International arbitrage and the extensive margin of trade between rich and poor countries*

Reto Foellmi[†], Christian Hepenstrick[‡], Josef Zweimüller[§]

January 16, 2017

forthcoming in: *Review of Economic Studies*

Abstract

We incorporate consumption indivisibilities into the Krugman (1980) model and show that an importer's per capita income becomes a primary determinant of "export zeros". Households in the rich North (poor South) are willing to pay high (low) prices for consumer goods; hence unconstrained monopoly pricing generates arbitrage opportunities for internationally traded products. Export zeros arise because some northern firms abstain from exporting to the South, to avoid international arbitrage. Rich countries benefit from a trade liberalization, while poor countries lose. These results hold also under more general preferences with both extensive and intensive consumption margins. We show that a standard calibrated trade model (that ignores arbitrage) generates predictions on relative prices that violate no-arbitrage constraints in many bilateral trade relations. This suggests that international arbitrage is potentially important.

JEL classification: F10, F12, F19

Keywords: Non-homothetic preferences, parallel imports, arbitrage, extensive margin, export zeros

*We thank Sandro Favre, Patrick DeCaes, Stefan Legge, Damian Osterwalder and Marco Schmid for excellent research assistance. We are grateful to Claudia Bernasconi for generous help in preparing and checking the Comtrade data and Michael Waugh for access to the data used in section 7 of the paper. We are particularly grateful to two anonymous referees, Olivier Blanchard, Josef Falkinger, Gene Grossman, Wilhelm Kohler and Marc Melitz for encouraging and very helpful comments. Financial support by the Swiss National Science Foundation (SNF) through the Sinergia-projects CRSI11_132272 "Economic Inequality and International Trade" and CRSII1_154446 "Inequality and Globalization: Demand versus Supply Forces and Local Outcomes" is gratefully acknowledged.

[†]University of St. Gallen and CEPR, SIAW-HSG, Bodanstrasse 8, CH-9000 St.Gallen, Tel: +41 71 224 22 69, Fax: +41 71 224 22 98, email: reto.foellmi@unisg.ch.

[‡]Swiss National Bank and University of Zurich, Boersenstrasse 15 Postfach, CH-8022 Zuerich, e-mail: Christian.Hepenstrick@snb.ch

[§]University of Zurich and CEPR, Department of Economics, Schoenberggasse 1, CH - 8001 Zuerich, Tel: +41 44 634 37 24, Fax: +41 44 634 49 07, e-mail: josef.zweimueller@econ.uzh.ch. Josef Zweimüller is also associated with CESifo and IZA.

1 Introduction

We study a model of international trade in which an importer's per capita income is a primary determinant of the extensive margin of international trade. Two facts motivate our analysis. *First*, there are huge differences in per capita incomes across the globe and these differences may have important consequences for patterns of international trade via the demand side. *Second*, per capita incomes of destination countries correlate strongly with the extensive margin of trade. In 2007, for example, the probability that the US exports a given HS 6-digit product to a high-income country was 63.4 percent, while the export probabilities to upper-middle, lower-middle, and low-income destinations were only 48.8 percent, 36.6 percent, and 13.6 percent, respectively. Furthermore, also US firm-level data show a positive correlation between export probabilities and destinations' per capita incomes (Bernard, Jensen, and Schott 2009).

Recent research has emphasized the presence of "zeros" in bilateral trade data, see e.g. Helpman, Melitz, and Rubinstein (2007) at the country pair level; Hummels and Klenow (2005) at the product level; and Bernard, Jensen, Redding, and Schott (2007) at the firm level. However, the literature did not systematically explore the role of per capita incomes. The standard explanation for export zeros relies on heterogeneous firms and fixed export-market entry costs (Melitz 2003, Chaney 2008, Arkolakis, Costinot and Rodriguez-Clare 2012). Export zeros arise if a firm's marginal costs are too high and/or export market size (in terms of aggregate GDP) is too low to cover the fixed export costs.[1] Importantly, there is no separate role for per capita incomes. Because of homothetic preferences it is irrelevant whether a given aggregate GDP arises from a large population and a low per capita income, or vice versa.[2]

Our paper provides an alternative approach to explain export zeros in which the demand side plays the crucial role. In particular, we elaborate the idea that low per capita incomes are associated with low willingnesses to pay for differentiated products, so that firms abstain from exporting to poor destinations. Our emphasis on the demand channel does not only lead to new predictions on trade patterns. It has also important implications for consumer welfare. Our model predicts that poor countries may lose from a trade liberalization, while rich countries always gain. This is different from standard models where gains from trade are more evenly distributed and all trading partners typically benefit from a trade liberalization.

We start out with a simple model that is identical to the basic Krugman (1980) framework, except that consumer goods are indivisible and households purchase either one unit of a particular product or do not purchase it at all. Such "0-1" preferences generate, in a straightforward way, a situation where a household's willingness to pay for differentiated products depends on household income. However, 0-1 preferences are very stylized, as households can adjust their consumption in response to price and income changes only through the extensive margin.[3] We then show that

[1]This heterogeneous-firm framework has proven to be useful in explaining firm-level evidence on export behavior. For a recent survey, see Bernard, Jensen, Redding, and Schott (2012).

[2]Baldwin and Harrigan (2011) find that real GDP per worker in the destination country is a significant determinant of zeros in US export data. While they consider real GDP per worker as a demand-related control, they do not systematically explore this result in the context of their theoretical model (which sticks to the assumption of homothetic preferences).

[3]Notice that "0-1" preferences and CES preferences can be considered as two polar cases. With 0-1 preferences, optimal consumption responds only along the extensive margin; with CES-preferences consumption responds only

the qualitative results of the 0-1 model carry over to more general settings where consumption is allowed to respond both along the extensive and the intensive margin.

Our paper makes two key contributions. The *first* is the recognition that firms from rich countries might not export to a poor country due to a threat of international arbitrage. Consider a US firm that sells its product both in the US and in China. Suppose this firm charges a price in China equal to the Chinese households' (low) willingness to pay and a price in the US equal to the US households' (high) willingness to pay. When price differences are large, arbitrage opportunities emerge: arbitrageurs can purchase the good cheaply on the Chinese market, ship it back to the US, and underbid local US producers. In equilibrium, US firms anticipate the threat of arbitrage and will adjust accordingly. To avoid arbitrage, a US exporter has basically two options: (i) charge a price in the US sufficiently low to eliminate arbitrage incentives; or (ii) abstain from selling the product in China (and other equally poor countries) thus eliminating arbitrage opportunities. These two options involve a trade-off between market size and prices: firms that export globally have a large market but need to charge a low price; firms that sell exclusively on the US market (and in other equally rich countries) can charge a high price but have a small market. In an equilibrium with ex-ante identical firms, the two options yield the same profit.

The *second* key contribution of our paper relates to gains from trade and the welfare effects of trade liberalizations. When per capita income gaps are small, firms are not constrained by arbitrage, and all goods are traded. In such a "full trade equilibrium", lower trade costs increase welfare in both countries. Lower losses during transport provide resources for production of more varieties from which consumers in both countries benefit. In the more interesting case of large per capita income gaps, firms are constrained by arbitrage, and not all goods are exported to the poor country. In such an "arbitrage equilibrium" lower trade costs increase welfare in the rich country but *decrease* welfare in the poor country. The reason is that lower trade costs tighten the arbitrage constraint. With lower trade costs, globally active firms in the rich country need to reduce prices on their home market. This will induce more firms to abstain from exporting to poor countries, thus avoiding international arbitrage. As a result, fewer varieties are exported to poor countries leading to lower consumption and welfare in these countries.

Our analysis highlights three further points. *First*, we make precise the differential consequences of an increase in aggregate GDP due to a higher per capita income and due to a larger population. A higher *per capita income* in the South raises poor households' willingness to pay, increasing northern firms' incentive to sell their products internationally. In equilibrium, a larger fraction of northern firms export their product to the South. In contrast, a larger *population* in the South leaves southern households' demand for varieties unchanged but allows for the production of more varieties. This increases the world's per capita consumption due to a scale effect; increases the volume of trade; and may or may not increase trade intensity. Moreover, a larger population in the poor country may or may not increase the probability that a northern firm exports to the South. In sum, our model predicts that per capita income has a stronger effect

along the intensive margin (because Inada conditions induce households to consume all goods, irrespective of prices and income). Clearly, the realistic scenario is in between these polar cases. We look at this case in Section 5.

than population size on the probability that a northern firm exports to the South.

A *second* point shows that the result of detrimental effects of trade liberalizations (on a poor country's welfare) needs to be qualified in a multi-country setting. When there are many rich and many poor countries, a multilateral trade liberalization still reduces North-South trade due to tighter arbitrage. However, it also stimulates South-South trade because the arbitrage constraint is not binding among trading partners with similar per capita incomes. Hence a multilateral trade liberalization increases the welfare of poor households if the increase in South-South trade overcompensates the fall in North-South trade. The multi-country setting is also useful because it delivers empirical predictions. The main prediction (on which we shed light empirically) is that a northern firm has a high probability to export to other northern countries, while the probability that it exports to a southern country is significantly lower and decreases in the per capita income gap between the North and South.

A *third* point analyzes the conditions under which the basic logic of our 0-1 preferences carries over to general (additive) preferences that allow for both an extensive and intensive margin of consumption. We assume a general, additive subutility function $v(c)$ and make precise the conditions on $v(c)$ under which international arbitrage can emerge. If these conditions are met, there will be export zeros, provided that per capita income differences between the trading partners are sufficiently large. In this sense, the predictions of the simple 0-1 model hold also under more general preferences.

There is compelling evidence that threats of arbitrage affect the pricing decisions of firms in many markets. Pharmaceutical industries are most prominent examples (WHO 2001, Ganslandt and Maskus 2004, Goldberg 2010). The WHO (2001) report argues that restraints on parallel trade between poor and rich countries would allow companies to supply the former. Consequently, a key WHO recommendation is a more comprehensive implementation of differential pricing strategies. Parallel trade is also relevant in other industries such as cars (Lutz 2004, Yeung and Mok 2013), consumer electronics (Feng 2013), DVDs and cinemas (Burgess and Evans 2005), and other markets, like clothing and cosmetics (NERA 1999).

To shed light on the quantitative relevance of the arbitrage channel, we proceed in two steps. We first provide reduced-form evidence from disaggregated trade data. We find that there is a strong and quantitatively large association between the export probability (of the US and other large countries) and the per capita income of a given destination (conditional on the destination's aggregate GDP and other commonly used determinants of international trade flows). This reduced-form evidence is consistent with binding arbitrage constraints, but could also be explained by a cost constraint: firms abstain from exporting because the marginal cost of exporting are too high.[4]

In a second step, we pursue a model-based approach to learn more about the potential relevance of arbitrage constraints. We start out with the framework of Simonovska (2015) who characterizes the general equilibrium in a world economy with many countries and heterogenous

[4] In the absence of fixed export costs, this requires a specification of preferences featuring a finite reservation price which increases in the destination's per capita income. When the destination is very poor the reservation price is very low, and export zeros arise because the price does not cover the marginal costs associated with exporting to that destination. The specification in Simonovska (2015) satisfies these features.

firms. She uses Stone-Geary preferences, which belong to the class of preferences where arbitrage constraints may, in principle, become binding. In this quite general environment, Simonovska (2015) is able to come up with predictions for the relative price of a given product between any two locations – under the assumption that firms do *not* have to take arbitrage constraints into account when setting their prices. This suggests a simple consistency check to assess the quantitative importance of arbitrage: With (i) predictions on relative (unconstrained) prices and (ii) estimates of trade costs between any two locations, we can assess whether arbitrage constraints are violated in an equilibrium that does not take into account arbitrage. In this numerical exercise, we find that arbitrage constraints are indeed violated in many trade relations.

Our analysis contributes to the literature on parallel trade (surveyed in Maskus, 2000 and Ganslandt and Maskus, 2007). In partial equilibrium models, the welfare effects of parallel imports in a rich country are typically ambiguous because there is a tradeoff between reduced innovation incentives on the one side and lower prices for consumers on the other side. To see the contrast, consider for example the recent contribution by Roy and Saggi (2012). They show that in an international duopoly, parallel trade induces the southern firm to charge an above monopoly price in the South in order to be able to charge a high price in the North. Softer competition in the North then induces the northern firm to sell only in its home market at a high price, which harms northern consumers. We show that considering the general equilibrium uncovers an opposing force working through the economy wide resource constraint: It is still true that a subset of northern firms will find it optimal to sell only in their home market at a high price. But this means that less northern resources are used to produce goods for the South, which increases the numbers of available varieties in the North and thus welfare. So the welfare effects of parallel trade rules go in the opposite direction when considering the general equilibrium.

Our paper is also related to the pricing-to-market literature, which focuses on the cross-country dispersion of prices of tradable goods. Atkeson and Burstein (2008) generate pricing-to-market in a model with Cournot competition and variable mark-ups. However, their focus is on the interaction of market structure and changes in marginal costs rather than on per capita income effects. Hsieh and Klenow (2007), Manova and Zhang (2012), and Alessandria and Kaboski (2011), among others, document that prices of tradable consumer goods show a strong positive correlation with per capita incomes in cross-country data. The papers by Markusen (2013), Sauré (2010), Behrens and Murata (2012a,b) and Bekkers, Francois, and Manchin (2012) provide frameworks in which richer consumers are less price-sensitive, so mark-ups and prices are higher in richer countries. Mrazova and Neary (2013) explore in a comprehensive way how deviations from CES preferences affect equilibrium outcomes in the Krugman model.[5] Variable

[5]Other related papers allowing for non-homothetic (or quasi-homothetic) preferences include Fajgelbaum, Grossman, and Helpman (2011), Hummels and Lugovskyy (2009), Desdoigts and Jaramillo (2009), Neary (2009), Melitz and Ottaviano (2008), Falkinger (1990), and Auer, Chaney and Sauré (2014). Many papers found empirical support for non-homotheticities, e.g. Hunter and Markusen (1988), Hunter (1991), Francois and Kaplan (1996), Choi, Hummels, and Xiang (2006), Dalgin, Mitra, and Trindade (2008), Fieler (2011), Hepenstrick and Tarasov (2015), Bernasconi (2013). Caron, Fally and Markusen (2014) show that non-homothetic preferences are quantitatively important for explaining the observed correlation between income elasticities and skill intensities at the sectoral level, resolving a substantial part of "missing trade puzzle" between rich and poor countries.

mark-ups and pricing-to-market driven by per capita income are also a crucial feature in our framework. Our paper extends this literature by showing that export zeros arise from the (threat of) international arbitrage, a feature not considered in previous papers.

The presence of a trade participation margin links the present paper to a recent literature that builds on Melitz (2003) and explores demand- and/or market-size effects in the context of heterogeneous firm models. Arkolakis (2010) incorporates marketing costs into that framework, generating an effect of population size on export markets in addition to aggregate income. Eaton, Kortum, and Kramarz (2011) extend this framework, allowing for demand shocks (in addition to cost shocks) as further determinants of firms' export behavior. Eaton, Kortum and Sotelo (2013) show that a standard heterogeneous-firm trade model with an integer number (rather than a continuum) of firms can reconcile the large share of a small number of firms in global trade, the many zeros in bilateral trade, and the observed trade volumes in cross-country data. These papers stick to homothetic preferences, hence arbitrage cannot arise. This is different from our paper where non-homotheticities and arbitrage incentives play a central role and may per se generate a trade participation margin.

The remainder of the paper is organized as follows. In the next section, we present the basic assumptions and discuss the autarky equilibrium. In Section 3, we use our basic framework to study trade patterns and trade gains in a two-country setting. Section 4 extends the analysis to many rich and poor countries. In Section 5, we introduce general preferences and show that arbitrage equilibria also arise when consumption responds both along the extensive and the intensive margin. Section 6 presents reduced form evidence on the association between US export probabilities and a destination's per capita income. In section 7 we show that a standard calibrated trade model (that ignores arbitrage) generates predictions on relative prices that violate no-arbitrage constraints in many bilateral trade relations. Section 8 concludes.

2 Autarky

We start by presenting the autarky equilibrium. The economy is populated by $\mathcal{P}$ identical households. Each household is endowed with L units of labor, the only production factor. Labor is perfectly mobile within countries and immobile across countries. The labor market is competitive and the wage is W. Production requires a fixed labor input F to set up a new firm and a variable labor input $1/a$ to produce one unit of output, the same for all firms. Producing good j in quantity $q(j)$ thus requires a total labor input of $F + q(j)/a$.

Consumers. Households spend their income on a continuum of differentiated goods. We assume that goods are indivisible and a given product j yields positive utility only for the first unit and zero utility for any additional units.[6] Thus consumption is a binary choice: either you buy or you don't buy. Let $x(j)$ denote an indicator that takes value 1 if good j is purchased and

[6]Preferences of this type were used, inter alia, by Murphy, Shleifer and Vishny (1989) to study demand composition and technology choices, by Matsuyama (2000) to explore non-homotheticities in Ricardian trade, and by Falkinger (1994) and Foellmi and Zweimüller (2006) to analyze inequality and growth.

value 0 if not. Then utility takes the simple form

$$U = \int_0^\infty x(j)dj, \qquad \text{where } x(j) \in \{0, 1\}\,. \tag{1}$$

Notice that utility is additively separable and that the various goods enter symmetrically. Hence the household's utility is given by the number of consumed goods.

