasing Third-Degree Price Discrimination," *The RAND Journal of Economics*, 2016, *47* (2).

Danzon, Patricia M and Andrew J Epstein, "Effects of Regulation on Drug Launch and Pricing in Interdependent Markets," *NBER Working Paper No. 14041*, 2008.

De Loecker, Jan and Jan Eeckhout, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2).

Demidova, Svetlana, "Trade Policies, Firm Heterogeneity, and Variable Markups," *Journal of International Economics*, 2017, *108.*

Dubois, Pierre and Morten Sæthre, "On the Effect of Parallel Trade on Manufacturers' and Retailers' Profits in the Pharmaceutical Sector," *Econometrica*, 2020, *88* (6), 2503–2545.

_ , Ashvin Gandhi, and Shoshana Vasserman, "Bargaining and International Reference Pricing in the Pharmaceutical Industry," *NBER Working Paper No. 30053*, 2022.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 2015, *105* (10).

Fajgelbaum, Pablo D and Amit K Khandelwal, "Measuring the Unequal Gains from Trade," *The Quarterly Journal of Economics*, 2016, *131* (3), 1113–1180.

Fajgelbaum, Pablo, Gene M. Grossman, and Elhanan Helpman, "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 2011, *119* (4).

_ , Gene M Grossman, and Elhanan Helpman, "A Linder Hypothesis for Foreign Direct Investment," *The Review of Economic Studies*, 2015, *82* (1).

Falkinger, Josef, "An Engelian Model of Growth and Innovation with Hierarchic Consumer Demand and Unequal Incomes," *Ricerche Economiche*, 1994, *48* (2), 123–139.

Feenstra, Robert C, "Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," *Journal of International Economics*, 2018, *110*, 16–27.

Fitzgerald, Doireann and Stefanie Haller, "Pricing-to-Market: Evidence from Plant-Level Prices," *Review of Economic Studies*, 2014, *81* (2), 761–786.

Föllmi, Reto and Josef Zweimüller, "Income Distribution and Demand-Induced Innovations," *The Review of Economic Studies*, 2006, *73* (4).

_ , Björn Hartmann, and Josef Zweimüller, "Income Distribution, Welfare, and the Patterns of Trade," *CESifo Working Paper No. 12107*, 2025.

_ , Christian Hepenstrick, and Josef Zweimüller, "International Arbitrage and the Extensive Margin of Trade between Rich and Poor Countries," *The Review of Economic Studies*, 2018, *85* (1), 475–510.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi, "Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade," *The Review of Economic Studies*, 2023, *90* (1), 331–375.

Ganslandt, Mattias and Keith E Maskus, "Parallel Imports and the pricing of pharmaceutical products: evidence from the European Union," *Journal of Health Economics*, 2004, *23* (5), 1035–1057.

Goldberg, Pinelopi Koujianou, "Intellectual Property Rights Protection in Developing Countries: The Case of Pharmaceuticals," *Journal of the European Economic Association*, 2010, *8* (2-3), 326–353.

Gopinath, Gita, Oleg Itskhoki, and Roberto Rigobon, "Currency Choice and Exchange Rate Pass-

Through," *American Economic Review*, 2010, *100* (1), 304–336.

Grossman, Gene M and Edwin L-C Lai, "Parallel Imports and Price Controls," *The Rand Journal of Economics*, 2008, *39* (2), 378–402.

Hallak, Juan Carlos, "A Product-Quality View of the Linder Hypothesis," *The Review of Economics and Statistics*, 2010, *92* (3), 453–466.

Harvey, David I, Neil M Kellard, Jakob B Madsen, and Mark E Wohar, "The Prebisch-Singer Hypothesis: Four Centuries of Evidence," *The Review of Economics and Statistics*, 2010, *92* (2), 367–377.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 2008, *123* (2).

Hummels, David and Peter J Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, 2005, *95* (3), 704–723.

