Geography, Competition, and Optimal Multilateral Trade Policy*

Antonella Nocco, Università del Salento
Gianmarco I.P. Ottaviano, Bocconi University, Baffi-CAREFIN, CEP, CEPR
Matteo Salto, European Commission

March 2019

Abstract

How should multilateral trade policy be designed in a world in which countries differ in terms of market access and technology, and firms with market power differ in terms of productivity? We answer this question in a model of monopolistic competition in which variable markups increasing in firm size are a key source of misallocation across firms and countries. We use 'disadvantaged' to refer to countries with smaller market size, worse state of technology (in terms of higher innovation and production costs), and worse geography (in terms of more remoteness from other countries). We show that, in a global welfare perspective, optimal multilateral trade policy should: promote the sales of low cost firms to all countries, but especially to disadvantaged ones; trim the sales of high cost firms to all countries, but especially to disadvantaged ones; reduce firm entry in all countries, but especially in disadvantaged ones. This would not only restore efficiency but also reduce welfare inequality between advantaged and disadvantaged countries if their differences in market size, state of technology and geography are large enough.

Keywords: International trade policy, monopolistic competition, firm heterogeneity, pricing to market, multilateralism.

J.E.L. Classification: D4, D6, F1, L0, L1.

---

*Nocco: University of Salento, Department of Management, Economics, Mathematics and Statistics, Ecotekne, via Monteroni, 73100, Lecce, ITALY, antonella.nocco@unisalento.it. Ottaviano: Bocconi University, Department of Economics, Via Roentgen 1, 20136, Milan, ITALY, gianmarco.ottaviano@unibocconi.it. Salto: European Commission, Rue de la Loi 200, B- 1049, BELGIUM, matteo.salto@ec.europa.eu. For useful comments and discussions we are grateful to Anders Akerman, Arnaud Costinot, Kyle Bagwell, Pamela Bombarda, Eric Bond, Paola Conconi, Carsten Eckel, Hartmut Egger, Peter Egger, Matthew Grant, Andreas Haufler, Elhanan Helpman, Mario Larch, Marc Melitz, Yasusada Murata, Antonio Navas, Peter Neary, Emanuel Ornelas, Ralph Ossa, Aurelien Poissonnier, Veronica Rappoport, Regis Renault, Thomas Sampson, Davide Suverato, Scott Taylor, Hylke Vandenbussche and Ben Zissimos as well as seminar participants at various institutions, conferences and workshops. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 789049-MIMAT-ERC-2017-ADG). The views expressed here are those of the authors and do not represent in any manner the EU Commission.

1 Introduction

How should multilateral trade policy be designed in a world in which countries differ in terms of market access and technology, and firms with market power differ in terms of productivity? Should trade policy differ across countries? Should worse performing (national) firms be protected from better performing (foreign) rivals? Should national product diversity be shielded against the potentially disruptive effects of cheaper imported goods? The answers to these questions crucially depend on market structure, demand characteristics and technological constraints. In particular, in the 'canonical' models of monopolistic competition with CES demand, iceberg trade friction, sunk entry costs, fixed production costs and constant marginal costs that are (inverse) Pareto distributed across firms, the free market equilibrium is efficient and there is, therefore, no room for welfare improving policy intervention: free trade is the best multilateral trade policy. More precisely, efficiency of the free market outcome is granted in models in which there is only the monopolistically competitive sector. When there is also another perfectly competitive ('outside good') sector, the relative size of the monopolistically competitive sector is inefficiently small due to markup pricing. Yet, as the markup is the same and constant across the monopolistic competitors, firms' sizes are efficient in both absolute and relative terms. This implies that the inefficiently small size of the monopolistically competitive sector materializes entirely through an inefficiently small number of firms (see, e.g., Melitz and Redding, 2014 and 2015).