Consider a household with income y who chooses among (a measure of) N goods supplied at prices $\{p(j)\}$.[7] The problem is to choose $\{x(j)\}$ to maximize the objective function (1) subject to the budget constraint $\int_0^N p(j)x(j)dj = y$. Denoting λ as the household's marginal utility of income, the first order condition can be written as

$$\begin{aligned} x(j) &= 1 \text{ if } 1 \geq \lambda p(j) \\ x(j) &= 0 \text{ if } 1 < \lambda p(j)\,. \end{aligned}$$

Rewriting this condition as $1/\lambda \geq p(j)$ yields the simple rule that the household will purchase good j if its willingness to pay $1/\lambda$ does not fall short of the price $p(j)$.[8] The resulting demand curve, depicted in Figure 1, is a step function which coincides with the vertical axis for $p(j) > 1/\lambda$ and equals unity for prices $p(j) \leq 1/\lambda$.

Figure 1

By symmetry, the household's willingness to pay is the same for all goods and equal to the inverse of λ, which itself is determined by the household's income and product prices. Intuitively, the demand curve shifts up when the income of the consumer increases (λ falls) and shifts down when the price level of all other goods increases (λ rises).

It is interesting to note the difference between consumption choices under these "0-1" preferences and the standard CES-case. With 0-1 preferences, the household chooses how many goods to buy, while there is no choice about the consumed quantity.[9] In contrast, a household has a choice with CES preferences about the quantities of the supplied goods, but finds it optimal to consume all varieties in positive amounts. This is because Inada conditions imply an infinite reservation price. In other words, 0-1 preferences shift the focus to the *extensive* margin of consumption, while CES preferences focus entirely on the *intensive* margin. It is important to note, however, that our central results below do not depend on the 0-1 assumption. In fact, we will show below that more general preferences – which allow for both the extensive and the intensive

[7] Notice that the integral in (1) runs from zero to infinity. While preferences are defined over an infinitely large measure of potential goods, the number of goods actually supplied is limited by firm entry, i.e. only a subset of potentially producible goods can be purchased at a finite price.

[8] Strictly speaking, the condition $1 \geq \lambda p(j)$ is necessary but not sufficient for $c(j) = 1$ and the condition $1 < \lambda p(j)$ is sufficient but not necessary for $c(j) = 0$. This is because purchasing all goods for which $1 = \lambda p(j)$ may not be feasible given the consumer's budget. For when N different goods are supplied at the same price p but $y < pN$ the consumer randomly selects which particular good will be purchased or not purchased. This case, however, never emerges in the general equilibrium.

[9] The discussion here rules out the case where incomes could be larger than pN, meaning that the consumer is subject to rationing (i.e. he would want to purchase more goods than are actually available at the available prices). While this could be a problem in principle, it will never occur in equilibrium.

margin of consumption – generate results that are qualitatively similar to those derived in the 0-1 case.

Equilibrium. Since both firms and households are identical, the equilibrium is symmetric. Similar to the standard monopolistic competition model, the information on other firms' prices is summarized in the shadow price λ. Hence, the pricing decision of a monopolistic firm depends only on λ. Moreover, the value of λ is unaffected by the firm's own price because a single firm is of measure zero.

Lemma 1 *There is a single price $p = 1/\lambda$ in all markets and all goods are purchased by all consumers.*

Proof. Aggregate demand for good j is a function of λ only. Consequently, the pricing decision of a monopolistic firm depends on the value of λ and not directly on the prices set by competitors in other markets. Thus, it is profit maximising to set $p(j) = 1/\lambda$ as long as $1/\lambda$ exceeds marginal costs. To prove the second part of the Lemma, assume to the contrary that only a fraction ν of consumers purchases the product at price $p(j) = p = 1/\lambda$. However, this cannot be an equilibrium, as the firm could undercut the price slightly and sell to all consumers. ■

Each monopolistic firm faces a demand curve as depicted in Figure 1. It will charge a price equal to the representative consumer's willingness to pay $p = 1/\lambda$ and sell output of quantity 1 to each of the $\mathcal{P}$ households. Without loss of generality, we choose labor as the numéraire, $W = 1$. Two conditions characterize the autarky equilibrium. The *first* is the zero-profit condition, ensuring that operating profits cover the entry costs but do not exceed them to deter further entry. Entry costs are $FW = F$ and operating profits are $[p - W/a]\,\mathcal{P} = [p - 1/a]\,\mathcal{P}$. The zero-profit condition can be written as $p = (aF + \mathcal{P})\,/a\mathcal{P}$. This implies a mark-up μ – a ratio of price over marginal cost – equal to

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

Notice that technology parameters a and F and the market size parameter $\mathcal{P}$ determine the mark-up.[10] We will show below that the mark-up is a crucial channel through which non-homothetic preferences affect patterns of trade and the international division of labor.

The *second* equilibrium condition is a resource constraint ensuring that there is full employment $\mathcal{P}L = FN + \mathcal{P}N/a$. From this latter equation, equilibrium product diversity (both in production and consumption) in the decentralized equilibrium is given by

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

[10] Notice that the determination of mark-ups is quite different between the 0-1 outcome and the standard CES-case. With 0-1 preferences, the mark-up depends on technology and market size parameters. With CES preferences, the mark-up is determined by the elasticity of substitution between differentiated goods, while it is independent of technology and market size. Notice further that, from the zero profit condition of the CES-model, we have $\omega F = (p - \omega/b)x\mathcal{P}$ (where x is the – endogenously determined – quantity of the representative product and $1/b$ is the unit labor requirement). Thus we can write the mark-up as $(b/x)(F/\mathcal{P}) + 1$, which compares to $a(F/\mathcal{P}) + 1$ in the 0-1 case. To achieve realistic mark-ups in empirical applications, the parameter a needs to be normalized appropriately, i.e. it has to assume an order of magnitude similar to the ratio b/x in the CES-model.

3 Trade between a rich and a poor country

Let us now consider a world economy where a rich and a poor country trade with each other. We denote variables of the rich country with superscript R and those of the poor country with superscript P. To highlight the relative importance of differences in per capita incomes and population sizes, we let the two countries differ along both dimensions, hence $L^R > L^P$ and $\mathcal{P}^R \gtreqless \mathcal{P}^P$. We assume trade is costly and of the standard iceberg type: for each unit sold to a particular destination, $\tau > 1$ units have to be shipped and $\tau - 1$ units are lost during transport.

3.1 Full trade equilibrium

When the income gap between the two countries is small, all goods are traded internationally. In such a *full trade equilibrium*, a firm's optimal price for a differentiated product in country $i = R, P$ equals the households' willingnesses to pay (see Figure 1), hence we have $p^R = 1/\lambda^R$ and $p^P = 1/\lambda^P$. Since country R is wealthier than country P, we have $\lambda^R < \lambda^P$ and $p^R > p^P$. By symmetry, the prices of imported and home-produced goods are identical within each country.

Solving for the full trade equilibrium is straightforward. Consider the resource constraint in the rich country. $N^R F$ labor units are needed for setting up the N^R firms. Moreover, $N^R \mathcal{P}^R/a$ and $N^R \mathcal{P}^P \tau/a$ labor units are employed in production to serve the home and the foreign market, respectively. Since each of the $\mathcal{P}^R$ households supplies L^R units of labor inelastically, the resource constraint is $\mathcal{P}^R L^R = N^R F + N^R \left(\mathcal{P}^R + \tau \mathcal{P}^P\right)/a$. Similarly, for the poor country. Solving for N^i $(i = R, P)$ lets us determine the number of active firms in the two countries

$$N^i = \frac{a\mathcal{P}^i}{aF + (\mathcal{P}^i + \tau\mathcal{P}^{-i})} L^i, \tag{2}$$

(where $-i = P$ if $i = R$ and vice versa).

Now consider the zero-profit conditions in the two countries. An internationally active firm from country i generates total revenues equal to $p^R\mathcal{P}^R + p^P\mathcal{P}^P$ and has total costs $W^i\left[F + (\mathcal{P}^i + \tau\mathcal{P}^{-i})/a\right]$. Using the zero-profit conditions of the two countries lets us calculate relative wages

$$\omega \equiv \frac{W^P}{W^R} = \frac{aF + \tau\mathcal{P}^P + \mathcal{P}^R}{aF + \mathcal{P}^P + \tau\mathcal{P}^R}. \tag{3}$$

When the two countries differ in population size, wages (per efficiency unit of labor) are higher in the larger country.[11] Why are wages higher in larger countries? The reason is that labor is more productive in a larger country. To see this, consider the amount of labor needed by a firm in country i to serve the world market. When country R is larger than country P, firms in country R need less labor to serve the world market because there are less iceberg losses during transportation, which is reflected in relative wages. There are two cases in which wages are equalized: (i) $\tau = 1$. When there are no trade costs, the productivity effect of country size

[11] While ω measures relative wages per efficiency unit of labor, $\omega L^P/L^R$ measures relative nominal per capita incomes. In principle, $\omega L^P/L^R > 1$ is possible, so that country P (with the *lower* labor endowment) has the *higher* per capita income. We show below that this can happen only in a full trade equilibrium but not in an arbitrage equilibrium. The latter case is the interesting one in the present context.

vanishes. (ii) $\mathcal{P}^P = \mathcal{P}^R$. When the two countries are of equal size, productivity differences vanish because iceberg losses become equally large. Note further that $\tau^{-1} < \omega < \tau$. When the poor country becomes very large, iceberg losses as a percentage of total costs become negligible, $\omega \to \tau$. Similarly, when the rich country becomes large, $\omega \to \tau^{-1}$.

Finally, let us calculate prices and mark-ups in the respective export destination. The budget constraint of a household in country i is $W^i L^i = p^i \left(N^R + N^P\right)$. Combining the zero-profit condition with these budget restrictions and the above equation for the number of firms, lets us express the price in country i as

$$p^i = W^i L^i \frac{aF + \mathcal{P}^R + \tau\mathcal{P}^P}{a\mathcal{P}^R L^R + a\omega\mathcal{P}^P L^P}, \; i = R, P. \tag{4}$$

By symmetry, prices for the various goods are identical within each country, irrespective of whether they are produced at home or abroad. Consequently, imported goods generate a lower mark-up than locally produced goods because exporters cannot pass trade costs through to consumers.[12] Marginal costs are W^i/a when the product is sold in the home market and $\tau W^i/a$ when the product is sold in the foreign market. Hence mark-ups (prices over marginal costs) are $\mu_D^i = p^i a/W^i$ in the domestic market and $\mu_X^i = p^j a/(W^i \tau)$ in the export market. Hence a full trade equilibrium is characterized as follows: (i) $N^P/N^R = \omega\mathcal{P}^P L^P / \left(\mathcal{P}^R L^R\right)$, i.e. differences in aggregate GDP lead to proportional differences in produced varieties; (ii) $p^P/p^R = \omega L^P/L^R$, i.e. differences in per capita incomes generate proportional differences in prices; and (iii) $\mu_D^P/\mu_D^R = \mu_X^P/\mu_X^R = L^P/L^R < 1$, i.e. differences in per capita endowments lead to proportional differences in mark-ups.

Patterns of international trade. Let us highlight how the volume and structure of international trade depend on relative per capita endowments L^P/L^R. We define "trade intensity" ϕ as the ratio between the value of world trade and world GDP. In a full trade equilibrium the value of world trade is given by $p^R N^P \mathcal{P}^R + p^P N^R \mathcal{P}^P$ while world income is $L^R\mathcal{P}^R + \omega L^P \mathcal{P}^R$. Trade intensity is given by

$$\phi = \frac{2L^R\mathcal{P}^R \cdot \omega L^P \mathcal{P}^P}{\left(L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P\right)^2}$$

When all goods are traded, the relative size of aggregate GDP matters for trade intensity. When GDP differs strongly across the two countries, trade intensity is small as most world production takes place in the large country and most of this production is also consumed in this country. Trade intensity is maximized when the two countries are of exactly equal size. We can now state the following proposition

Proposition 1 *Assume the two countries are in a full trade equilibrium. a) All goods are traded. b) Trade intensity ϕ increases with both the per capita endowment L^P and population size $\mathcal{P}^P$ if $\omega L^P\mathcal{P}^P < L^R\mathcal{P}^R$. c) The impact on ϕ of $\mathcal{P}^P$ is stronger than the one of L^P. d) A trade liberalization increases trade intensity if $\omega L^P\mathcal{P}^P < L^R\mathcal{P}^R$.*

[12] This is different from CES preferences, where transportation costs are more than passed through to prices as exporters charge a fixed mark-up on marginal costs (including transportation). Notice that limited cost pass-through has been documented in a large body of empirical evidence.

Proof. See Appendix A. ■

Trade and welfare. Let us finally consider the gains from trade and the welfare effects of a trade liberalization in a full trade equilibrium. Since all firms sell to all households worldwide, consumption and welfare levels are equalized across rich and poor countries. Gains from trade are higher for the country with lower product variety under autarky. Product variety in autarky is $N^i = a\mathcal{P}^i L^i / \left(aF + \mathcal{P}^i\right)$. The country with a smaller population $\mathcal{P}^i$ and/or lower per capita endowment (lower L^i) gains more from trade. Here we are interested in how bilateral trade liberalizations affect welfare and the distribution of trade gains between the two countries. A trade liberalization is modeled as a reduction in iceberg transportation costs τ.

In a full trade equilibrium, households in both countries purchase all goods produced worldwide. Hence the welfare levels are identical in both countries despite their unequal endowment with productive resources

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

Firms' price setting behavior drives this result. R-consumers are willing to pay higher prices than P-consumers because their income is higher. In the full trade equilibrium, higher nominal incomes translate one-to-one into higher prices, welfare is therefore identical. To see the mechanism by which welfare is equalized, consider mark-ups in the special case when the two countries are equally large. When $\mathcal{P}^P = \mathcal{P}^R$, prices are higher in country R, while costs are the same for each country. In other words, country-R households bear a larger share of total costs. In this case, the poor country's welfare is lower under autarky.[13] We summarize this in the following proposition.

Proposition 2 *In a full trade equilibrium, welfare levels are equalized. A trade liberalization (a lower τ) increases welfare for both countries.*

Proof. In text. ■

3.2 "Arbitrage" equilibrium with non-traded goods

Full trade ceases to be an equilibrium when per capita income differences $\omega L^P/L^R$ become large. The reason is a threat of arbitrage. Consider a US firm that sells its product both in the US and in China. Suppose the firm charges a price in China that equals the Chinese households' willingness to pay $p^P = 1/\lambda^P$ and a price in the US that equals the US households' willingness to pay $p^R = 1/\lambda^R$. If the difference between $1/\lambda^P$ and $1/\lambda^R$ is large, arbitrage opportunities emerge. Arbitrageurs can purchase the good cheaply on the Chinese market, ship it back to the US, and underbid the producer on the US market. A threat of arbitrage also concerns Chinese firms which both produce for the local market and export to the US. When these firm charge

[13] This continues to hold as long as $\mathcal{P}^P$ is not too much larger larger than $\mathcal{P}^R$. When $\mathcal{P}^P \gg \mathcal{P}^R$, so that $\omega L^P > L^R$, prices become higher in country P. In that case, country-P bears the larger share in total costs.

too high prices in the US, arbitrage traders purchase the cheap products in China and parallel export them to the US.

Firms anticipate this arbitrage opportunity and adjust their pricing behavior accordingly. Notice that the threat of parallel trade only contains firms operating on the world market. Firms that abstain from selling the product in the poor country and focus exclusively on the market of the rich country do not face such a threat. Adopting this latter strategy implies a smaller market but lets firms exploit the rich households' high willingness to pay. In equilibrium, firms are indifferent between the two strategies. Notice that concentrating sales exclusively on the rich market country is, in principle, an option both for producers in the rich and in the poor country. In equilibrium, however, only by rich-country producers adopt this strategy. While total revenues are independent of the producer's location, total costs are not. To serve households in the rich country, country-R producers face marginal costs $W^R a$, while country-P exporters face marginal costs $W^R \omega\tau/a$ (they have to bear transportation costs). Since $\omega\tau > 1$, country-P producers have a competitive disadvantage in serving the rich country even when the poor country has lower wages $\omega < 1$.

An *arbitrage equilibrium* looks as follows. A subset of rich-country producers sells their product exclusively in the rich country, while the remaining rich-country producers sell their product both in the rich and in the poor country. All poor-country producers sell their product worldwide. To see why this is an equilibrium, consider the alternative situation in which all rich-country producers trade their products internationally. If all firms charged a price that prevents arbitrage, all goods would be priced below rich households' willingness to pay. In that case, however, rich households do not spend all their income, generating an infinitely large willingness to pay for additional products. This would induce country-R firms to sell their product only on the home market. Thus, in equilibrium, both types of firms will exist. Notice that all firms are ex-ante identical (i.e. all firms have the same cost- and demand functions). Notice that there is an indeterminacy concerning the selection of firms into export status. Clearly, this is an artefact of the symmetry-assumption, which would disappear once asymmetries (i.e. firm heterogeneities) are added to the model.