Ivus, Olena, "Do Stronger Patent Rights Raise High-Tech Exports to the Developing World?," *Journal of International Economics*, 2010, *81* (1), 38–47.

_ **and Edwin L-C Lai**, "Patent Exhaustion Regime and International Production Sharing: Winners and Losers?," *CESifo Working Paper No. 6644*, 2017.

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

_ , "Pricing to Market when the Exchange Rate Changes," *NBER Working Paper No. 1926*, 1986.

Lanjouw, Jenny, "Patents, Price Controls, and Access to New Drugs: How Policy Affects Global Market Entry," *NBER Working Paper No. 11321*, 2005.

Li, Changying and Keith E Maskus, "The Impact of Parallel Imports on Investments in Cost-Reducing Research and Development," *Journal of International Economics*, 2006, *68* (2), 443–455.

Linder, Staffan Burenstam, *An Essay on Trade and Transformation*, Almqvist & Wiksell Stockholm, 1961.

Lutz, Matthias, "Pricing in Segmented Markets, Arbitrage Barriers, and the Law of One Price: Evidence from the European Car Market," *Review of International Economics*, 2004, *12* (3), 456–475.

Malueg, David A and Marius Schwartz, "Parallel Imports, Demand Dispersion, and International Price Discrimination," *Journal of International Economics*, 1994, *37* (3-4).

Maskus, Keith E, "Parallel Imports," *The World Economy*, 2000, *23* (9).

Matsuyama, Kiminori, "A Ricardian Model with a Continuum of Goods under Nonhomothetic Preferences: Demand Complementarities, Income Distribution, and North-South Trade," *Journal of Political Economy*, 2000, *108* (6), 1093–1120.

McFadden, Daniel, "Modelling the Choice of Residential Location," in Anders Karlqvist, Lars Lundqvist, Folcke Snickars, and Jorgen Weibull, eds., *Spatial Interaction Theory and Planning Models*, Amsterdam, 1978.

Melitz, Marc J, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

_ **and Gianmarco I. P. Ottaviano**, "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 2008, *75.*

Mongey, Simon and Michael E Waugh, "Discrete Choice, Complete Markets, and Equilibrium," *NBER Working Paper No. 32135*, 2024.

Murphy, Kevin M, Andrei Shleifer, and Robert Vishny, "Income Distribution, Market Size, and Industrialization," *The Quarterly Journal of Economics*, 1989, *104* (3), 537–564.

NERA, "The Economic Consequences of the Choice of Regime in the Area of Trademarks," Technical Report, Final Report for DG XV of the European Commission 1999.

Roy, Santanu and Kamal Saggi, "Equilibrium Parallel Import Policies and International Market Structure," *Journal of International Economics*, 2012, *87* (2), 262–276.

Schmalensee, Richard, "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *The American Economic Review*, 1981, *71* (1), 242–247.

Schwartz, Marius, "Third-Degree Price Discrimination and Output: Generalizing a Welfare Result," *The American Economic Review*, 1990, *80* (5), 1259–1262.

Simonovska, Ina, "Income Differences and Prices of Tradables: Insights from an Online Retailer," *The Review of Economic Studies*, 2015, *82* (4), 1612–1656.

Singer, H.W., "The Distribution of Gains between Investing and Borrowing Countries," *American Economic Review*, 1950, *40* (2).

Varian, Hal R, "Price Discrimination and Social Welfare," *The American Economic Review*, 1985, *75* (4).

Waugh, Michael E, "Heterogeneous Agent Trade," *NBER Working Paper No. 31810*, 2023.

WTO, "Agreement on Trade-Related Aspects of Intellectual Property Rights," 1994. Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994).

Yeung, Godfrey and Vincent Mok, "Manufacturing and Distribution Strategies, Distribution Channels, and Transaction Costs: The Case of Parallel Imported Automobiles," *Managerial and Decision Economics*, 2013, *34* (1), 44–58.