The aim of the present paper is to show how all this ceases to hold once the CES assumption is removed, leading to new implications in terms of multilateral trade policy aimed at maximizing the joint welfare of all trade partners.[1] In doing so, we focus on a specific deviation from CES known as 'Marshall's Second Law of Demand' (MSLD), according to which demand becomes more inelastic with consumption (Mrazova and Neary, 2013). As we discuss below, this assumption is both theoretically and empirically appealing. We show that under MSLD the free trade allocation of resources fails to be efficient in terms of product range, product selection and product mix with the extent of misallocation varying across countries depending on market size, state of technology and geography. For conciseness, we use 'advantaged' ('disadvantaged') to refer to countries with larger (smaller) domestic market size, better (worse) state of technology in terms of lower (higher) innovation and production costs, and better (worse) geography in terms of closer proximity to other countries. Our findings can then be summarized as follows. First, from a welfare point of view, too many products are sold to advantaged countries and too few are sold to disadvantaged ones ('inefficient product range'). Second, conditional on range, relatively too many high cost products are sold to any country ('inefficient product selection'). This inefficiency is, however, more severe for disadvantaged countries. Third, conditional on range and selection, the quantities of high cost products sold to any country are too large and those of low cost products are too small ('inefficient product mix'). Also this inefficiency is more severe for disadvantaged countries. As a result, the free market provides an inefficiently high degree of welfare inequality between advantaged and disadvantaged countries if their

[1] Our aim is normative rather than positive: to inform rather than explain actual trade policy choices. This normative focus also sets us apart from recent works that compare the 'gains from trade' and the effects of counterfactual trade policies in quantitative trade models with firm heterogeneity in alternative demand and supply side setups (see, e.g., Arkolakis, Costinot and Rodriguez-Clare, 2012; Costinot and Rodriguez-Clare, 2014; Arkolakis et al, 2015; Jung, Simonovska and Weinberger, 2015; Bertoletti, Etro and Simonovska, 2016).

differences in market size, state of technology and geography are large enough. There is, therefore, room for welfare improving multilateral policy intervention that: increases sales of low cost firms to all countries but especially to disadvantaged ones; decreases sales of high cost firms to all countries but especially to disadvantaged ones; reduces firm entry in all countries but especially in disadvantaged ones.

In our analytical framework market inefficiency stems from four types of externalities (Nocco, Ottaviano and Salto, 2014; Behrens et al, 2016; Dhingra and Morrow, 2019). First, firms neglect their impact on product variety. Due to 'love of variety', the product range enters utility as a direct utility-enhancing argument on top of the quantities consumed. This acts as a driver towards too few varieties. Second, by keeping price above marginal cost, firms leave too much room for entry. This acts as a driver towards too many varieties. Third, firms neglect the negative impact of their entry on rivals' profits. This also acts as a driver towards too many varieties. These three externalities are the traditional ones already highlighted in earlier models of monopolistic competition (Spence, 1976; Dixit and Stiglitz, 1977) and operate also when firms are not heterogeneous. Their net effect on product range is generally ambiguous depending on the cross-elasticities of demand. A special case arises with CES demand: the opposite externalities exactly offset each other so that the free market outcome is efficient (without the 'outside good'). The introduction of firm heterogeneity does not alter this property as CES demand implies the same constant markup for all firms so that also the product mix is efficient (Melitz and Redding, 2015). The presence of a fourth type of externality is tied to MSLD as, with MSLD but not with CES, firm heterogeneity becomes an additional driver of inefficiency. The fact that demand becomes more inelastic with consumption is reflected in larger markups for firms with lower marginal cost. As a result, these firms do not fully trasmit their cost advantage to prices. By softening competition, this generates a positive externality in favor of firms with higher marginal cost. The externality works at both the intensive and the extensive margins. At the intensive margin, higher marginal cost firms are inefficiently large relative to lower marginal costs firms. At the extensive margin, by keeping price above marginal cost more than their higher marginal cost rivals, lower marginal cost firms leave inefficiently larger room for entry.