We are now ready to solve for the arbitrage equilibrium. Denote the price in the rich country of traded and non-traded goods by p_T^R and p_N^R, respectively. The price of non-traded goods is $p_N^R = 1/\lambda^R$. Anticipating the threat of parallel trade, the price of traded goods may not exceed and exactly equals the price in the poor country (plus trade costs), $p_T^R = \tau/\lambda^P$, in equilibrium. The price of a product in the poor country is still given by $p^P = 1/\lambda^P$. The following lemma proves that this is a Nash equilibrium.

Lemma 2 *In an arbitrage equilibrium, firms that sell their product in both countries (i) set $p^P = 1/\lambda^P$ in country P and $p_T^R = \tau p^P$ in country R, and (ii) sell to all households in both countries.*

Proof. (i) Assume $1/\lambda^P$ exceeds marginal costs of exporting. In that case, the profit maximization problem of an exporting firm reduces to maximize total revenue $\mathcal{P}^P p^P(j)+\mathcal{P}^R p^R(j)$ s.t. $\tau p^P(j) \geq p^R(j)$ and $p^i(j) \leq 1/\lambda^i$. Applying Lemma 1, it is profit maximizing to set

$p^i(j) = 1/\lambda^i$ if $\tau/\lambda^P \geq \lambda^R$ (full trade equilibrium). If $\tau/\lambda^P < \lambda^R$, the arbitrage constraint is binding $\tau p^P(j) = p^R(j) = p_T^R$ and revenues are maximized when $p^P(j) = 1/\lambda^P$. (ii) Assume to the contrary that only a fraction ν of consumers purchases the product at price $p^P(j) = 1/\lambda^P$. As in Lemma 1, this cannot be an equilibrium, as the firm would lower $p^P(j)$ and $p^R(j)$ slightly and gain the whole market in the poor country. ■

The zero-profit condition for an internationally active country-i producer is $p_T^R \mathcal{P}^R + p^P \mathcal{P}^P = W^i \left[F + (\mathcal{P}^i + \tau \mathcal{P}^{-i})/a\right]$. These firms' total revenues do not depend on the location of production, but the required labor input depends on location. Differences in population sizes generate differences in (total) transport costs, and relative wages equalize these differences. From the zero-profit conditions we see that relative wages ω are still given by equation (3). The zero-profit conditions also let us derive the prices for the various products. Using $p_T^R = \tau p^P$, we get

$$p_T^R = \frac{\tau}{a} \frac{aF + \mathcal{P}^R + \tau \mathcal{P}^P}{\tau \mathcal{P}^R + \mathcal{P}^P} \quad \text{and} \quad p^P = \frac{1}{a} \frac{aF + \mathcal{P}^R + \tau \mathcal{P}^P}{\tau \mathcal{P}^R + \mathcal{P}^P},$$

where we have set $W^R = 1$. (We use this normalization throughout the paper.) The zero-profit condition for an exclusive rich-country producer is $p_N^R \mathcal{P}^R = F + \mathcal{P}^R/a$, from which we calculate the equilibrium price of a non-traded variety

$$p_N^R = \frac{aF + \mathcal{P}^R}{a\mathcal{P}^R}.$$

Notice that, due to the arbitrage constraint on exporters' pricing behavior, prices do not depend on L^P and L^R. This is quite different from the full-trade equilibrium, where price differences reflect differences in per capita endowments.

The resource constraint in country P is the same as that in the full trade equilibrium, so N^P is still given by (2). The resource constraint in country R is now different, however, because there are traded and non-traded products. Denoting the range of traded and non-traded goods produced in the rich country by N_T^R and N_N^R, respectively, the resource constraint of country R is given by $\mathcal{P}^R L^R = N_T^R \left(F + (\mathcal{P}^R + \tau \mathcal{P}^P)/a\right) + N_N^R \left(F + \mathcal{P}^R/a\right)$. Together with the trade balance condition $N_T^R p^P \mathcal{P}^P = N^P p_T^R \mathcal{P}^R$ and the terms of trade $p_T^R/p^P = \tau$ we get

$$N_T^R = \frac{a\mathcal{P}^R}{aF + \tau \mathcal{P}^R + \mathcal{P}^P} \tau L^P, \quad \text{and } N_N^R = \frac{a\mathcal{P}^R}{aF + \mathcal{P}^R} \left(L^R - \tau \omega L^P\right). \tag{5}$$

Patterns of international trade. Let us now describe volume and structure of international trade in an arbitrage equilibrium. The value of traded goods is $p_T^R N^P \mathcal{P}^R + p^P N_T^R \mathcal{P}^P$ (while world income still is $L^R \mathcal{P}^R + \omega L^P \mathcal{P}^R$). Using equations (2) and (5) we calculate the trade intensity in an arbitrage equilibrium.

$$\phi = \frac{2\tau}{\tau + (\mathcal{P}^P/\mathcal{P}^R)} \cdot \frac{\omega L^P \mathcal{P}^P}{L^R \mathcal{P}^R + \omega L^P \mathcal{P}^P} \tag{6}$$

Equation (6) shows that per capita incomes differences and differences in population sizes affect trade intensity in different ways. Consider first the impact of a given change in per capita income of country P. The above expression for ϕ reveals that a higher per capita income of the poor country unambiguously increases the intensity of trade. This is reminiscent of the Linder-hypothesis (Linder 1961) postulating that a higher similarity in per capita incomes is associated higher trade between trading partners. The intuition for this result is straightforward. When L^P increases by 10 percent, the range of exported goods increases by 10 percent while prices remain unchanged. Hence the aggregate value of trade $p_T^R N^P \mathcal{P}^R + p^P N_T^R \mathcal{P}^P$ increases by 10 percent as well. In contrast, increasing L^P by 10 percent (while leaving L^R unchanged) increases world GDP by less than 10 percent. Trade intensity, the ratio between world trade and world GDP, thus rises unambiguously.

Now consider a change in population-size of country P. It turns out that a change in $\mathcal{P}^P$ has a smaller effect on trade intensity than an increase in relative per capita incomes that increases GDP by the same magnitude, i.e. we have $\partial \text{log}\phi / \partial \log \mathcal{P}^P < \partial \log \phi / \partial \log L^P$. This can be seen from looking at the volume of world trade which is equal to $2p^P N_T^R \mathcal{P}^P$. An increase in $\mathcal{P}^P$ has a direct and an indirect effect on world trade. The direct effect increases trade in proportion to country P's population. The indirect effect lowers per capita imports. Notice that imports per capita in country-P are equal to $p^P N_T^R = \left[\tau / (\tau + \mathcal{P}^P / \mathcal{P}^R)\right] \omega L^P$. From the point of view of country R, a larger population in country P requires fewer exports to each country-P households to cover a given amount of own imports. Hence country-P imports (and world trade) increase with $\mathcal{P}^P$ less than proportionately.[14]

Proposition 3 *Assume per capita income differences are large, so that the world economy is in an arbitrage equilibrium. a) Some firms in country R do not export. b) An increase in per capita endowment L^P raises trade intensity ϕ, while an increase in population size $\mathcal{P}^P$ may increase or decrease ϕ. c) The impact on ϕ of $\mathcal{P}^P$ is weaker than the one of L^P. d) A trade liberalization decreases trade intensity.*

Proof. See Appendix B. ■

Trade and welfare. We proceed by looking at the impact of trade liberalizations. We first consider the case where trade costs are symmetric and explore a bilateral liberalization. (We look at the unilateral case below). In an arbitrage equilibrium, a bilateral trade liberalization lets consumers' welfare levels in the two countries diverge. Country-P households' welfare equals $N_T^R + N^P$, while country-R households' welfare equals $N^P + N_T^R + N_N^R$. Using (2) and (5), these welfare levels are given by

$$U^P = \frac{aL^P(\mathcal{P}^P + \tau\mathcal{P}^R)}{aF + \tau\mathcal{P}^R + \mathcal{P}^P} \quad \text{and} \quad U^R = \frac{aL^P(\mathcal{P}^P + \tau\mathcal{P}^R)}{aF + \tau\mathcal{P}^R + \mathcal{P}^P} + \frac{a\mathcal{P}^R\left(L^R - \tau L^P\right)}{aF + \mathcal{P}^R}.$$

It is straightforward to verify, that $\partial U^P / \partial \tau > 0$ while $\partial U^R / \partial \tau < 0$. We are now able to state the following proposition.

[14] Notice that an increase in $\mathcal{P}^P$ also increases ω. It is shown in the proof of proposition 2 (see Appendix) that taking the impact of $\mathcal{P}^P$ on ω into account, an increase in $\mathcal{P}^P$ still reduces per capita imports.

Proposition 4 *In an arbitrage equilibrium, a trade liberalization increases the welfare of country-R households, but decreases it for country-P households.*

Proof. In text. ■

Proposition 4 shows the crucial role of trade costs for welfare. Unequal countries have different preferred trade barriers (or different preferred degrees of trade liberalizations). Consumers in the rich country are essentially free-traders, whereas consumers in the poor country are harmed by liberalizations. What is the intuition behind this result? The reason is country-R firms' pricing behavior. As higher trade costs imply a less tight arbitrage constraint, country-R firms can charge higher prices for traded goods relative to non-traded goods. This induces country-R firms to export rather than sell exclusively to domestic customers. The result is an increase in trade intensity which benefits the poor country. Put differently, poor country households are against a trade liberalization because a lower τ decreases trade and welfare in country P.

Unilateral trade liberalization. Up to now we have assumed symmetric trade costs across countries. However, policy makers can influence trade costs through tariffs and regulations. This is interesting in the present context because, in an arbitrage equilibrium, the poor country has an incentive to increase trade barriers and relax the arbitrage constraint. This increases the supply of northern varieties and hence may raise welfare in the South. It is therefore interesting to look at an unilateral trade liberalization. Assume that trade costs differ between countries, with τ^i denoting iceberg costs for imports into country i. While total revenues of exporters are still $p_T^R \mathcal{P}^R + p^P \mathcal{P}^P$, now total costs do not only vary as a result of unequally large populations but also because of differences in transportation costs, $W^i \left[F + (\mathcal{P}^i + \tau^{-i}\mathcal{P}^{-i})/a\right]$. From the zero-profit condition we derive relative wages ω as

$$\omega \equiv \frac{W^P}{W^R} = \frac{aF + \tau^P \mathcal{P}^P + \mathcal{P}^R}{aF + \mathcal{P}^P + \tau^R \mathcal{P}^R},$$

which implies that $\left(\tau^R\right)^{-1} < \omega < \tau^P$. Assume that income differences are sufficiently large, $\omega L^P/L^R < \tau^R$, so that an arbitrage equilibrium prevails. To prevent arbitrage, the price of traded goods in the rich country may not exceed the price in the poor country plus transportation costs, hence firms will charge $p_T^R = \tau^R p^P$ in the rich country.[15]

Using zero-profit conditions and resource constraints is it is straightforward to calculate welfare in the two countries as

$$U^P = \frac{a\mathcal{P}^P + a\tau^R \mathcal{P}^R}{aF + \tau^R \mathcal{P}^R + \mathcal{P}^P} L^P \quad \text{and} \quad U^R = U^P + \frac{a\mathcal{P}^R}{aF + \mathcal{P}^R}\left(L^R - \tau^R \omega L^P\right).$$

Interestingly, a unilateral trade liberalization by the poor country (a fall in τ^P) does not have any effect on poor households, but affects rich households through a fall in ω. Lower costs of exporting to the poor country makes producers in country R more productive, improving their

[15] To make sure that such an equilibrium exists, we also assume that country-R exporters can charge a price in country P that covers (production plus transportation) costs, $p^P > \tau^P/a$. This implies $\tau^P \tau^R < aF/\mathcal{P}^R + 1$. If this condition is satisfied also country-P exporters will export, $p_T^R > \tau^R/a$ because $p_T^R = \tau^R p^P$.

terms of trade while leaving the arbitrage constraint unaffected. This saves resources for country R which are employed to produce non-traded goods. This raises welfare of rich consumers. In contrast, an unilateral increase in trade barriers into the rich country (a larger τ^R) harms country-R but benefits country-P households. Hence, our model predicts that a poor country has an incentive to levy an export tax. This relaxes the arbitrage constraint and increases the supplied varieties and hence welfare in country P.

3.3 Existence of equilibria

The conditions under which the threat of parallel trade becomes binding and the economy switches from a full trade to a partial trade equilibrium are straightforward. In a full trade equilibrium, relative prices equal relative per capita incomes $p^P/p^R = \omega L^P/L^R$. In that case, differences in willingnesses to pay must be small enough, $\lambda^P/\lambda^R \leq \tau$, so that the threat of parallel trade is not binding. In contrast, when differences in willingnesses to pay become large, $\lambda^P/\lambda^R > \tau$, the parallel trade constraint kicks in. This happens when

$$\frac{\omega L^P}{L^R} > \tau^{-1}. \tag{7}$$

In other words, a full trade equilibrium emerges when per capita incomes are similar, while an arbitrage equilibrium emerges when the gap in per capita incomes is large.

Up to now we have implicitly assumed that trade costs are sufficiently low so that the two countries will engage in trade. The following proposition proves existence of a general equilibrium with trade.

Proposition 5 *When $\tau \leq \tau^* \equiv \sqrt{aF/\mathcal{P}^R + 1}$, the two countries will trade with each other for all $L^P/L^R \in (0, 1]$.*

Proof. See Appendix C. ■

The trade condition in the proposition makes sure prices in country P are sufficiently high to induce country-R firms to export their product. Notice that, with $\tau \leq \tau^*$, country-P firms are also willing to export since they can charge a price $p_T^R > p^P < \tau/a$. The trade condition is quite intuitive. Trade is more valuable when fixed costs are high, as these costs are spread out over a larger market. For the same reason, trade is more valuable if the local market is small. Hence the critical value of iceberg costs τ^* is increasing in F and falling in $\mathcal{P}^R$. Notice also that the trade condition makes a statement about the relative size of trade costs and the square root of the mark-up. One could argue that empirically observed mark-ups are often lower than observed trade costs, thus contradicting the trade condition. Notice, however, that $aF/\mathcal{P}^R + 1$ is the mark-up of a (northern) firm that sells its product exclusively on the home market, while the average mark-up in the economy is a weighted average of these non-exporting firms (with high mark-ups) and exporting firms (with low mark-ups). Hence, our simple model can accommodate a situation where there is trade, even though the economy-wide mark-up falls short of trade costs.

Figure 2 shows the relevant equilibria in $(L^P/L^R, \tau)$ space. There is full trade in region **F** which emerges at high values of L^P/L^R and intermediate values of τ. An arbitrage equilibrium prevails in region **A** which arises at low trade costs and high income differences. Figure 2 also shows what happens when population size in the poor country increases. In that case, the downward-sloping branch that separates regions **F** and **A** shifts to the left. When the poor country is larger, τ^* is unaffected and there are more parameter constellations $(L^P/L^R, \tau)$ under which a full trade equilibrium emerges. In this sense, a larger population in the poor country fosters trade.[16]

Figure 2

Figure 3 shows the welfare responses of changes in τ across the various regimes graphically. Panel a) is drawn for relatively low per capita income differences $\omega L^P/L^R > \tau^*$. In that case, an arbitrage equilibrium emerges with low trade costs, while a full trade equilibrium emerges with moderate trade costs. Panel b) is drawn for higher per capita income differences $\omega L^P/L^R \leq \tau^*$ so that a full trade equilibrium is not feasible. Country-R welfare (the bold graph) is monotonically decreasing in τ in both panels of Figure 3. Hence the R-consumer reaches his maximum welfare when trade costs are at their lowest possible level $\tau = 1$. In contrast, the impact of τ on country-P welfare (the dotted graph) interacts with per capita income differences. When these differences are low (panel a), country-P welfare increases in τ when $\tau < \left(\omega L^P/L^R\right)^{-1}$ and decreases in τ when $\tau \geq \left(\omega L^P/L^R\right)^{-1}$. Welfare is maximized at $\tau = \left(\omega L^P/L^R\right)^{-1}$ (when the equilibrium switches from a full-trade to an arbitrage equilibrium). When per capita income differences are large (panel b), country-P welfare decreases monotonically in τ (full trade is not feasible) and welfare is maximized at $\tau = \tau^*$.

Figure 3

4 Many rich and poor countries

In an arbitrage equilibrium with two countries, all firms in the poor country are exporters while only a subset of firms in the rich country exports. Moreover, a trade liberalization that relaxes the arbitrage constraint always hurts poor consumers. We now show that these predictions need to be qualified in a multi-country world. The effect of moving from two to many countries can be most easily shown when there are n identical rich countries and m identical poor countries, i.e. a world with a fragmented rich North and a fragmented poor South. As before, we assume that countries differ in per capita endowments (and population size) but are identical in all other respects.