Appendix

A Proof of Proposition 2

Part (a): This follows from differentiating p^* with respect to π

$$\frac{\partial p^*}{\partial \pi} = \frac{\partial}{\partial \pi}\left[\frac{(1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*}{a\big(\mathcal{P}^*+(1-\pi)\tau\mathcal{P}\big)}\right] = \frac{-\mathcal{P}^*(aF+\mathcal{P}-\tau^2\mathcal{P})}{a(\mathcal{P}^*+(1-\pi)\tau\mathcal{P})^2} < 0$$

Part (b):

$$\frac{\partial \xi(\tau,\pi)}{\partial \pi} = \frac{\mathcal{P}^*}{\mathcal{P}}\frac{(aF+\mathcal{P}+\tau\mathcal{P}^*)(aF+\mathcal{P}-\tau^2\mathcal{P})}{\Big((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*\Big)^2} > 0$$

Part (c):

$$\frac{\partial(N_T+N_N)}{\partial \pi} = \frac{a\omega\mathcal{P}^*L^*(aF+\mathcal{P}-\tau^2\mathcal{P})(1-aF-\mathcal{P}-\tau\mathcal{P}^*)}{(aF+\mathcal{P})\big((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*\big)^2} < 0$$

Part (d): Note that $\text{sign}(\partial\phi/\partial\pi) = \text{sign}(\partial\log\phi/\partial\pi)$. Then, we can write:

$$\frac{\partial \log \phi}{\partial \pi} = \frac{\mathcal{P}^{*2}(aF+\mathcal{P}-\tau^2\mathcal{P})}{a\mathcal{P}(\mathcal{P}^*+(1-\pi)\tau\mathcal{P})} > 0$$

where the inequality sign follows from the trade condition (12).

B Proof of Proposition 3

By appendix A, we already know that $\frac{\partial(N_T+N_N)}{\partial\pi} < 0$. Hence, what remains to be shown is that:

$$\frac{\partial N_T}{\partial \pi} = \frac{\mathcal{P}N^*}{\mathcal{P}^*}\frac{\partial \xi}{\partial \pi} = \frac{a\omega\mathcal{P}^*L^*(aF+\mathcal{P}-\tau^2\mathcal{P})}{((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*)^2} > 0$$

$$\frac{\partial N_N}{\partial \pi} = -\frac{a\omega\mathcal{P}L^*}{aF+\mathcal{P}}\frac{\partial \xi}{\partial \pi} = -\frac{a\omega\mathcal{P}^*L^*}{aF+\mathcal{P}}\frac{(aF+\mathcal{P}+\tau\mathcal{P}^*)(aF+\mathcal{P}-\tau^2\mathcal{P})}{((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*)^2} < 0$$

$$\frac{\partial N^*}{\partial \pi} = 0$$

C Proof of Proposition 4

Part (a): The two corner solutions are given by $\left.\frac{\partial p^*}{\partial \tau}\right|_{\pi=0} = \frac{1}{a}\left[\frac{\mathcal{P}^{*2}-\mathcal{P}^2-aF\mathcal{P}}{(\mathcal{P}^*+\tau\mathcal{P})^2}\right] \gtrless 0$ and $\left.\frac{\partial p^*}{\partial \tau}\right|_{\pi=1} = \frac{1}{a} > 0$. For the case of $\pi = 0$, the derivative is negative as long as the relative population sizes are not too different. The cross-derivative is

$$\frac{\partial^2 p^*}{\partial\tau\partial\pi} = \frac{1}{a}\left[\frac{2\mathcal{P}\mathcal{P}^*\big((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*\big)}{\big(\mathcal{P}^*+(1-\pi)\tau\mathcal{P}\big)^3}\right] > 0$$