Analyzing the MSLD case is important in many respects both theoretically and empirically. As pointed out by Mrazova and Neary (2013), Marshall (1920) argues this case represents the normal behavior of demand, an opinion shared also by Spence (1976), Dixit and Stiglitz (1977) and Krugman (1979). Subsequent studies have vindicated this view. MSLD plays a crucial role for some of the key traditional (non-CES) implications of trade models with monopolistic competition, including: 'pro-competitive' effects, through which trade liberalization reduces firms' markups (Krugman, 1979); 'pricing to market', through which firms set c.i.f. prices in each market they sell to rather than simply setting a single f.o.b. price in the market they sell from (Krugman, 1987); 'dumping', through which firms accept a lower profit margin per unit sold in foreign than in home markets (Brander and Krugman, 1983); and incomplete 'pass through', through which differences in firms' production costs traslate in less than proportionate price differences (Dornbusch, 1987).[2] MSLD also underpins some newer implications of those models in the presence of firm heterogeneity. In particular, better

[2]While some of these properties have been initially discussed in the case of oligopoly, later studies have shown that they also hold in the case of monopolistic competition under MSLD. See, e.g., Ottaviano, Tabuchi and Thisse (2002) and the discussion in Mayer, Melitz and Ottaviano (2016).

performing firms (those with lower marginal cost and larger market shares) are predicted to set higher markups (Melitz and Ottaviano, 2008; Mayer, Melitz and Ottaviano, 2014 and 2016).[3] Last but not least, several of the implications of MSLD are supported by mounting empirical evidence on firm performance based on price data (Berman, Martin and Mayer, 2012; De Loecker and Goldberg, 2014; De Loecker, Goldberg, Khandelwal and Pavcnik, 2016) as well as on revenue data (Mayer, Melitz and Ottaviano, 2014 and 2016).[4]

We perform our normative analysis of the free market outcome within the general equilibrium framework proposed by Melitz and Ottaviano (2008). This model with an 'outside good' and quasi-linear quadratic utility exhibits several features useful for our purposes.[5] As it exhibits linear demand, it satisfies MSLD and thus features pro-competitive effects, pricing to market, dumping, and incomplete pass-through as well as higher markups for better performing firms.[6] As it is analytically solvable with asymmetries in market size, technology and accessibility for an arbitrary number of countries, it allows for transparent comparative statics in a multi-country setup. As the marginal utility of income is constant and utility is transferable, it allows for a consistent efficiency analysis based on a straightforward definition of global welfare for an economy with heterogeneous countries as the sum of all individuals' indirect utilities. While one may note that the absence of income effects gives our investigation some partial equilibrium flavor, the framework still shares its focus on social surplus with a large body of trade policy analyses that abstract from distributive issues (Bagwell and Staiger, 2016).[7]

Our analysis contributes to three main literatures. The first is the literature on optimal trade policy under imperfect competition (Grossman and Helpman, 1989; Grossman, 1992).[8] This literature usually does not feature more than two countries. Its findings with homogeneous firms are summarized by Felbermayr, Jung and Larch (2013): tariffs can correct for the distortion due to markup pricing (Flam and Helpman, 1987); tariffs can induce welfare-enhancing additional entry (Venables, 1987); tariffs can improve the terms of trade (Gros, 1987).[9] With CES demand and monopolistic competition

[3] Mayer, Melitz and Ottaviano (2016) show that MSLD entails an increasing relationship between output and markup, and thus the *level* of pass-through. They also show that, when a stronger restriction (which they call MSLD') holds, there is an additional connection between changes in output and changes in markups, and thus *differences* in pass-through: the pass-through rate is lower for better than for worse performing firms. Specifically, MSLD' requires marginal revenue to become more inelastic with consumption and this implies MLSD.

[4] Due to its far-reaching implications, MSLD has also attracted renewed interest in the contemporary debate on the qualitative and quantitative effects of trade liberalization, though often disguised under different headings: "log-concavity in log-prices" (Arkolakis et al, 2015); "sub-convexity" (Neary and Mrazova, 2013); "increasing relative love of variety" (Zhelobodko et al, 2012); "decreasing elasticity of substitution" (Bertoletti and Epifani, 2014); "Adjustable pass-through" (Fabinger and Weyl, 2014). See Mayer, Melitz and Ottaviano (2016) for a discussion of mappings between these concepts.

[5] Irrespective of quasi-linearity, as pointed out by Ossa (2011), models with a freely traded 'outside good' generate a perfectly elastic labor supply curve and thus isolate the effects of trade policies on firm location. Models with no 'outside good' generate, instead, a perfectly inelastic labor supply curve and hence isolate the effects of trade policy on the terms of trade.

[6] As linear demand also satisfies MSLD', it also features lower pass-through rate for better performing firms (Mayer, Melitz and Ottaviano, 2016).