The general equilibrium has a structure very similar to that of the two-country case. From the zero-profit conditions for internationally active firms, it is straightforward to show that relative

[16] Notice that there is international trade even when income differences become extremely large and L^P/L^R becomes very small. The range of traded goods approaches zero, however, when L^P/L^R goes to zero.

wages are now given by

$$\omega \equiv \frac{W^R}{W^P} = \frac{aF + \tau\mathcal{P}^{-R} + \mathcal{P}^R}{aF + \tau\mathcal{P}^{-P} + \mathcal{P}^P},$$

where $\mathcal{P}^{-R} = (n-1)\mathcal{P}^R + m\mathcal{P}^P$ and $\mathcal{P}^{-P} = n\mathcal{P}^R + (m-1)\mathcal{P}^P$ are rest-of-the-world populations from the perspective of country R and country P, respectively. In full world trade equilibrium, relative prices of southern relative to northern markets are determined by relative per capita incomes, $p^P/p^R = \omega L^P/L^R$, and the ratio of produced varieties still reflects differences in aggregate GDP, $N^P/N^R = \omega L^P\mathcal{P}^P/L^R\mathcal{P}^R$.

The interesting case is when income differences are sufficiently large, so that $\omega L^P/L^R > \tau^{-1}$. In that case, the arbitrage constraint is binding, limiting trade between the rich North and the poor South. A northern firm now has two options: either export worldwide or export only to other northern countries. Notice that, unlike in the two-country case, all northern firms are now exporters. Firms that export exclusively to the North have a smaller market but can charge higher prices. Firms that export to all countries worldwide set low prices but have the large world market. While large differences in per capita incomes limit trade *across* regions, there is full trade *within* regions. As there are no income differences within a region, all goods produced in that region are also sold to other countries in that region.

The arbitrage equilibrium can now be solved in a straightforward way (for details see Appendix E). We first study how differences in per capita incomes and population sizes affect trade intensity. It is straightforward to calculate

$$\phi = 2\frac{m\omega L^P\mathcal{P}^P}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}\frac{(m-1)\mathcal{P}^P + \tau\mathcal{P}^R}{m\mathcal{P}^P + n\tau\mathcal{P}^R} + 2\frac{(n-1)L^R\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}.$$

which readily reduces to the expression derived in the last section when $n = m = 1$. An increase in L^P increases world trade intensity (and reduces North-North trade with exclusive goods). It can also be shown that a larger population in the South has a weaker effect on trade intensity than a larger per capita income. Hence with respect to per capita incomes and populations sizes, the results of the two-country case carry over to the multi-country framework.

In contrast to the two-country case, the effect of a trade liberalization on welfare is now ambiguous. There are two effects. On the one hand, a lower τ implies a tighter arbitrage constraint for globally active producers. Lower prices for globally traded products (relative to products exclusively sold in the North) induce former northern world-market producers to concentrate their sales on northern markets only. This reduces trade intensity between the North and the South. On the other hand, a reduction in τ stimulates trade within regions. While South-South trade increases less than North-South trade falls (first term of above equation increases in τ), North-North trade unambiguously increases (second term decreases in τ). A trade liberalization is more likely to stimulate trade if there are more countries with a region. Within-regional trade is more strongly affected in this case and dominates the reduction in North-South trade. A trade liberalization is also more likely to stimulate North-South trade, the larger is the North relative to the South. In that case, North-North trade (which is positively affected) comprises the bulk of world trade. When the North is much larger than the South, positive effects on North-North trade of a trade liberalization dominate negative effects on North-South trade flows.

It turns out that the welfare level of a country-P household is given by

$$U^P = mN^P + nN_N^R = \frac{aL^P(m\mathcal{P}^P + \tau n\mathcal{P}^R)}{aF + \tau\mathcal{P}^{-P} + \mathcal{P}^P}.$$

It is straightforward to see that $\partial U^P/\partial\tau < 0$ if $aF < (m-1)\mathcal{P}^P(1 + (m/n)(\mathcal{P}^P/\mathcal{P}^R))$. This means that a trade liberalization may raise welfare in country P and is more likely to do so the higher is $m\mathcal{P}^P$. A reduction in τ has two opposing effects. The arbitrage channel is still at work and induces northern firms to abstain from selling to southern households. This is harmful for southern welfare. However, a lower τ stimulates South-South trade, which has a beneficial effect on southern welfare. Households in the poor country gain from a trade liberalization when there are many poor countries and when poor countries are large. In such a situation, there is a lot to gain from South-South trade because there are many trade barriers and because the southern markets are large.

We summarize the above discussion in the following

Proposition 6 *Assume there are m identical poor countries and n identical rich countries, with $L^P/L^R < (\omega\tau)^{-1}$. a) All northern firms export, but some of them export only to other northern countries; b) A trade liberalization (a lower τ) unambiguously increases welfare of rich households. It increases welfare of poor households if $aF < (m-1)\mathcal{P}^P(1+(m/n)(\mathcal{P}^P/\mathcal{P}^R))$ and decreases it otherwise. The increase in welfare is larger in the North than in the South.*

Proof. In text. ■

Notice that the multi-country model generates an empirically testable hypothesis. The model predicts a positive correlation between the export probability of a northern firm and the per capita income of a potential destination. In our simple model, the probability that a firm from a rich country exports to another rich country is 100 percent. (The prediction that 100 percent of all firms export is clearly an artefact arising from the assumed absence of firm heterogeneity.) In contrast, the probability that a northern firms exports to a poor country is less than 100 percent and is lower the poorer a potential destination. Below, we will test this prediction by investigating whether and to which extent US export probabilities of HS6-digit product categories are indeed positively related to potential destinations' per capita incomes.

5 General preferences

The assumption of 0-1 preferences yields a tractable framework with closed-form solutions. However, the focus is entirely on the extensive margin of consumption. This contrasts with the standard CES case where all adjustments happen along the intensive margin. We go beyond these two polar cases in this section by studying general preferences. We show that the qualitative characteristics of the equilibria under 0-1 preferences carry over to general preferences featuring non-trivial intensive *and* extensive margins of consumption.[17] In particular, we precisely define the conditions under which an arbitrage equilibrium with non-traded goods exists and also provide a simple calibration exercise showing that arbitrage equilibria emerge under reasonable parameter values and that there is a quantitatively strong relationship between per capita incomes on the extensive margin of trade.

5.1 Utility and prices

Let us go back to the setup of Section 3 with two countries that differ in per capita income and population size. However, let household welfare take the general form

$$U = \int_0^\infty v(c(j))dj, \tag{8}$$

where $c(j)$ denotes the consumed quantity of good j. It is assumed that the subutility $v()$ satisfies $v' > 0$, $v'' < 0$ and $v(0) = 0$. Beyond these standard assumptions, we make two further assumptions on the function $v()$: (i) $v'(0) < \infty$, (ii) $v''(0) > -\infty$, and (iii) $-v'(c)/[v''(c)c]$ is decreasing in c. The first assumption implies that reservation prices are finite, generating a non-trivial extensive margin of consumption; the second ensures that an arbitrage equilibrium exists when per capita income differences are sufficiently high (see below); and the third implies a price elasticity of demand decreasing along the demand curve. Monopolistic pricing leads to $p = (1 + v''(c)c/v'(c))^{-1}b$, where b denotes marginal cost. To simplify notation, we denote the mark-up by $\mu(c) \equiv (1 + v''(c)c/v'(c))^{-1}$. Assumptions (i)-(iii) imply that $\mu(0) = 1$ and $\mu'(c) > 0$.[18]

How does firms' price setting behavior change when there are consumer responses along the intensive margin? With 0-1 preferences, the monopoly price equals the representative household's willingness to pay and does not depend on marginal production costs. With general preferences, however, firms solve the standard profit maximization problem: the price equals marginal costs times a mark-up that depends on the price elasticity of demand. This implies an important difference to the case of 0-1 preferences. With general preferences, there are price differences between imported and domestically produced goods. While symmetric utility implies that importers and local producers within a given location face the same demand curve, marginal costs differ since importers have to bear transportation costs and since wages vary by location. To allow for such differences, we denote by p^i_j, c^i_j and b^i_j, respectively, the price, quantity and marginal cost of a good produced in country j and consumed in country i. Unconstrained monopoly pricing implies $p^i_j = \mu(c^i_j)b^i_j$.

[17] Li (2012) and Bronnenberg (2015) propose other ways to study the trade-off between intensive and extensive margins of consumption by sticking to CES preferences but allowing for fixed purchasing costs.

[18] $\mu'(c) > 0$ follows directly from assumption (iii). To see why $\mu(0) = 1$ we use l'Hopital's rule $\lim_{c\to 0} v'(c)c/v(c) = \lim_{c\to 0}(1 + v''(c)c/v'(c))$. However, $\lim_{c\to 0} v'(c)c/v(c) = v'(0) \cdot \lim_{c\to 0} c/v(c) = v'(0)/v'(0) = 1$. This implies $\lim_{c\to 0} v''(c)c/v'(c) = 0$ and hence $\lim_{c\to 0}\mu(c) = 1$. Since the monopolist optimally chooses a price along the elastic part of the demand curve, no further restrictions on the $\mu(c)$-function are needed.

5.2 The arbitrage equilibrium

The arbitrage equilibrium features a situation in which (i) only a subset of country-R producers sell their product worldwide at sufficiently low prices to avoid arbitrage; (ii) the remaining country-R firms sell their product exclusively in the rich country at the unconstrained monopoly price; (iii) all poor-country producers export their products, also at prices that avoid arbitrage. The discussion in this section focuses on the conditions under which an arbitrage equilibrium exists. (Appendix E provides the full system of equations that characterize such an equilibrium.)

The arbitrage constraints for country-R and country-P producers, respectively, are now given by

$$1/\tau \leq p_R^R/p_R^P \leq \tau \text{ and } 1/\tau \leq p_P^P/p_P^R \leq \tau.$$

A necessary condition for the existence of an arbitrage equilibrium is that these constraints are binding, so that $p_R^R/p_R^P = p_P^R/p_P^P = \tau$. This happens to be the case if the gap in per capita incomes becomes sufficiently large. As L^R/L^P, and hence c_R^R/c_R^P, get large the ratio of (unconstrained) monopoly prices eventually exceeds trade costs, or $\mu(c_R^R)/\mu(c_R^P) > \tau^2$. (Recall that $\mu'(c) > 0$.) Notice, however, that a binding arbitrage constraint does not necessarily imply that there are non-traded goods. The reason is that adjustment now does not only occur at the extensive margin but also at the intensive margin. Hence there are full trade equilibria where the arbitrage constraint binds.

To verify the existence of an arbitrage equilibrium with non-traded goods, we look at incentives of country-R firms to sell exclusively on the home market rather than selling their products worldwide. A country-R producer's profit is given by (to ease notation we write $p_R^R \equiv \tau p$ and $p_R^P \equiv p$)

$$\pi = \mathcal{P}^R \left(\tau p - 1/a\right) c_R^R + \mathcal{P}^P \left(p - \tau/a\right) c_R^P.$$

The corresponding demand curves are given by the first order conditions $v'(c_R^R) = \lambda^R \tau p$ and $v'(c_R^P) = \lambda^P p$ for households in country-R and and country-P, respectively. This yields $dc_R^R/dp = (1/p)v'(c_R^R)/v''(c_R^R)$ and $dc_R^P/dp = (1/p)v'(c_R^P)/v''(c_R^P)$. The first order condition of the monopolistic firm's price setting choice is given by

$$\frac{\tau p - 1/a}{p}\left(-\frac{v'(c_R^R)}{v''(c_R^R)}\right) + \frac{p - \tau/a}{p}\left(-\frac{v'(c_R^P)}{v''(c_R^P)}\right)\frac{\mathcal{P}^P}{\mathcal{P}^R} = \tau c_R^R + c_R^P \frac{\mathcal{P}^P}{\mathcal{P}^R}.$$

To examine whether an arbitrage equilibrium exists, let L^P and therefore c_R^P approach zero, all other exogenous variables (including $\mathcal{P}^P/\mathcal{P}^R$) remain fixed. The first order condition then becomes

$$\frac{\tau p - 1/a}{\tau p}\left(-\frac{v'(c_R^R)}{v''(c_R^R)c_R^R}\right) + \frac{p - \tau/a}{\tau p c_R^R}\left(-\lim_{c_R^P \to 0}\frac{v'(c_R^P)}{v''(c_R^P)}\right)\frac{\mathcal{P}^P}{\mathcal{P}^R} = 1.$$

Now consider the optimal decision of a country-R firm whether to produce exclusively for the home market. Denoting by p^N and c_R^N price and quantity of non-traded goods, the first order condition for exclusive producers is

$$\frac{p^N - 1/a}{p^N}\left(-\frac{v'(c_R^N)}{v''(c_R^N)c_R^N}\right) = 1.$$

When τ is sufficiently low, so that $p > \tau/a$, comparing the last two equations shows that the price of a non-exporting firm p^N is strictly larger than the price of an exporting firm τp. (This is because, by assumption, $-v'(0)/v''(0) > 0$.) Since $c_R^P \to 0$ when $L^P \to 0$, export revenues are zero. Hence non-exporters charge higher prices and their profits are larger than those of exporters. This implies that an outcome where all firms export cannot be an equilibrium. We summarize our discussion in

Proposition 7 *There is a critical income gap Δ such that, for all $L^P/L^R < \Delta$, an equilibrium emerges in which only a subset of goods is traded.*

Proof. In text. ■

The above proposition implicitly assumes an equilibrium where the two countries trade with each other. This is not a priori clear because the countries may also remain in autarky. The following proposition shows that transportation costs need to fall short of a certain limit to make sure that trade will take place in equilibrium.

Proposition 8 *Denote by c_a^R consumption per variety under autarky in the rich country. There will be trade in equilibrium, if $\tau < \mu(c_a^R)v'(0)/v'(c_a^R)$ where $aF/\mathcal{P}^R = c_a^R(\mu(c_a^R) - 1)$.*

Proof. See Appendix F. ■

An important result we derived under 0-1 preferences holds that population size has a weaker effect than per capita incomes in determining trade patterns. We now demonstrate that this is also true with general preferences. The previous proposition showed that, starting from a full trade equilibrium, increasing the gap in per capita incomes will eventually generate an arbitrage equilibrium with non-traded goods. We now show this is *not* necessarily the case, when we increase relative population size.

To make this point, we proceed as follows. We first observe that, starting from a full trade equilibrium, an increase in L^P/L^R beyond unity eventually leads to a "reversed" arbitrage equilibrium, in which some country-P producers sell only on the domestic market while all country-R producers export. We now show that such a reversed arbitrage equilibrium *cannot* emerge from a successive increase in $\mathcal{P}^P/\mathcal{P}^R$ (keeping $L^P/L^R < 1$ constant), because this does *not* generate price differences sufficiently large to escape a full trade equilibrium. In other words, increasing $\mathcal{P}^P/\mathcal{P}^R$, we cannot reach a situation where both arbitrage constraints are violated, $\mu(c_R^P) \geq \mu(c_R^R)$ and $\mu(c_P^P) \geq \mu(c_P^R)\tau^2$. To see this, consider the households' budget constraints

$$\begin{aligned} aL^R &= N_P\mu(c_P^R)c_P^R\tau\omega + N_R\mu(c_R^R)c_R^R \\ aL^P &= N_P\mu(c_P^P)c_P^P + N_R\mu(c_R^P)c_R^P\tau/\omega, \end{aligned}$$

and take the difference between the two equations. If both arbitrage conditions are violated, the budget constraints can only hold if $\omega > \tau$. However, if $\omega > \tau$ the zero-profit condition is violated in at least one country in a full trade equilibrium (where firms charge the unconstrained monopoly price). In such an equilibrium, the zero-profit condition in country j is given by

$$\mathcal{P}^R c_j^R(\mu(c_j^R) - 1)/a + \mathcal{P}^P c_j^P(\mu(c_j^P) - 1)\tau/a = F,$$

where $p_P^i > p_R^i$ and $c_P^i < c_R^i$, since country P has higher marginal cost than country R, both on the domestic and the export market. However, this implies $c_P^i(\mu(c_P^i) - 1) < c_R^i(\mu(c_R^i) - 1)$ for both $i = P$ *and* $i = R$, i.e. country R-producers make strictly larger profits on both markets. It follows that, when the zero-profit condition holds in country R, it must be violated in country P, and vice versa, if country P is the low-wage country. In the latter case, we must have $\omega > 1/\tau$ to ensure that both zero-profit conditions can hold simultaneously. Hence we have $\omega \in (1/\tau, \tau)$ in a full trade equilibrium with unconstrained price setting. In sum, we always have $\omega < \tau$ in a full trade equilibrium. But this implies that households' budget constraints continue to hold simultaneously when $\mathcal{P}^P/\mathcal{P}^R$ gets very large. Thus, unlike a successive increase in L^P/L^R (beyond unity), it is not possible to reach a "reversed" arbitrage equilibrium with a successive increase in $\mathcal{P}^P/\mathcal{P}^R$. In this sense, the difference in population sizes has a weaker effect on trade patterns than the difference in per capita endowments. We summarize our discussion in the following

Proposition 9 *Per capita endowments are more important for trade patterns than population sizes. Starting from a full trade equilibrium with unconstrained price setting, successive increases in L^P/L^R (beyond unity) lead to a "reversed" arbitrage equilibrium, while successive increases in $\mathcal{P}^P/\mathcal{P}^R$ cannot generate a reversed arbitrage equilibrium.*

Proof. In text. ■

6 Reduced-form empirical evidence

Before we start with a model-based approach to shed light on the empirical relevance of arbitrage, it seems useful to test the prediction (developed at the end of section 4) of a positive relationship between a rich country's export probability and a destination's per capita income. If the arbitrage constraint is at work, such a positive relation between the extensive margin of exports and the per capita income of a potential destination should be observed in the data.