Part (b): The two corner solutions are $\left.\frac{\partial \xi}{\partial \tau}\right|_{\pi=0} = \frac{(aF+\mathcal{P})(aF+\mathcal{P}+2\tau\mathcal{P}^*)+\tau^2\mathcal{P}^{*2}}{(aF+\mathcal{P}+\tau\mathcal{P}^*)^2} > 0$ and $\left.\frac{\partial \xi}{\partial \tau}\right|_{\pi=1} = -\frac{(aF+\mathcal{P})}{\tau^2\mathcal{P}} < 0$. The cross-derivative is given by:

$$\frac{\partial^2 \xi}{\partial \tau \partial \pi} = \frac{\mathcal{P}^{*2}\big((aF+\mathcal{P}-\tau^2\mathcal{P})((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*)-2(aF+\mathcal{P})(aF+\mathcal{P}+\tau\mathcal{P}^*)(1+\tau(1-\pi))\big)}{P\big((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*\big)^3} < 0$$

Part (c): The case of $\pi = 0$ was already proven in proposition 3 of FHZ. For $\pi = 1$, we can show that:

$$\left.\frac{\partial \phi}{\partial \tau}\right|_{\pi=1} = -\frac{2\mathcal{P}^2 L^*(aF+\mathcal{P})}{(L+L^*)(aF+(1+\tau)\mathcal{P})^2} < 0$$

D Proof of Proposition 5

The welfare effects of trade costs τ in the case of no parallel trade restrictions are given by:

$$\frac{\left.\partial U(\pi,\tau)\right|_{\pi=0}}{\partial \tau} = -\frac{a\mathcal{P}^* L^*}{aF+\mathcal{P}}\left[\frac{\mathcal{P}\big(\tau^2\mathcal{P}+2a\tau F+2\tau\mathcal{P}^*+aF+\mathcal{P}\big)}{\big(aF+\mathcal{P}^*+\tau\mathcal{P}\big)^2}\right] < 0$$

$$\frac{\left.\partial U^*(\pi,\tau)\right|_{\pi=0}}{\partial \tau} = \frac{a^2 F\mathcal{P}L^*}{(aF+\mathcal{P}^*+\tau\mathcal{P})^2} > 0$$

In the case of forbidden parallel trade, the welfare effects of τ are given by:

$$\frac{\left.\partial U^*(\pi,\tau)\right|_{\pi=1}}{\partial \tau} = -\frac{aL^*\big(\tau^2\mathcal{P}\mathcal{P}^*+2a\tau F\mathcal{P}+2\tau\mathcal{P}^2+a^2F^2+aF\mathcal{P}^*+aF\mathcal{P}+\mathcal{P}\mathcal{P}^*\big)}{\tau^2(aF+\mathcal{P}^*+\tau\mathcal{P})^2} < 0$$

In the general case, the effect of trade costs on welfare is:

$$\frac{\partial U}{\partial \tau} = \frac{\partial N^*}{\partial \tau}\Big(1+\xi\frac{\mathcal{P}}{\mathcal{P}^*}\Big) + \frac{\partial \xi}{\partial \tau}\Big(\frac{\partial N_N}{\partial \xi} + N^*\frac{\mathcal{P}}{\mathcal{P}^*}\Big) + \frac{\partial N_N}{\partial \omega}\frac{\partial \omega}{\partial \tau} \leq 0$$

$$\frac{\partial U^*}{\partial \tau} = \frac{\partial N^*}{\partial \tau}\Big(1+\xi\frac{\mathcal{P}}{\mathcal{P}^*}\Big) + \frac{\partial \xi}{\partial \tau}N^*\frac{\mathcal{P}}{\mathcal{P}^*} \gtreqless 0$$

where the inequalities follow from appendix C. Hence, what remains to be shown is $\partial^2 U/\partial\tau\partial\pi > 0$ and $\partial^2 U^*/\partial\tau\partial\pi < 0$. From (7), (10) and (11), we get:

$$\frac{\partial^2 N_T}{\partial \tau \partial \pi} = \frac{a\mathcal{P}L^*}{(aF+\mathcal{P}^*+\tau\mathcal{P})^2}\Big((aF+\mathcal{P}^*+\tau\mathcal{P})\xi_{\tau,\pi} - \mathcal{P}\xi_\pi\Big) < 0$$