[7] The assumption of quasi-linear utility, under which income transfers are utility transfers, is also frequently made in political economy models of trade policy (Grossman and Helpman, 2001).

[8] For a recent overview of optimal trade policy under perfect competition, see the introductory discussion in Costinot, Donaldson, Vogel and Werning (2015).

[9] Flam and Helpman (1987), Gros (1987) and Venables (1987) all rely on variants of the CES two-country model by Krugman (1980). In a multi-country set-up involving the six major players in recent GATT/WTO negotiations (Brazil, China, the EU, India, Japan, and the US), Ossa (2011) shows that a calibrated version of that model predicts noncooperative tariffs of the same order of magnitude as the tariffs observed during the tariff war following Smoot-Hawley.

à la Krugman (1980) the incentives for a non-cooperative trade policy arise from the desire to eliminate monopolistic distortions and to improve domestic terms of trade (Campolmi, Fadinger and Forlati, 2014). More recently, firm heterogeneity has been introduced in models of monopolistic competition. When demand is CES as in Melitz (2003) and tariffs are either set unilaterally by a small open economy (Demidova and Rodriguez-Claire, 2009; Haaland and Venables, 2016) or by a large open economy facing another large open economy (Felbermayr, Jung and Larch, 2013), trade barriers have beneficial effects on the protectionist country. By raising the country's wage, an import tariff produces an improvement in its terms-of-trade. When product variety is inefficiently poor, a tariff on imports, or a subsidy to domestic sales, increases the number of varieties offered by the market also correcting the mark-up distortion. Costinot, Rodriguez-Clare and Werning (2016) analyze the effects of firm-specific unilateral intervention by a large open economy and show that its welfare is maximized by optimal import taxes that discriminate against the most profitable foreign exporters and optimal export taxes that are uniform across domestic exporters. Campolmi, Fadinger and Forlati (2018) show that with two large countries the Nash equilibrium when both domestic and trade policies are available is characterized by first-best-level labor subsidies that achieve production efficiency, and inefficient import subsidies and export taxes that aim at improving domestic terms of trade. Non-CES demand à la Melitz and Ottaviano (2008) is considered by Bagwell and Lee (2015), who show that in the case of two symmetric countries there is an incentive for a country to introduce a small unilateral import tariff. They also identify the conditions under which two symmetric countries have unilateral incentives to introduce beggar-thy-neighbor export subsidies. Moreover, in the case of symmetric trade policies, they find that global free trade is generally inefficient. Within the same framework but without the outside good, Demidova (2017) shows that a unilateral reduction in a 'wasteful' import tariff (i.e. a frictional tariff that does not generate any tax revenue) increases the protectionist country's welfare both in the case of two large economies and in the case of a small open economy. Differently, when the import tariff is 'non-wasteful' (i.e. it generates tax revenues as in the other foregoing studies), in both cases unilateral trade liberalization reduces the country's welfare. Our contribution to this literature is the analysis of multilateral trade policy with heterogeneous firms for an arbitrary number of asymmetric countries that cooperatively maximize global welfare when demand is non-CES. For completeness we also investigate the incentives for a country to deviate from multilateral cooperation: they are consistent with the tradeoffs already highlighted by the existing literature on unilateral trade policy.[10]

The second literature we contribute to studies optimal product variety in models of monopolistic competition without firm heterogeneity (Spence, 1976, and Dixit and Stiglitz, 1977) and with firm heterogeneity (Melitz and Redding, 2014; Nocco, Ottaviano and Salto, 2014 and 2017; Dhingra and Morrow, 2019). This literature focuses on a closed (or 'perfectly integrated') economy or on open economies with symmetric countries.[11] We extend this literature by investigating the role of country asymmetries in terms of market size, geographical barriers to trade and state of technology.

[10] Ossa (2014) studies noncooperative and cooperative trade policy in a calibrated multi-country multi-industry general equilibrium model with inter-industry trade in the Ricardian tradition as well as intra-industry trade in the wake of Krugman (1980) and thus with CES demand. The model is richer than ours but cannot be solved analytically for an arbitrary number of asymmetric countries. A recent overview of the economics literature on trade agreements, under perfect and imperfect competition, can be found in Bagwell, Bown and Staiger (2016).