We analyze the following empirical model

$$D(i,k) = \alpha_0 + \alpha_1 \ln GDP(k) + \alpha_2 \ln y(k) + X(i,k)\beta + \phi(i) + e(i,k),$$

where $D(i,k)$ indicates whether the US exports product i to country k, $GDP(k)$ is aggregate GDP of country k, and $y(k)$ denotes per capita income of country k. $X(i,k)$ is a vector of controls,[19] $\phi(i)$ is a product-fixed effect, and $e(i,k)$ is an error term.

[19]Control variables include: log of distance between exporter's and importer's capital, dummy for a common

We use UN Comtrade data complied by Gaulier and Zignago (2010) containing yearly unidirected bilateral trade flows at the 6-digit-level of the Harmonized System (1992) for the year 2007. We observe 5,018 product categories at the 6-digit level. We look only at consumer goods (according the BEC classification). This leaves us with 1,263 product categories from which we exclude those 11 categories the US did not export in 2007. Our data set includes 135 potential export destinations. Information on per capita incomes (2005 PPP-adjusted USD) and population sizes are taken from Heston et al. (2006). We exclude all bilateral trade flows with negative quantities and set $D(i,k) = 0$ when the observed quantity falls short of USD 2,000. We end up with 169,020 potential export flows (1,252 products × 135 potential importers). 39.1 percent of these potential export flows actually materialized in 2007.

A crucial prediction of our model is $\alpha_2 > 0$: a destination's per capita income is a significant determinant for the export probability, conditional on the destination's aggregate GDP. In the standard homothetic model we should have $\alpha_2 = 0$, since there is no extra role for a destination's per capita income once aggregate GDP is controlled for. The estimates of Table 1, column 1, clearly indicate that α_2 is positive, statistically significant, and qualitatively large: doubling a destination's per capita income, holding its aggregate GDP constant, increases the US export probability of a HS6 digit product category by 8.5 percentage points. The estimates of column 1 also imply that a destination's per capita income is more important than its population size. To see this, denote by $Pop(k)$ destination k's population size, so that $\ln Pop(k) = \ln GDP(k) - \ln y(k)$. This lets us rewrite the empirical model as $D(i,k) = \alpha_0 + (\alpha_1 + \alpha_2) \ln y(k) + \alpha_1 \ln Pop(k) + ...$ Hence doubling the per capita income (holding population size constant) increases the export probability by 14.9 (= 8.5 + 6.4) percentage points, while doubling population size (holding per capita income constant) increases it by only 6.4 percentage points. We conclude that the empirical evidence is consistent with the predictions of our model, according to which per capita income play a more important role than populations size to explain the extensive margin of international trade (see Propositions 3 and 9).

Table 1

To check the robustness of our estimates, columns 2-6 of Table 1 replace the regressor $\ln y(k)$ by a set of dummy variables to allow for a more flexible impact of per capita incomes on export probabilities.[20] In column 2, we use the same sample as in the log specification of column 1; column 3 excludes very small destination (with population size less than 1 million); and column 4 and 5 aggregate export probabilities to HS4 and HS2 digit levels, respectively. Results indicate

border, dummy for importer being an island, dummy for importer being landlocked, dummy for importer and exporter ever having had colonial ties, dummy for currency union between importer and exporter, dummy for importer and exporter sharing a common legal system, dummy for religious similarity, dummy for importer and exporter having a free trade agreement, and dummy for importer and exporter sharing a common language.

[20] The estimated model is $D(i,k) = \tilde{\alpha}_0 + \tilde{\alpha}_1 \ln GDP(k) + \sum_{n=1}^{6} \tilde{\alpha}_{2n} D_y(k,n) + X(i,k)\tilde{\beta} + \tilde{\phi}(i) + \tilde{e}(i,k)$, where $D_y(k,n)$ indicates whether of not destination k falls into per capita income category n. We classify countries into 7 per capita income groups: (i) lower than USD 1000; (ii) USD 1000-1999; (iii) USD 2000-3999; (iv) USD 4000-7999; (v) USD 8000-15999; (vi) USD 16000-31999; and (vii) USD 32000 or larger. The group with per capita income larger than USD 32,000 serves as the reference group.

a monotonic impact of per capita income, which is robust across specifications.[21]

Obviously, a positive relationship between export probabilities and per capita incomes does not mean that only arbitrage constraints are driving the empirical evidence. Export zeros could also be the results of cost constraints (i.e. the marginal cost of exporting to particular destination exceeds that destination's reservation price). Standard heterogeneous-firm models can accommodate the results of Table 1 only with additional parameters, e.g. by assuming heterogeneous market-entry costs depending on the destination's per capita income. Simple specifications of non-homothetic preferences that have been frequently used in the earlier literature (e.g. Bergstrand 1990, Mitra and Trindade 2005) typically work with a homogenous and a differentiated good. Such specifications of non-homotheticities generate a bang-bang solution: either consumers buy no differentiated goods or all of them. Again, additional parameters like preference heterogeneity across countries would have to be introduced to produce export zeros. More recent papers mentioned in the introduction combine non-homothetic preferences and supply heterogeneities to generate non-negativity constraints so that export zeros can emerge. Clearly, the empirical evidence of Table 1 cannot discriminate between export zeros arising from a reservation price falling short of marginal production costs; and export zeros arising from international arbitrage due to large per capita income differences. Therefore, in the next section, we go one step further and check quantitatively whether binding arbitrage can arise in a calibrated standard model, thus speaking more specifically to the empirical relevance of the arbitrage channel.

7 Are no-arbitrage constraints binding in a calibrated model?

In this section, we calibrate a multi-country trade model with non-homothetic preferences. This model makes predictions about prices in the various destination but rules out arbitrage by assumption. The idea is to check whether the price distribution predicted in such a model is such that arbitrage opportunities exist. If this is the case for many exporters (and a high export volume), we take it as evidence that points to the potential relevance of international arbitrage.[22]

Arbitrage in a multi-country framework. In a multi-country framework, arbitrage opportunities can not only arise *directly* within a country pair but also *indirectly* through a third country. A *direct* arbitrage opportunity occurs through parallel-imports – when the relative price between origin i and destination j exceeds trade costs τ_{ji} (so that arbitrageurs can buy cheaply in j and undercut the price in i); or through parallel-exports – when the relative price between

[21] In Tables A1 and A2 of the Online Appendix, we provide further robustness checks. In Table A1, we show that US regressions for each single year 1997-2006 yield estimates very close to those obtained for the year 2007, which we report in Table 1. In Table A2 we look at the 14 largest consumer goods exporters (rather than only the US). For all these large exporting countries we find a significant effect of destinations' per capita income, holding destinations' GDP constant (the only exception being China and Mexico where the effect is not statistically significant). These results are also in line with the evidence in Baldwin and Harrigan (2011) and Bernasconi and Wuergler (2012) where, respectively, output per worker and income per capita are included as control variables in extensive margin regressions.

[22] We thank an anonymous referee for making this suggestion.

destination j and origin i exceeds trade costs τ_{ij} (so that arbitrageurs can buy cheaply in i and undercut in j).

Importantly, however, in a world economy with many countries, the arbitrage channel may also work *indirectly* through third countries. Consider the example of a US firm that sells its product in Korea. Korean consumers are poorer than US consumers. Therefore, the US firm wants to charge a lower price in Korea. It can do so because trade costs from Korea to the US are relatively high and as a consequence arbitrageurs do not find it profitable to ship the goods back to the US. However, if the same US firm also serves the Chinese market, it is likely to be indirectly constrained on the Korean market. The reason is that arbitrageurs may purchase cheaply on the Chinese market and underbid in Korea. Such a scenario is likely since income differences between Korea and China are high, while trade costs are comparably low. In fact, such a constellation emerges in the numerical analysis discussed below.

The above discussion suggests the following simple rule to identify whether a producer from country i is subject to (direct or indirect) arbitrage in destination j: we need to check, for all possible country triples (i, j, k), whether the following inequality holds

$$\frac{p_{ij}}{p_{ik}} > \tau_{kj}, \tag{9}$$

where p_{ij} is the price charged by the producer from country i in destination j and τ_{kj} are the trade costs from country k to country j.

Equation (9) encompasses both cases of direct and indirect arbitrage. Indirect cases are represented by $i \neq j \neq k$. Direct cases are represented by $k = i \neq j$ (parallel exports) and $i = j \neq k$ (parallel imports). The $I - 1$ countries from where a threat of arbitrage potentially originates include the $I-2$ third countries (indirect arbitrage) and one of the two trading partners (direct arbitrage). A threat of indirect arbitrage emerges when $i \neq j \neq k$ (purchasing country i goods cheaply in destination k and undercutting in destination j). A threat of parallel exports occurs when the above inequality holds for $k = i \neq j$ (purchasing cheaply in the home market i and underbid in destination j). A threat of parallel imports occurs when the above inequality holds for $i = j \neq k$ (purchasing cheaply in destination k and shipping it back to origin j). To sum up, exporter i is constrained in destination j, if the above inequality holds for *at least one* of the other $I - 1$ destination markets k (excluding market j). Conversely, a producer from i is unconstrained in j, if and only if the above inequality does not hold (and hence the arbitrage constraint is not binding) for *any* of the $I - 1$ other destinations.

Numerical analysis. We use the model by Simonovska (2015) who introduces Stone-Geary preferences into the multi-country heterogeneous firm model. Her model provides a useful framework for our purpose, since Stone-Geary preferences belong to the class of utility functions (studied in section 5) where arbitrage constraints may, in principle, be binding. Moreover, there is now a well-established approach to quantifying this type of models (exemplified by Simonovska and Waugh, 2014, whose data we use). Notice that the prices will be unconstrained monopoly prices, as Simonovska (2015) does not consider arbitrage. We will use the quantified model to compute implied relative prices and trade costs and to check then if and how often the arbitrage constraint (9) holds.

The Simonovska (2015) model features many countries, indexed by $i = 1, ..., I$, that differ in population size $\mathcal{P}_i$ and that face exporter-destination-specific trade costs τ_{ij}, where we set $\tau_{ii} = 1$. It also allows for heterogenous firms within countries, which differ in productivity a drawn from a (country-specific) Pareto distribution, $G_i(a) = 1 - (b_i/a)^{\theta}$. All countries have the same θ but differ in the parameter b_i. Ceteris paribus, countries with a higher b_i are more productive (and hence richer) than other countries. The utility function takes the Stone-Geary form $v(c) = \log(c + \gamma) - \log \gamma$, where $\gamma > 0$ is a normalization constant to ensure $v(0) = 0$. Utility is the same for all goods and countries.

Since arbitrage is ruled out, firms set the unconstrained monopoly price in each country. Under the maintained assumptions the price in destination j charged by a country-i firm with productivity a is

$$p_{ij}(a) = (\tau_{ik} W_i/a)^{\frac{1}{2}} \left(W_j L_j F(\theta+1)(1+2\theta)/\gamma\right)^{\frac{1}{2(1+\theta)}} \left(\sum_{\nu=1}^{I} L_\nu \mathcal{P}_\nu (b_\nu)^{\theta} (\tau_{\nu j} W_\nu)^{-\theta}\right)^{-\frac{1}{2(1+\theta)}}$$

where $W_i L_i$ is the income level of the representative consumer in country i. From this expression, the relative price in destinations j and k of a firm producing in country i is given by

$$\frac{p_{ij}(a)}{p_{ik}(a)} = \left(\frac{W_j L_j}{W_k L_k}\right)^{\frac{1}{2(\theta+1)}} \left(\frac{\tau_{ij}}{\tau_{ik}}\right)^{\frac{1}{2}} \left(\frac{\sum_{\nu=1}^{I} L_\nu \mathcal{P}_\nu b_\nu^{\theta} (\tau_{\nu j} W_\nu)^{-\theta}}{\sum_{\nu=1}^{I} L_\nu \mathcal{P}_\nu b_\nu^{\theta} (\tau_{\nu k} W_\nu)^{-\theta}}\right)^{-\frac{1}{2(\theta+1)}}, \tag{10}$$

where the last term in equation (10) captures relative market sizes of j and k. Note that the relative price in (10) is independent of the firm's productivity a. This is a consequence of the Stone-Geary assumption and does not hold generally with non-homothetic preferences. Simonovska (2015) shows that the expenditure share in j for goods produced in i is

$$\lambda_{ij} = \frac{L_i \mathcal{P}_i b_i^{\theta} (\tau_{ij} W_i)^{-\theta}}{\sum_{v=1}^{I} L_\nu \mathcal{P}_v b_v^{\theta} (\tau_{vj} W_v)^{-\theta}}, \tag{11}$$

which allows us rewrite equation (10) as

$$\frac{p_{ij}(a)}{p_{ik}(a)} = \left(\frac{W_j L_j}{W_k L_k}\right)^{\frac{1}{2(\theta+1)}} \left(\frac{\tau_{ij}}{\tau_{ik}}\right)^{\frac{2\theta+1}{2(\theta+1)}} \left(\frac{\lambda_{ij}}{\lambda_{ik}}\right)^{\frac{1}{2(\theta+1)}}. \tag{12}$$

Equations (10) and (12) will be central in what follows since they allow us to compute the price that a country-i producer charges in destination j relative to the price charged in destination k under the assumption that there is no international arbitrage. We can then compare this relative price to the trade costs between k and j to check if there is an incentive to buy the good cheaply in k, ship it to j and undercut there the country-i producer. Note that (10) and (12) correspond to equations (22) and (24) in Simonovska (2015).

Data and baseline numerical analysis. In our baseline numerical analysis we use equation (12). We therefore need values for the parameters/variables W_jL_j, τ_{ij}, λ_{ij}, and θ. We use the dataset of Simonovska and Waugh (2014) that contains $I = 123$ countries for the year 2004.[23] To measure the above parameters/variables we proceed as follows: (i) as a measure of W_jL_j, we use per capita income of country j, which can be taken directly from the dataset. (ii) As a measure of λ_{ij}, we use imports of j from i as a fraction of the manufacturing absorption of country j (i.e. total manufacturing output minus net exports). Simonovska and Waugh (2014) compute these shares based on source data from the UN Comtrade database and UNIDO. (iii) Trade costs τ_{ij} cannot be directly observed; we obtain estimates for τ_{ij} by using the gravity approach of Simonovska and Waugh (2014).[24] In principle, this method may generate trade costs below unity. When this happens, we set $\tau_{ij} = 1$. Note that in our context, this is a conservative choice, since lower trade costs would make arbitrage more likely. (iv) We are agnostic about the specific value of the Pareto parameter θ. We provide results for $\theta = 2$, 4, 6, 8, which covers the range of estimates provided by the literature.

In the online appendix, we perform a number of robustness checks to explore the sensitivity of our findings with respect to alternative choices for (i) - (iii) above. We also present the results based on equation (10) instead of (12). (Data and programs to derive these estimates and the results shown below are provided in an online package.)

Based on the above measures, we compute the relative prices according to (12) for all country triples (i, j, k) by combining the observed per capita incomes and expenditure shares with the estimated trade costs. We then use these relative prices and the estimated trade costs to check if the arbitrage condition (9) is binding. Note that the price ratio (12) goes toward infinity whenever the trade flow from i to k is zero. In this case the arbitrage condition automatically becomes binding. To avoid inflating our results by this, we focus in the baseline specification on positive trade flows.

Baseline results. An exporter from country i is subject to arbitrage in destination j, if the arbitrage condition (9) holds for at least one of the $I - 1$ other destinations. This suggests the

[23] We thank Michael Waugh for providing the dataset. For details on the dataset see the data appendix of Simonovska and Waugh (2014).