$$\frac{\partial^2 N_N}{\partial \tau \partial \pi} = -\frac{a\mathcal{P}L^*}{aF+\mathcal{P}}\Big(\omega_\tau \xi_\pi + \omega\xi_{\tau,\pi}\Big) > 0$$

where $\xi_\pi = \frac{\partial \xi}{\partial \pi} > 0$ and $\xi_{\tau,\xi} = \frac{\partial^2 \xi}{\partial \tau \partial \pi} < 0$. From this, it can be verified that $\frac{\partial^2 (N_N+N_T)}{\partial \tau \partial \pi} > 0$, which concludes the proof.

E Optimal Trade Policy

Given that $\frac{\partial \xi}{\partial \pi} > 0$ by appendix A, all that remains to be shown is that

$$\begin{aligned}
\mathcal{P}\frac{\partial N_N}{\partial \xi} + \frac{\mathcal{P}(\mathcal{P}+\mathcal{P}^*)}{\mathcal{P}^*}N^* &= -\frac{a\mathcal{P}L^*}{aF+\mathcal{P}}\frac{aF+\mathcal{P}+\tau\mathcal{P}^*}{aF+\mathcal{P}^*+\tau\mathcal{P}} + \frac{\mathcal{P}+\mathcal{P}^*}{\mathcal{P}^*}\frac{a\mathcal{P}^*L^*}{aF+\mathcal{P}^*+\tau\mathcal{P}} \\
&= \frac{aL^*}{(aF+\mathcal{P}^*+\tau\mathcal{P})(aF+\mathcal{P})}\Big[-\mathcal{P}(aF+\mathcal{P}+\tau\mathcal{P}^*) + (\mathcal{P}+\mathcal{P}^*)(aF+\mathcal{P})\Big] \\
&= \frac{a\mathcal{P}^*L^*}{(aF+\mathcal{P}^*+\tau\mathcal{P})(aF+\mathcal{P})}\Big[aF+\mathcal{P}-\tau\mathcal{P}\Big] > 0
\end{aligned}$$

where the inequality follows from the trade condition (12).

F Proof Proposition 6

Part (a):

$$\frac{\partial \xi}{\partial P^*/P} = \frac{(1-\pi)\Big(\frac{aF+P}{P}\Big)\Big(\frac{aF+P}{P}-\tau^2\Big)}{\Big((1-\pi)\Big(\frac{aF+P}{P}\Big)+\tau\frac{\mathcal{P}^*}{\mathcal{P}}\Big)^2} \geq 0$$

Part (b): It can be shown that $\frac{\partial N^*}{\partial L^*} = \frac{a\mathcal{P}^*}{aF+\mathcal{P}^*+\tau\mathcal{P}} > 0$ and $\frac{\partial N_T}{\partial L^*} = \frac{aP\xi}{aF+\mathcal{P}^*+\tau\mathcal{P}} > 0$, which ensures that $\partial U^*/\partial L^* > 0$. For the North, we additionally have $\frac{\partial N_N}{\partial L^*} = -\Big(\frac{a\mathcal{P}}{aF+\mathcal{P}}\Big)\Big(\frac{aF+\mathcal{P}+\tau\mathcal{P}^*}{aF+\mathcal{P}^*+\tau\mathcal{P}}\Big)\xi < 0$. However, in sum:

$$\frac{\partial U}{\partial L^*} = a\mathcal{P}^*\left[\frac{((1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*)(aF+\mathcal{P}) - \tau\mathcal{P}\Big(\big(aF+\mathcal{P}\big)\big(\pi\frac{\mathcal{P}^*}{\mathcal{P}}+(1-\pi)\tau\big)+(1-\pi)\mathcal{P}^*\tau^2\Big)}{(1-\pi)(aF+\mathcal{P})+\tau\mathcal{P}^*}\right] \geq 0$$