[11] See, e.g., Nocco, Ottaviano and Salto (2017) for a discussion of the main developments in this literature, and for a study of the impact of different degrees of firm heterogeneity on the extent of market inefficiencies.

Third and last, we contribute to the growing literature on 'misallocation' in the wake of Hsieh and Klenow (2009), who use a closed economy CES model of monopolistic competition to show how output and capital distortions give rise to 'wedges' in marginal revenue products between firms, and how the welfare losses from those distortions can be quantified through the measurement of the corresponding wedges.[12] As discussed by De Loecker and Goldberg (2014), trade policies fall nicely into this framework. While output tariffs and subsidies distort output markets due to their effects on competition, input tariffs and subsidies directly distort capital and intermediates markets. With CES demand and no 'outside good', reducing these distortions through trade liberalization necessarily improves welfare through a more efficient allocation of resources across firms. Our contribution to this literature is to show that, when demand is non-CES, free trade is not efficient and trade liberalization does not necessarily improve welfare. In particular, with asymmetric countries there are situations in which trade liberalization may actually increase the misallocation of resources towards less productive firms and countries.[13] This can happen because the global welfare maximizing trade policy is not free trade.[14]

The rest of the paper is organized as follows. Section 2 introduces the model and Section 3 derives the free market outcome. Section 4 characterizes the efficient outcome. Section 5 compares the two outcomes, discussing the inefficiency of the former in terms of product range, product selection and product mix. It also analyzes the sources of the inefficiencies and the implications of the two outcomes for international inequality. Section 6 describes the first best multilateral trade policies that can be implemented to attain efficiency at the market equilibrium when policy tools are unconstrained. It also discusses second and third best policies when there are constraints on available tools. Section 7 looks at unilateral deviation from the efficient outcome. Section 8 concludes.

2 Multi-Country Economy

We follow Melitz and Ottaviano (2008, Appendix) and consider a global economy consisting of M countries, indexed by $l = 1, ..., M$. Country l is populated by L_l consumers, each endowed with one unit of labor, inelastically supplied in a perfectly competitive labor market. Preferences of consumers in l are defined over a 'traditional' homogeneous good 0 and a continuum of varieties of a horizontally differentiated 'modern' good. We use Ω_l to denote this continuum and index varieties by $i \in \Omega_l$. All

[12] See Hopenhayn (2014) for a recent appraisal of the broader literature on the role that firm heterogeneity and the allocation of resources across firms play in determining aggregate productivity.

[13] There are very few contributions that explicitly look at misallocation through the lenses of the markup heterogeneity implied by non-CES demand. Epifani and Gancia (2011) focus on heterogeneity across industries, and thus on between-industry misallocation, relying on a reduced-form 'markup function' that encompasses different underlying model structures (including Bernard et al., 2003; Atkeson and Burstein 2008; Melitz and Ottaviano, 2008). They show that in a symmetric multi-country open economy model trade can affect (and in some cases reduce) welfare by changing the cross-sectoral dispersion of market power. See also Holmes, Hsu and Lee (2014) and Edmond, Midrigan and Xu (2015). Peters (2018) proposes a closed-economy dynamic model of firm growth that generates an endogenous within-industry stationary (Pareto) distribution of markups. The model is applied to the study of the effects that barriers to entry and product market expansion have on aggregate productivity through changes in the markup distribution.

[14] This is a classical second-best welfare result in the tradition of Bhagwati and Ramaswami (1963). See Srinivasan (1996) for an appraisal of the ensuing literature.

consumers share the same utility function

$$U_l = q_{0l}^{\varepsilon} + \alpha \int_{i\in\Omega_l} q_l^{\varepsilon}(i)\, di - \frac{1}{2}\gamma \int_{i\in\Omega_l} \left(q_l^{\varepsilon}(i)\right)^2 di - \frac{1}{2}\eta \left(\int_{i\in\Omega_l} q_l^{\varepsilon}(i)\, di\right)^2, \tag{1}$$

where q_{0l}^{ε} and $q_l^{\varepsilon}(i)$ refer to the individual consumption levels of the traditional good and variety i of the modern good respectively. Parameters α, η and γ are all positive: γ is a measure of 'love for variety'; α and η capture the intensity of preferences for the modern good relative to the traditional one. All consumers have an initial endowment $\overline{q}_{0l}^{\varepsilon}$ of the traditional good, which is assumed to be large enough for its consumption to be strictly positive.