[24] Simonovska and Waugh (2014) specify trade costs as $\log \hat{\tau}_{ij} = \alpha_0 d_{k,ij} + \alpha_1 b_{ij} + \alpha_2 ex_i$ where $d_{k,ij}$ measures the distance between countries i and j (measured in 6 discrete intervals); b_{ij} indicates a common border; and ex_i indicates that i is the exporter in the respective trade relation (allowing for asymmetric trade cost). Using $\tau_{jj} = 1$ yields an empirically implementable gravity equation

$$\log\left(\frac{\lambda_{ij}}{\lambda_{jj}}\right) = \underbrace{\log\left(L_i \mathcal{P}_i b_i^{\theta} W_i^{-\theta}\right)}_{\log S_i} - \underbrace{\log\left(L_j \mathcal{P}_j b_j^{\theta} W_j^{-\theta}\right)}_{\log S_j} - \theta\alpha_0 d_{k,ij} - \theta\alpha_1 b_{ij} - \theta\alpha_2 ex_i + \nu_{ij},$$

where the first two terms are captured by a country effect and v_{ij} is an error term. Note that because the country effect is restricted to be the same when a country is the exporter as when it is the importer, one can separately identify the exporter fixed effect, $\theta\alpha_2$, and the country effect, $\log S_i$. Estimating this equation yields estimates $\widehat{\theta\alpha_0}$, $\widehat{\theta\alpha_1}$, $\widehat{\theta\alpha_2}$, and $\widehat{\log S_i}$. The estimated coefficients exactly replicate the results of Simonovska and Waugh (2014) presented in their online appendix. With an assumption for the Pareto parameter θ, we can back out $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2$ and predict trade costs as $\hat{\tau}_{ij} = \exp\left(\hat{\alpha}_0 d_{k,ij} + \hat{\alpha}_1 b_{ij} + \hat{\alpha}_2 ex_i\right)$. Also, by combining the predicted trade cost (scaled by the Pareto parameter) with the exponential of the estimated country effects we can compute the element $L_\nu \mathcal{P}_\nu b_\nu^{\theta} \left(\tau_{\nu j} W_\nu\right)^{-\theta} = \hat{S}_\nu \hat{\tau}_{\nu j}^{-\theta}$ for all ν and j. These elements can be then be used when implementing equation (10) and also when computing model-implied expenditure shares according to equation (11).

following two measures to quantify the relevance of the arbitrage channel:

(i) a *count measure*: the fraction of exporter-destination pairs where the exporter is subject to arbitrage, and

(ii) a *volume measure:* the value of exports among constrained exporter-destination pairs relative to the total global volume of exports.

There are in total $I \times (I-1) = 15,006$ potential bilateral trade relations, of which $10,636$ flows feature positive (non-zero) trade. Among these exporter-destination pairs, according to the baseline numerical analysis shown in Panel a of Table 2, more than 20 percent are price constrained. This estimate varies little with the choice of the Pareto parameter θ. When we set $\theta = 2$ (close to Jung et al. 2015), we find that $2,199$ out of the $10,636$ export flows, or 20.7 percent are price constrained. When we increase this value to $\theta = 8$ (close to Eaton and Kortum 2002), the estimate increases only slightly to 21.9 percent. Changing θ has little quantitative impact, since both sides (9) move in the same direction. When θ increases, both the elasticity of relative prices with respect to per capita income differences, $(2(\theta+1))^{-1}$, and trade costs decrease.

Table 2

Panel a also displays both the number of direct and indirect cases, that lead to constrained exporters. When $\theta = 4$, inequality (9) holds for 32 country pairs (direct arbitrage) and for $3,726$ country triples (indirect arbitrage). The fact that there are few directly and many indirectly binding arbitrage constraints is per se not surprising as the number of country triples (where indirect arbitrage can potentially occur) is an order of magnitude larger than the number of country pairs.[25] Notice also that there are more indirect arbitrage possibilities than constrained exporter-destination pairs. This is because one exporter-destination pair can be subject to indirect arbitrage through several third countries. The numerical analysis finds that the average constrained exporter-destination pair is price constrained through about 1.5 third countries.

In Panel b of Table 2, we calculate the export volume that is subject to a binding arbitrage constraint. Irrespective of the particular value of θ, our numerical analysis suggests that about 45 percent of total world trade is subject to arbitrage. Interestingly, the volume measure indicates that arbitrage is of even higher importance than indicated by the count measure. This suggests that large export flows are more likely to be constrained.

We conclude that the numerical analysis of a multi-country model, in which preferences are such that the arbitrage constraint can in principle become binding, indicates that the arbitrage channel is quantitatively relevant: the typical exporter is price-constrained on 20 percent of its destinations and about 45 percent of world trade occurs among constrained exporter-destination pairs.

In the online appendix, we also show separately for each of the 123 exporters in the data set, the number and volume of constrained destinations. For instance, the US, Japan, and China

[25] Of course, indirect arbitrage is only counted, when it emerges through a third country k that is actually served by producer i. This means that indirect arbitrage threats are only considered for the subset of the $I \times (I-1) \times (I-1) = 1,830,732$ triples, where both the flows from i to j and from i to k is non-zero.

export to all other (122) countries and are constrained in 8, 10, and 13 destinations, respectively. In contrast, the large European exporters, Germany, France, and the UK also export to all other countries and are constrained in 67, 97 and 85 destinations, respectively. The larger number of binding destinations for European exporters is partly due to the high integration (low trade costs) within Europe. This leads to a disproportionate share of indirect arbitrage via cheaper European destinations.

Robustness. In the online appendix, we check the robustness of our results with respect to alternative ways to measure the crucial parameters/varibles. (i) To calculate expenditure shares from observed data, we can use GDP in the denominator rather than manufacturing absorption.[26] (ii) For the trade costs, we can allow fitted trade costs $\hat{\tau}_{ij}$ to fall short of unity (while the baseline sets estimated trade costs below unity to 1). (iii) To quantify $W_i L_i$, we can calculate model-consistent wage rates using the balanced trade condition, $W_i L_i \mathcal{P}_i = \sum_{j=1}^{I} \lambda_{ij} W_j L_j \mathcal{P}_j$. This can be done using empirically observed expenditure shares λ_{ij} or model-implied expenditure shares $\hat{\lambda}_{ij}$ (which are computed using the procedure sketched in footnote 24 above). (iv) We also present the results when using equation (10) instead of (12), which is equivalent to replacing the empirically observed expenditure shares in equation (12) with the model-implied expenditure shares $\hat{\lambda}_{ij}$. (v) Finally, for each of the alternative choices in (i)-(iv) we iterate over the different values of the Pareto parameter $\theta = 2, 4, 6, 8$. In total, this yields $2 \times 2 \times 3 \times 2 \times 4 = 96$ alternative ways to calculate the outcomes of interest (including our baseline specification).

While the particular values for the count and volume measures vary across the various specification, two robust conclusions emerge. First, there is always a substantial fraction of constrained exporters, ranging from 12.5 to 50.8 percent. Second, in all specifications, the share of constrained exports (volume measure) is substantial and larger than the share of constrained markets (count measure). This suggests that arbitrage is more important for larger exporters, ranging from 43.0 to 75.4 percent. We conclude that the results of our baseline numerical analysis are robust.

Discussion. Notice that the above numerical exercise is tentative and clearly needs to be taken with caution. Here we emphasize three points which have to be kept in mind when interpreting the results. First, the baseline numerical analysis is based on observed and strictly positive (non-zero) trade flows. This is clearly in the spirit of our arbitrage model, which emphasizes the limits to price setting for exporters with big price differences and low trade costs. However, our model of arbitrage also emphasizes the relevance of arbitrage threats for export zeros (that let firms abstain from exporting). In the online appendix, we compute relative prices for zero trade flows using equation (10) and country fixed effects from the gravity regression (the approach is sketched out at the end of footnote 24). Relative prices for zero exports can be calculated from equation (10) because the gravity approach of Simonovska and Waugh (2014) yields estimates for τ_{ij} even when i does not export to j. This allows us to test whether inequality (9) holds (or does not hold) for export zeros. We find that the fraction of export zeros that feature binding arbitrage lies between 6.8 percent ($\theta = 2$) and 5.4 percent ($\theta = 8$). One reason for the lower

[26] Notice that GDP is consistent with the model, but inconsistent with the data, since trade are gross flows while GDP is value added.

frequency of arbitrage among exports zeros are the rather high estimated trade costs among these exporter-destination pairs.

This brings us to the second issue, the rather large size of estimated iceberg costs. With $\theta = 8$, about 55 percent of bilateral trade relationships have iceberg trade costs larger than 3, with $\theta = 6, 4, 2$ this number increases to 75, 90, and 95 percent, respectively. High estimated iceberg costs may reflect real-world barriers not adequately captured by the model, such as fixed costs (for establishing a trade relationship, infrastructure, distribution networks, marketing, etc.). These costs, while potentially substantial for the incumbent firm, maybe do not accrue for arbitrageurs (who may, at least partly, free-ride on incumbents). In this case, the trade costs relevant for arbitrageurs could be lower than those captured in the numerical analysis, suggesting that our procedure tends to underestimate the importance of arbitrage. (This could partly explain the lower relevance of arbitrage among export zeros). Clearly, other real-world features not captured in the model may work in the opposite direction. In particular, incumbent firms may lobby for restricting parallel trade (pharmaceutical products are one prominent example), differentiate products (horizontally or vertically), or use other strategies that limit the usefulness/desirability of a product sold in country k for consumers in destination j, thereby reducing arbitrage threats.

A third point concerns the role of firm heterogeneity. While the Stone-Geary specification provides an elegant and simple framework, it implies that relative prices of a country-i exporter do not depend on the firm's productivity, a, see equation (12). This leads to the prediction that, if one exporter from country i is price-constrained in market j, all other exporters from i are also constrained in that market. This artefact of the Stone-Geary specification abstracts from important margins (product characteristics, product-specific trade costs) which may be relevant in practice.

8 Conclusions

This paper studies a model of international trade in which an importer's per capita income determines export zeros and prices of exported goods. This is due to a demand effect: consumers in poor countries have lower willingnesses to pay for differentiated products than consumers in rich countries. As a result, northern firms have a low incentive to export to a southern destination. Our model generates export zeros from non-homothetic preferences and does not rely on firm heterogeneity and/or fixed export-market entry costs. Hence our analysis is complementary to standard heterogeneous-firm approaches which focus on the supply side.

A key insight of our analysis is that export zeros arise from a threat of international arbitrage. Globally active firms cannot simultaneously set low prices in the South and high prices in the North because this triggers arbitrage opportunities. Northern firms have two options to avoid arbitrage: (i) set a sufficiently low price in the North; or (ii) abstain from exporting to the South. These two options involve a trade-off between market size and price: firms that export globally have a large market but need to charge a low price; firms that sell exclusively to northern markets can charge a high price but have a smaller market. The equilibrium of our model is characterized by Linder-effects, a situation where similarity in per capita incomes increases trade intensity between two countries. The model also generates interesting welfare effects. While rich countries always gain from a trade liberalization, poor countries may lose. Lower trade costs tighten the arbitrage constraint and this induces northern firms *not* to export to poor destinations, thus reducing the menu of supplied goods and harming welfare of households in the South.

We argue that the arguments put forth in this paper are potentially empirically relevant. Our model predicts a positive relationship between the export probability of rich exporters and the per capita income of a potential destination. While this is clearly supported by the data, it does not necessarily arise from a threat of arbitrage but could also be due too high export costs.

To shed more specific light on the potential relevance of the arbitrage constraint, we pursue a model-based approach. We explore a model of non-homothetic preferences featuring many countries and heterogenous firms (Simonovska 2015). This model makes predictions for the relative price of a given product between any two locations – under the assumption that firms do not take arbitrage constraints into account. We pursue a simple consistency check and explore whether and to which extent the predicted relative prices violate the arbitrage constraint. We find that arbitrage constraints are indeed violated in many trade relations which suggests that international arbitrage is potentially relevant.

The above consistency check is based on the predictions of a model that does not take arbitrage constraints into account. Solving a full-fletched multi-country model with supply and demand heterogeneities that takes arbitrage constraints into account turns out complex and is beyond the scope of this paper.[27] However, our analysis suggests that accounting for arbitrage in models of monopolistic competition and international trade is an interesting direction for future research. The challenge is to appropriately disentangle supply and demand effects and assess their relative importance. This seems particularly relevant for a better understanding of how rapidly growing per capita incomes in China, India and other emerging markets affect trade patterns and the international division of labor.

[27] If arbitrage is ruled out, a firm optimizes country by country. In contrast, if arbitrage is allowed, the firm optimizes for all countries simultaneously. In fact, it does not only optimize the price in each country taking into account potential arbitrage flows, but also optimizes the set of countries that it supplies. While this problem can be clearly characterised, we were not able to derive general equilibrium expressions that take empirically implementable forms.

References

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

[2] Arkolakis, Costas (2010), "Market Penetration Costs and the New Consumer Margin of Trade," *Journal of Political Economy* 118: 1151-1199.

[3] Arkolakis, Costas, Arnaud Costinot and Andrés Rodriguez-Clare (2012), "New Trade Models, Same Old Gains?," *American Economic Review* 102: 94-130.

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

[5] Auer, Raphael, Thomas Chaney, and Philip Sauré (2014), "Quality Pricing-to-Market," CEPR Discussion Paper 100053.

[6] Baldwin, Richard and James Harrigan (2011), "Zeros, Quality, and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics* 3: 60–88.

[7] Behrens, Kristian and Yasusada Murata (2012a), "Globalization and Individual Gains from Trade," *Journal of Monetary Economics* 59: 703–720.

[8] Behrens, Kristian and Yasusada Murata (2012b), "Trade, Competition, and Efficiency," *Journal of International Economics* 87: 1–17.

[9] Bekkers, Eddy, Francois, Joseph and Miriam Manchin (2012), "Import prices, Income, and Inequality," *European Economic Review* 56: 848-869.

[10] Bergstrand, Jeffrey H. (1990), "The Heckscher-Ohlin Samuelson Model, the Linder Hypothesis and the Determinants of Bilateral Intra-Industry Trade," *The Economic Journal* 100: 1216-1229.

[11] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott (2007), "Firms in International Trade," *Journal of Economic Perspectives* 21: 105-130.

[12] Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott (2012), "The Empirics of Firm Heterogeneity and International Trade," *Annual Review of Economics* 4: 283-313.

[13] Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott (2009), "Importers, Exporters and Multinationals: A Portrait of Firms in the US that Trade Goods," in T. Dunne, J.B. Jensen and M.J. Roberts (eds.), *Producer Dynamics: New Evidence from Micro Data,* University of Chicago Press.

[14] Bernasconi, Claudia (2013), "Similarity of Income Distributions and the Extensive and Intensive Margins of Bilateral Trade Flows," *mimeo,* University of Zurich.

[15] Bernasconi, Claudia and Tobias Wuergler (2012), "Per capita Income and the Quality and Variety of Imports," *mimeo,* University of Zurich.

[16] Bronnenberg, Bart J. (2015), "The Provision of Convenience and Variety by the Market," *RAND Journal of Economics* 46: 480-498.

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

[18] Caron, Justin, Thibault Fally, and James R. Markusen (2014), "International Trade Puzzles: A Solution Linking Production and Preferences," *Quarterly Journal of Economics* 129: 1501-1552.

[19] Chaney, Thomas (2008), "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review* 98: 1707-1721.

[20] Choi, Yo Chul, David Hummels, and Chong Xiang (2006), "Explaining Export Variety and Quality: The Role of the Income Distribution," NBER Working Paper No. 12531.

[21] Dalgin, Muhammed, Devashish Mitra, and Vitor Trindade (2008), "Inequality, Nonhomothetic Preferences, and Trade: A Gravity Approach," *Southern Economic Journal* 74: 747-774.

[22] Desdoigts Alain and Fernando Jaramillo (2009), "Trade, Demand Spillovers, and Industrialization: The Emerging Global Middle Class in Perspective," *Journal of International Economics* 79: 248-258.

[23] Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography, and Trade," *Econometrica* 70: 1741-1779.

[24] Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2011), "An Anatomy of International Trade: Evidence from French Firms," *Econometrica* 79: 1453-1498.

[25] Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo (2013), "International Trade: Linking Micro and Macro," in: Advances in Economics and Econometrics: 1st ed. Vol 2. Cambridge: Cambridge University Press, 329-370.

[26] Fajgelbaum, Pablo, Gene Grossman, and Elhanan Helpman (2011), "Income Distribution, Product Quality and International Trade," *Journal of Political Economy* 119: 721–765.

[27] Falkinger, Josef (1994), "An Engelian Model of Growth and Innovation with Hierarchic Consumer Demand and Unequal Incomes," *Richerche Economiche* 48: 123-139.

[28] Falkinger, Josef (1990), "Innovator-Imitator Trade and the Welfare Effects of Growth," *Journal of the Japanese and International Economies* 4: 157-172.

[29] Feng, Chen Ng (2013), "Arbitrage Across Borders: Why Itunes Gift Cards Sell Above Face Value On Ebay," *Economic Inquiry* 51:1299-1310.

[30] Fieler, Ana (2011), "Non-Homotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," *Econometrica* 79: 1069–1101.

[31] Foellmi, Reto and Josef Zweimüller (2006), "Income Distribution and Demand-Induced Innovations," *Review of Economic Studies* 73: 941-960.

[32] Francois, Joseph F. and Seth Kaplan (1996), "Aggregate Demand Shifts, Income Distribution, and the Linder Hypothesis," *Review of Economics and Statistics* 78: 244-250.

[33] Ganslandt, Mattias and Keith E. Maskus (2004), "Parallel Imports and the Pricing of Pharmaceutical Products: Evidence from the European Union," *Journal of Health Economics* 23: 1035-1037.

[34] Ganslandt, Mattias and Keith E. Maskus (2007), "Intellectual Property Rights, Parallel Imports and Strategic Behavior," Working Paper Series 704, Research Institute of Industrial Economics.

[35] Gaulier, Guillaume and Soledad Zignago (2010), BACI: International Trade Database at the Product-Level. CEPII.

[36] Goldberg, Pinelopi K. (2010), "Intellectual Property Rights Protection in Developing Countries: the Case of Pharmaceuticals," *Journal of the European Economic Association* 8:326-353.

[37] Helpman, Elhanan, Marc Melitz, and Yona Rubinstein (2007), "Trading Partners and Trading Volumes," *Quarterly Journal of Economics* 123: 441- 487.

[38] Hepenstrick, Christian and Alexander Tarasov (2015), "Per capita incomes and the extensive margin of bilateral trade," *Canadian Journal of Economics* 48: 1561-1599.

[39] Heston, Alan, Robert Summers and Bettina Aten (2006), Penn World Table Version 6.2, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

[40] Hsieh, Chang-Tai and Peter J. Klenow (2007), "Relative Prices and Relative Prosperity," *American Economic Review* 97: 562-585.

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

[42] Hummels, David and Volodymyr Lugovskyy (2009), "International Pricing in a Generalized Model of Ideal Variety," *Journal of Money, Credit and Banking* 41: 3–33.

[43] Hunter, Linda C. (1991), "The Contribution of Nonhomothetic Preferences to Trade," *Journal of International Economics* 30: 345-358.

[44] Hunter, Linda C. and James R. Markusen (1988), "Per Capita Income as a Basis for Trade," in Robert Feenstra, ed., *Empirical Methods for International Trade*, MIT Press, Cambridge MA and London.

[45] Jung, Wook Jae, Ina Simonovska, and Weinberger, Ariel (2015), "Exporter Heterogeneity and Price Discrimination: A Quantitative View," NBER Working Paper No. 21408.

[46] Krugman, Paul R. (1980), "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review* 70: 950-959.

[47] Li, Nicholas (2012), "An Engel curve for Variety," *mimeo*, University of Toronto.

[48] Linder, Staffan B. (1961), *An Essay on Trade and Transformation*, Almqvist and Wiksells, Uppsala.

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

[50] Manova, Kalina and Zhiwei Zhang (2012), "Export Prices and Heterogeneous Firm Models," *Quarterly Journal of Economics* 127: 379-436.

[51] Markusen, James R. (2013), "Putting per Capita Income Back into Trade Theory," *Journal of International Economics* 90: 255-265.

[52] Maskus, Keith E. (2000), "Parallel Imports," *The World Economy* 23: 1269-1284.

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

[54] Melitz, Marc J. (2003), "The Impact of Trade on Intra-Industry Allocation and Aggregate Industry Productivity," *Econometrica* 71: 1695-1725.

[55] Melitz, Marc J. and Gianmarco Ottaviano (2008), "Market Size, Trade, and Productivity," *Review of Economic Studies* 75: 295-316.

[56] Mitra, Devashish and Vitor Trindade (2005) "Inequality and Trade," *Canadian Journal of Economics* 38: 1253-1271.

[57] Mrazova, Monika and J. Peter Neary (2013), "Not So Demanding: Preference Structure, Firm Behavior, and Welfare," *mimeo,* Oxford University.

[58] Murphy, Kevin, Andrei Shleifer, and Robert Vishny (1989), "Income Distribution, Market Size, and Industrialization," *Quarterly Journal of Economics* 104: 537-564.

[59] Neary, J. Peter (2009), "International Trade in General Oligopolistic Equilibrium," *mimeo*, University of Oxford.

[60] NERA (1999), *The Economic Consequences of the Choice of Regime in the Area of Trademarks*, Final Report for DG XV of the European Commission.

[61] Roy, Santanu and Kamal Saggi (2012), "Equilibrium Parallel Import Policies and International Market Structure," *Journal of International Economics* 87: 262-276.

[62] Sauré, Philip (2010), "Productivity Growth, Bounded Marginal Utility, and Patterns of Trade," *mimeo*, Swiss National Bank.

[63] Simonovska, Ina (2015), "Income Differences and Prices of Tradables: Insights from an Online Retailer," *Review of Economic Studies* 82: 1612-1656.

[64] Simonovska, Ina and Michael Waugh (2014), "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics* 92: 34-50.

[65] World Health Organization (2001), *Macroeconomics and Health: Investigating in Health for Economic Development*, Report for the Commission of Macroeconomics and Health, WHO, Geneva.

[66] Yeung, Godfrey and Vincent Mok (2013), "Manufacturing and Distribution Strategies, Distribution Channels, and Transactions Costs: The Case of Parallel Imported Automobiles," *Managerial and Decision Economics* 34: 44-58.

A Proof of Proposition 1

Part b). This follows from calculating the derivatives of ϕ with respect to L^P

$$\frac{\partial \phi}{\partial L^P} = \frac{2L^R\mathcal{P}^R\omega\mathcal{P}^P}{(L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P)^2} - \frac{4L^R\mathcal{P}^R\omega L^P\mathcal{P}^P}{(L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P)^3}\omega\mathcal{P}^P = \frac{\phi}{L^P}\left(1 - \frac{2\omega L^P\mathcal{P}^P}{L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P}\right),$$

and with respect to $\mathcal{P}^P$

$$\frac{\partial \phi}{\partial \mathcal{P}^P} = \frac{\phi\left(1 + \frac{\partial \omega}{\partial \mathcal{P}^P}\frac{\mathcal{P}^P}{\omega}\right)}{\mathcal{P}^P}\left(1 - \frac{2\omega L^P\mathcal{P}^P}{L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P}\right),$$

where $\frac{\partial \omega}{\partial \mathcal{P}^P}\frac{\mathcal{P}^P}{\omega} > 0$.

Part c). A given increase in $L^P\mathcal{P}^P$ has a stronger impact on ϕ if it comes from $\mathcal{P}^P$ rather than from L^P if $\partial\phi/\partial\log L^P = \left(\partial\phi/\partial L^P\right)L^P < \partial\phi/\partial\log\mathcal{P}^P = \left(\partial\phi/\partial\mathcal{P}^P\right)\mathcal{P}^P$. This is true since $\frac{\partial \omega}{\partial \mathcal{P}^P}\frac{\mathcal{P}^P}{\omega} > 0$.

Part d). This follows from the derivative of ϕ with respect to τ

$$\frac{\partial \phi}{\partial \tau} = \frac{\phi\frac{\partial \omega}{\partial \tau}}{\omega}\left(1 - \frac{2\omega L^P\mathcal{P}^P}{L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P}\right).$$

B Proof of Proposition 2

Part b) follows from the derivatives of ϕ with respect to L^P and $\mathcal{P}^P$. It is straightforward to see that $\partial\phi/\partial L^P > 0$. To calculate $\partial\phi/\partial\mathcal{P}^P$ we need to take into account that ω depends on $\mathcal{P}^P$

$$\frac{\partial \phi}{\partial \mathcal{P}^P} = \frac{-1}{\mathcal{P}^R\left(\tau + \mathcal{P}^P/\mathcal{P}^R\right)}\phi + \frac{1 + \frac{\partial \omega}{\partial \mathcal{P}^P}\frac{\mathcal{P}^P}{\omega}}{\mathcal{P}^P}\frac{L^R\mathcal{P}^R}{L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P}\phi,$$

hence an increase in $\mathcal{P}^P$ increases trade intensity ϕ when $\mathcal{P}^P$ is small and vice versa.

Part c). We need to show that the volume of trade increases with $\mathcal{P}^P$ less than proportionally. The argument in the text was made without considering that $\mathcal{P}^P$ increases ω. It remains to show that, taking account of the impact of $\mathcal{P}^P$ on ω, an increase in $\mathcal{P}^P$ reduces per capita imports. We have $\text{sign}(\partial p^P N_T^R/\partial\mathcal{P}^P) = \text{sign}(\partial\log(p^P N_T^R)/\partial\mathcal{P}^P) < 0$. Calculating $p^P N_T^R$, taking logs and the derivative with respect to $\mathcal{P}^P$ reveals that $\partial p^P N_T^R/\partial\mathcal{P}^P < 0$ if

$$-\frac{1}{\tau\mathcal{P}^R + \mathcal{P}^P} + \frac{\tau}{aF + \mathcal{P}^R + \tau\mathcal{P}^P} - \frac{1}{aF + \tau\mathcal{P}^R + \mathcal{P}^P} < 0.$$

Multiplying by $aF + \mathcal{P}R + \tau\mathcal{P}^P$ yields

$$-\frac{aF + \mathcal{P}^R + \tau\mathcal{P}^P}{\tau\mathcal{P}^R + \mathcal{P}^P} + \frac{\tau(aF + \mathcal{P}^R + \tau\mathcal{P}^P)}{aF + \mathcal{P}^R + \tau\mathcal{P}^P} - \frac{aF + \mathcal{P}^R + \tau\mathcal{P}^P}{aF + \tau\mathcal{P}^R + \mathcal{P}^P} < 0 \Longleftrightarrow -p^P + \frac{\tau - \omega}{a} < 0,$$

which holds true because $p^P > \tau/a$.

Part d). We note that sign$(\partial\phi/\partial\tau) =$ sign$(\partial \log\phi/\partial\tau)$. Taking logs of the expression for ϕ and the derivative with respect to τ yields

$$\operatorname{sign}(\partial\phi/\partial\tau) = \operatorname{sign}\left(\frac{1}{\tau} - \frac{1}{\tau + \frac{\mathcal{P}^R}{\mathcal{P}^P}} + \frac{\omega'(\tau)}{\omega(\tau)} - \frac{\omega'(\tau)}{\omega(\tau)} \cdot \frac{\frac{\omega(\tau)L^P}{L^R}}{\frac{\omega(\tau)L^P}{L^R} + \frac{\mathcal{P}^R}{\mathcal{P}^P}}\right) > 0,$$

which, using $\omega(\tau)L^P/L^R < \tau$ and $\frac{\omega\prime(\tau)\tau}{\omega(\tau)} > -1$, implies

$$\operatorname{sign}(\partial\phi/\partial\tau) = \operatorname{sign}\left[\left(1 + \frac{\omega'(\tau)\tau}{\omega(\tau)}\right)\left(1 - \frac{\tau}{\tau + \frac{\mathcal{P}^R}{\mathcal{P}^P}}\right)\right] > 0.$$

C Proof of Proposition 3

Part a). In an arbitrage equilibrium we have $p^P = \left(aF + \mathcal{P}^R + \tau\mathcal{P}^P\right) / \left(a\tau\mathcal{P}^R + a\mathcal{P}^P\right)$. Country -$R$ firms export if $p^P \geq \tau/a$ or, equivalently, $\left(aF + \mathcal{P}^R + \tau\mathcal{P}^P\right)\left(\tau\mathcal{P}^R + \mathcal{P}^P\right)^{-1} \geq \tau$. Solving that latter equation for τ yields the trade condition. (Notice that, if the trade condition holds for country -R firms, it also holds for country -P firms, as we have $p^R_T = \tau p^P > p^P$.)

Part b). Under full trade we have $p^P = \omega LP\left(aF + \mathcal{P}^R + \tau\mathcal{P}^P\right)\left(a\mathcal{P}^R L^R + a\omega\mathcal{P}^P L^P\right)^{-1} \geq \tau/a$ or $\left(\omega L^P/L^R\right)\left(aF/\mathcal{P}^R + 1\right) \geq \tau$. But since full trade occurs only when $\omega L^P/L^R \geq 1/\tau$, the trade condition follows.

D Two regions: n rich and m poor countries

In an arbitrage equilibrium, the price of globally traded goods is $p^R_T = \tau p^P$ in the North. Zero profit constraints of globally traded goods are

$$p^P m\mathcal{P}^P + \tau p^P n\mathcal{P}^R = \left(F + \frac{\mathcal{P}^i + \tau\mathcal{P}^{-i}}{a}\right)W^i, \quad i = R, P.$$

where where $\mathcal{P}^{-R} = (n-1)\mathcal{P}^R + m\mathcal{P}^P$ and $\mathcal{P}^{-P} = n\mathcal{P}^R + (m-1)\mathcal{P}^P$. The prices of globally traded goods can be directly calculated $p^P = (aF + \mathcal{P}^R + \tau\mathcal{P}^{-R})/(am\mathcal{P}^P + a\tau n\mathcal{P}^R)$ and $p^R_T = \tau p^P$. The zero profit conditions for goods exclusively traded in the North are

$$p^R_N n\mathcal{P}^R = \left(F + \frac{\mathcal{P}^R + \tau(n-1)\mathcal{P}^R}{a}\right)W^R,$$

and the price of these goods follows immediately $p^R_N = W^R(aF + \mathcal{P}^R + \tau(n-1)\mathcal{P}^R)/(an\mathcal{P}^R)$. From the zero profit conditions of globally traded goods we can calculate relative wages between North and South

$$\omega \equiv \frac{W^P}{W^R} = \frac{aF + \tau\mathcal{P}^{-R} + \mathcal{P}^R}{aF + \mathcal{P}^P + \tau\mathcal{P}^{-P}}.$$

The resource constraints are

$$\begin{aligned} L^P\mathcal{P}^P &= N^P\left(F+\frac{\mathcal{P}^P+\tau\mathcal{P}^{-P}}{a}\right) \qquad \text{for a poor country, and} \\ L^R\mathcal{P}^R &= N_T^R\left(F+\frac{\mathcal{P}^R+\tau\mathcal{P}^{-R}}{a}\right)+N_N^R\left(F+\frac{\mathcal{P}^R+\tau(n-1)\mathcal{P}^R}{a}\right) \text{ for a rich country.} \end{aligned}$$

Each R-country imports all goods produced worldwide, while each P-country imports only a subset of these goods. Hence the aggregate trade balance between the North and the South has to be balanced in equilibrium.[28] This implies

$$\tau N^P\mathcal{P}^R = N_T^R\mathcal{P}^P.$$

From the resource constraints and the trade balance condition we get closed-form solutions for N_P, N_R^T, and N_N^R. This gives welfare of rich and poor households

$$U^R(\tau) = mN^P + nN_T^R + nN_N^R = \frac{aL^P\left(m\mathcal{P}^P+\tau n\mathcal{P}^R\right)}{aF+\mathcal{P}^P+\tau\mathcal{P}^{-P}} + \frac{a\left(L^R-\tau\omega L^P\right)n\mathcal{P}^R}{aF+\mathcal{P}^R+\tau(n-1)\mathcal{P}^R},$$

and

$$U^P(\tau) = mN^P + nN_T^R = \frac{aL^P\left(m\mathcal{P}^P+\tau n\mathcal{P}^R\right)}{aF+\mathcal{P}^P+\tau\mathcal{P}^{-P}}.$$

We see that that $\partial U^R(\tau)/\partial\tau < 0$ and $\partial U^S(\tau)/\partial\tau \leq 0$ when $aF < (m-1)\mathcal{P}^P(1+m\mathcal{P}^P/(n\mathcal{P}^R))$. It also follows that $\partial U^R(\tau)/\partial\tau < \partial U^S(\tau)/\partial\tau$, i.e. a trade liberalization benefits the rich country more.

Finally, let us calculate trade intensity. The value of North-North trade is given by

$$2(n-1)\left(p_N^RN_N^R+p_T^RN_T^R\right)n\mathcal{P}^R = 2(n-1)\left(L^R-\omega\tau L^P\frac{m\mathcal{P}^P}{m\mathcal{P}^P+\tau n\mathcal{P}^R}\right)\mathcal{P}^R.$$

The value of South-South trade is

$$2(m-1)p^PN^Pm\mathcal{P}^P = 2(m-1)\frac{m\mathcal{P}^P}{m\mathcal{P}^P+\tau n\mathcal{P}^R}\omega L^P\mathcal{P}^P$$

and the value of North-South trade is

$$2mp^PN^Pn\mathcal{P}^R = 2m\frac{n\tau\mathcal{P}^R}{m\mathcal{P}^P+\tau n\mathcal{P}^R}\omega L^P\mathcal{P}^P$$

[28] Due to the symmetry of our set-up, the volume of bilateral trade is undetermined. One of the Northern countries could produce predominantly (or exclusively) goods that are consumed only in the North, while the other Northern country produces mainly (or exclusively) goods that are consumed worldwide. In that case, the first Northern country runs a trade surplus with the other Northern country and a trade deficit with both Southern countries taken together. Such trade imbalances cannot occur between the Southern countries, since each Southern country consumes all goods the other Southern country produce, meaning that the South-South trade flows are of the same magnitude in either direction. However, each Southern country may run a surplus with one of the Northern countries that is balanced by a deficit with the other Northern country. Notice further that all bilateral trade flows are equalized in a full trade equilibrium since all households in each country consume all goods that are produced worldwide.

This allows us to calculate trade intensity ϕ, the value of world trade relative to world GDP as

$$\phi = 2\frac{(n-1)\left(p_N^R N_N^R + p_T^R N_T^R\right) n\mathcal{P}^R + (m-1)p^P N^P m\mathcal{P}^P + mp^P N^P n\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}$$

which, using the above formulas, can be expressed as

$$\phi = 2\frac{m\omega L^P\mathcal{P}^P}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}\left(\frac{(m-1)\mathcal{P}^P + n\tau\mathcal{P}^R}{m\mathcal{P}^P + n\tau\mathcal{P}^R}\right) + 2\frac{(n-1)\left(L^R - \tau\omega L^P\frac{m\mathcal{P}^P}{m\mathcal{P}^P + \tau n\mathcal{P}^R}\right)\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}.$$

$$\phi = 2\frac{m\omega L^P\mathcal{P}^P}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}\left(\frac{(m-1)\mathcal{P}^P + n\tau\mathcal{P}^R}{m\mathcal{P}^P + n\tau\mathcal{P}^R}\right) - 2\frac{(n-1)\tau\omega L^P\frac{m\mathcal{P}^P}{m\mathcal{P}^P + \tau n\mathcal{P}^R}\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P} + 2\frac{(n-1)L^R\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}.$$

$$\phi = 2\frac{m\omega L^P\mathcal{P}^P}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}\frac{(m-1)\mathcal{P}^P + \tau\mathcal{P}^R}{m\mathcal{P}^P + n\tau\mathcal{P}^R} + 2\frac{(n-1)L^R\mathcal{P}^R}{nL^R\mathcal{P}^R + m\omega L^P\mathcal{P}^P}.$$

When $m = 1$ and $n = 1$ we get

$$\phi = 2\frac{\omega L^P\mathcal{P}^P}{L^R\mathcal{P}^R + \omega L^P\mathcal{P}^P}\frac{\tau\mathcal{P}^R}{\tau\mathcal{P}^R + \mathcal{P}^P}.$$

Unlike in the arbitrage equilibrium of the two-county case, trade intensity may decrease in τ. This is when a reduction in τ increases South-South and North-North trade more strongly than it reduces North-South trade.

E Equilibrium with general preferences

The arbitrage equilibrium. Here we state the full system of equations that characterize an arbitrage equilibrium with non-traded goods. *Households* choose consumption levels to maximize utility. This implies marginal rates of substitution

$$\frac{v'(c_R^R)}{v'(c_P^R)} = \frac{p_R^R}{p_P^R}, \qquad \frac{v'(c_R^P)}{v'(c_P^P)} = \frac{p_R^P}{p_P^P}, \qquad \frac{v'(c_R^R)}{v'(c_R^N)} = \frac{p_R^R}{p_R^N}.$$

Firms set prices to maximize profits. Firms that sell exclusively on the home market set the unconstrained monopoly price

$$p_R^N = \mu(c_R^N)\frac{1}{a}.$$

Exporting firms set prices to avoid arbitrage

$$p_P^R = \tau p_P^P, \qquad p_R^R = \tau p_R^P.$$

which leads to first-order conditions[29]

$$\tau\frac{p_P^P-\omega/a}{p_P^P}\left(-\frac{v'(c_P^R)}{v''(c_P^R)}\right)+\frac{p_P^P-\omega/a}{p_P^P}\left(-\frac{v'(c_P^P)}{v''(c_P^P)}\right)\frac{\mathcal{P}^P}{\mathcal{P}^R}=\tau c_P^R+c_P^P\frac{\mathcal{P}^P}{\mathcal{P}^R},$$

$$\frac{\tau p_R^P-1/a}{\tau p_R^P}\left(-\frac{v'(c_R^R)}{v''(c_R^R)}\right)+\frac{p_R^P-\tau/a}{p_R^P}\left(-\frac{v'(c_R^P)}{v''(c_R^P)}\right)\frac{\mathcal{P}^P}{\mathcal{P}^R}=c_R^R+\tau c_R^P\frac{\mathcal{P}^P}{\mathcal{P}^R}.$$

The *resource constraints* are

$$L^P=N_P\left(F+\mathcal{P}^R\tau c_P^R/a+\mathcal{P}^P c_P^P/a\right),$$

$$L^R=N_R^T\left(F+\mathcal{P}^R c_R^R/a+\mathcal{P}^P\tau c_R^P/a\right)+N_R^N\left(F+\mathcal{P}^R c_R^N/a\right),$$

the *trade balance* condition is

$$p_R^P N_R^T\mathcal{P}^P c_R^P=p_P^R N_P\mathcal{P}^R c_P^R,$$

and the *zero-profit conditions* are

$$\mathcal{P}^P c_P^P\left(p_P^P-\omega/a\right)+\mathcal{P}^R c_P^R\left(p_P^R-\tau\omega/a\right)=\omega F.$$

$$\mathcal{P}^R c_R^R\left(p_R^R-1/a\right)+\mathcal{P}^P c_R^P\left(p_R^P-\tau/a\right)=F,$$

$$\mathcal{P}^R c_R^N\left(p_R^N-1/a\right)=F,$$

In sum, the arbitrage equilibrium has 14 equations in 14 unknowns: quantities (c_P^P, c_R^P, c_R^R, c_P^R, c_R^N), prices (p_P^P, p_R^P, p_R^R, p_P^R, p_R^N), firm measures (N_P, N_R^T, N_R^N), and the relative wage ω.

Full trade equilibria. As mentioned in the main text, a binding arbitrage constraint is a necessary though not sufficient condition for an arbitrage equilibrium with non-traded goods since consumers can now respond also along the intensive margin. There are three types of full trade equilibria: (i) both P- and R-firms are price-constrained; (ii) P-firms are price-constrained while R-firms set the monopoly price; and (iii) firms in both countries set the monopoly price.[30]

ad (i). When both firms are price-constrained but all goods are traded, all equations are identical except that $N_R^N=c_R^N=0$ and p_R^N do not exist. The system reduces to 11 equations.

ad (ii). When P-firms are price constrained but R-firms are not, we have $p_R^R=\mu(c_R^R)/a$ and $p_R^P=\mu(c_R^P)\tau/a$ while p_P^R and p_P^P are still determined as above.

ad (iii). When firms in both countries are unconstrained, also P-firms set the monopoly price $p_P^P=\mu(c_P^P)\omega/a$ and $p_P^R=\mu(c_P^R)\omega\tau/a$.

[29] These conditions derive from maximizing the profit functions for country-P and country-R producers, i.e. $\mathcal{P}^P c_P^P\left(p_P^P-\omega/a\right)+\mathcal{P}^R c_P^R\left(p_P^R-\tau\omega/a\right)$ and $\mathcal{P}^R c_R^R\left(p_R^R-1/a\right)+\mathcal{P}^P c_R^P\left(p_R^P-\tau/a\right)$, subject to the above arbitrage constraints. Moreover, we use the fact that households' demand functions derive from $v(c)=\lambda p$ which implies $\partial c/\partial p=(1/p)v'(c)/v''(c)$.

[30] Notice that the (unconstrained) price gap of country-P firms between market P and market R is higher than the corresponding price gap for country-R firms. This is because country-P firms have low (high) costs and low (high) demand on the home (foreign) market. This is different from the situation of country-R firms. They have high (low) costs and low (high) demand on the foreign (home) market. This implies that country-P firms get price-constrained first, and an equilibrium, where country-R firms are price-constrained - but country-P firms are not - cannot exist.

F Proof of Proposition 7

We determine the autarky equilibrium and ask under which conditions an entrepreneur has incentives to sell his products abroad. Setting $W = 1$, optimal monopolistic pricing implies $p = \mu(c)/a$. With free entry, profits $\mathcal{P}^R(p_a^R - 1/a)c_a^R$ must equal set up costs F

$$aF/\mathcal{P}^R = \left(\mu(c_a^R) - 1\right) c_a^R$$

The equilibrium is symmetric for all firms, hence the resource constraint reads

$$L^R = N_a^R \left(F + \mathcal{P}^R c_a^R/a\right)$$

Solving for c_a^R and N_a^R, we see that c_a^R does not depend on L^R. Hence when the two countries differ only in L^i but have equal populations, intensive consumption levels under autarky are identical between the two countries, $c_a^R = c_a^P$. Selling one marginal unit abroad at price $v'(0)/\lambda_a^P$, allows the purchase of $v'(0)/(\lambda_a^P p_a^P)$ foreign goods. Since $\lambda_a^P = v'(c_a^P)/p_a^P$ and $c_a^R = c_a^P$ this is equal to $v'(0)/v'(c_a^R) > 1$. Reselling this (new) product at home, yields a price $v'(0)p_a^R/v'(c_a^R)$ minus trade costs. Hence, this strategy is profitable if $\left[v'(0)p_a^R/v'(c_a^R)\right] \cdot \left[v'(0)/v'(c_a^R)\right] > \tau^2$. Expressing p_a^R in terms of c_a^R, we get the condition of the Proposition.

Figure 1: Demand function

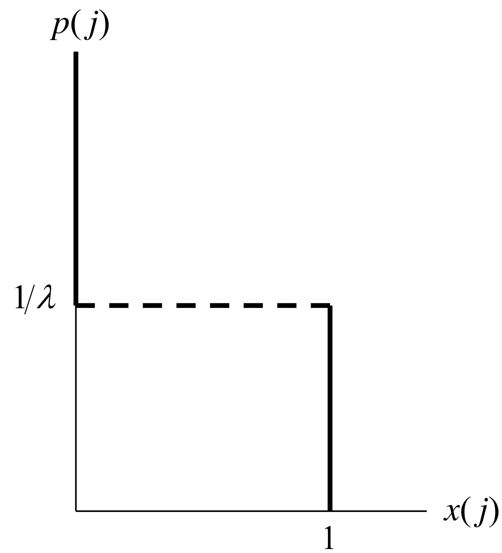


Figure 2: Full trade and arbitrage equilibrium

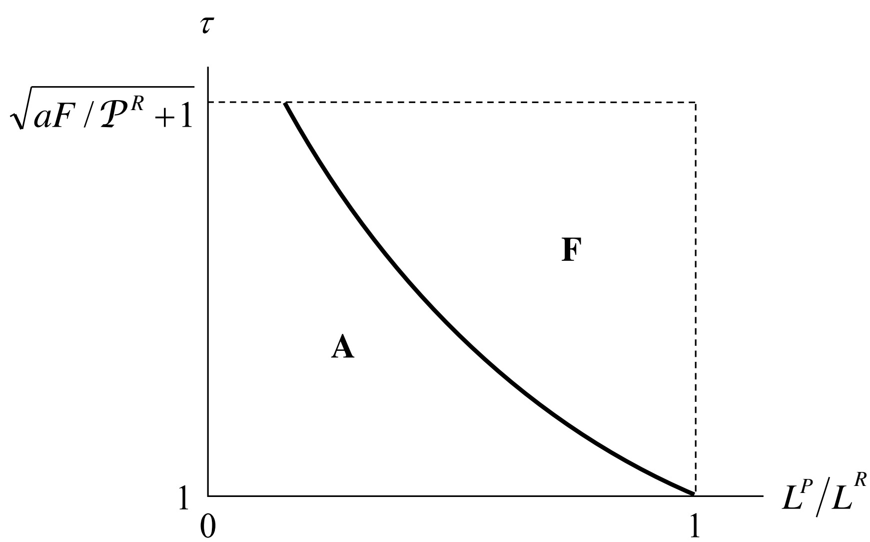
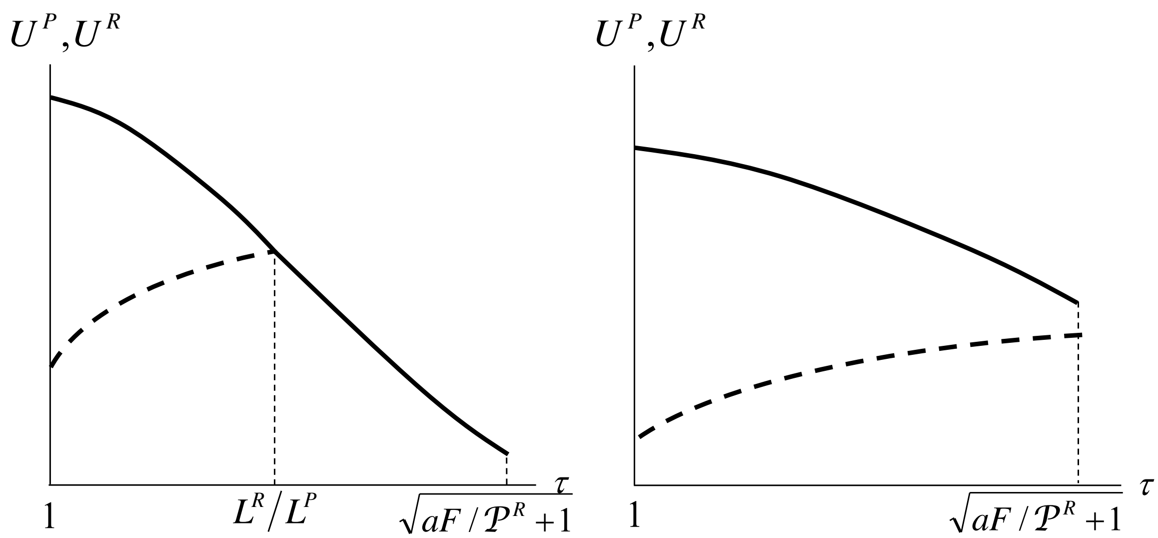


Figure 3: Welfare and trade costs



(a) (b)

Table 1: Extensive margin of U.S. exports, 2007

	(1) All HS6	(2) All HS6	(3) Pop>1m HS6	(4) All HS4	(5) All HS2
Mean dependent variable	0.391	0.391	0.399	0.576	0.731
Log importer GDP	0.064***	0.070***	0.089***	0.076***	0.061***
	(0.011)	(0.010)	(0.010)	(0.011)	(0.009)
Log importer GDP per capita	0.085***				
	(0.014)				
Per capita income USD 385–999		-0.265***	-0.205***	-0.270***	-0.178***
		(0.064)	(0.066)	(0.067)	(0.065)
Per capita income USD 1,000–1,999		-0.262***	-0.226***	-0.249***	-0.145**
		(0.067)	(0.068)	(0.071)	(0.060)
Per capita income USD 2,000–3,999		-0.237***	-0.233***	-0.232***	-0.163***
		(0.065)	(0.063)	(0.066)	(0.049)
Per capita income USD 4,000–7,999		-0.116*	-0.146**	-0.076	-0.035
		(0.069)	(0.066)	(0.068)	(0.047)
Per capita income USD 8,000–15,999		-0.128*	-0.149**	-0.104	-0.089*
		(0.068)	(0.064)	(0.067)	(0.051)
Per capita income USD 16,000–31,999		-0.034	-0.030	-0.017	-0.008
		(0.064)	(0.061)	(0.060)	(0.041)
Per capita income USD 32,000–		ref	ref	ref	ref
Trade cost indicators	Yes	Yes	Yes	Yes	Yes
HS fixed effects	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.424	0.423	0.442	0.431	0.417
N	169,020	169,020	147,736	42,255	9,045

Notes: Estimates based on a linear probability model, *, **, *** denotes statistical significance at the 10%, 5%, 1% level, respectively. Standard errors are clustered on importer level. Year is 2007. Omitted category of income per capita groups is above USD 32,000. Trade cost indicators include log of distance between exporter's and importer's capital, dummy for a common border, dummy for importer being an island, dummy for importer being landlocked, dummy for importer and exporter ever having had colonial ties, dummy for currency union between importer and exporter, dummy for importer and exporter sharing a common legal system, dummy for religious similarity, dummy for importer and exporter having a free trade agreement, and dummy for importer and exporter sharing a common language.

Table 2: Number of constrained exporters and constrained export volume

	$\theta = 2$	$\theta = 4$	$\theta = 6$	$\theta = 8$
Panel a: Number of constrained exporters				
Number positive trade flows	10636	10636	10636	10636
Number constrained flows	2199	2239	2295	2324
Fraction constrained	0.207	0.211	0.216	0.219
Direct arbitrage	28	32	37	37
Indirect arbitrage	3466	3726	3903	4022
Panel b: Constrained export volume				
World trade volume (m)	6829931	6829931	6829931	6829931
Constrained volume (m)	3113948	3077943	3106796	3136772
Share constrained volume	0.456	0.451	0.455	0.459
Direct constrained volume (m)	1315237	1375793	1489080	1489080
Share direct constrained volume	0.193	0.201	0.218	0.218

Notes: The data is described in the Data Appendix of Simonovska and Waugh (2014). Bilateral trade-flow data are from UN Comtrade for the year 2004 (for a sample of 123 countries, restricted to manufacturing trade flows only). Manufacturing production data for 2004 is from UNIDO where it exists and imputed for all other countries.