Labor is the only input. It is employed in the production of the traditional good under perfect competition and constant returns to scale with unit labor requirement equal to one. It is also employed in the production of the modern good under monopolistic competition with a one-to-one relation between firms and varieties. In country l the supply of a variety of this good faces two type of costs: a sunk 'innovation' requirement of $f_l > 0$ units of labor to design the blueprint of the variety; and a 'production' requirement of c units of labor per unit of output. The latter is drawn from a continuous distribution with cumulative density function

$$G_l(c) = \left(\frac{c}{c_{M,l}}\right)^k, \; c \in [0, c_{M,l}]. \tag{2}$$

This corresponds to the usual case in which marginal productivity $1/c$ is Pareto distributed with shape parameter $k \geq 1$ over the support $[1/c_{M,l}, \infty)$. For $k = 1$ the distribution is uniform on its support $[0, c_{M,l}]$. As k rises, density is skewed towards the upper bound of the support. In the limit, as k goes to infinity, the distribution becomes degenerate at $c_{M,l}$. Together with f_l, $c_{M,l}$ defines the 'state of technology' in country l. Comparing the cumulative density functions $G_l(c)$ and $G_h(c)$ of two countries l and h with the same shape parameter k but different supports $c_{M,l} < c_{M,h}$ shows that the former first-order stochastically dominates the latter as it cumulates more density on the lower part of the overlapping segment of the supports. Accordingly, given that the traditional good's unit labor requirement equals one, $c_{M,h}/c_{M,l} > 1$ can be interpreted as a measure of country l's 'comparative advantage' in the production of the modern good with respect to country h.

Exchange of varieties of the modern good is hampered by 'iceberg frictions' for international shipments: $\tau_{lh} > 1$ units have to be shipped from country l for one unit to arrive in country $h \neq l$. These frictions are determined by geographical and technological factors. Crucially, they are not trade policy variables. National shipments do not face, instead, any friction ($\tau_{ll} = 1$).

3 Market Outcome

In the equilibrium consumers maximize utility subject to their budget constraints, firms maximize profits subject to their technological constraints (for both production and trade), and all markets clear. Choosing the traditional good as numeraire, perfect competition in its market together with free trade implies that both its price and the wage of workers equal one in all countries.[15] Quasi-

[15] Unit wage allows us to interpret the parameters of input requirements as costs, which we will do henceforth.

linearity of utility (1) then implies that workers decide how much to spend of their unit wage on the varieties of the modern good, leaving whatever residual budget to the consumption of the traditional good. The first order condition for utility maximization gives individual inverse demand for variety i

$$p_l(i) = \alpha - \gamma q_l^{\varepsilon}(i) - \eta Q_l^{\varepsilon}, \tag{3}$$

for $q_l^{\varepsilon}(i) \geq 0$, with $p_l(i)$ denoting the price of variety i in country l and $Q_l^{\varepsilon} = \int_{i\in\Omega^l} q_l^{\varepsilon}(i)\,di$ denoting total individual demand of the differentiated varieties. Aggregation of (3) across consumers leads to aggregate demand of variety i in country l

$$q_l(i) \equiv q_l^{\varepsilon}(i) L_l = \frac{\alpha L_l}{\eta N_l + \gamma} - \frac{L_l}{\gamma} p_l(i) + \frac{\eta N_l}{\eta N_l + \gamma} \frac{L_l}{\gamma} \bar{p}_l \;\forall i \in \Omega_{*,l}, \tag{4}$$

where the set $\Omega_{*,l}$ is the largest subset of Ω_l such that demand in l is positive for variety i, N_l is the measure ('number') of varieties in $\Omega_{*,l}$ (given by the sum of domestic and imported varieties), and $\bar{p}_l = (1/N_l)\int_{i\in\Omega_{*,l}} p_l(i)\,di$ is their average price. Variety i belongs to this set when

$$p_l(i) \leq \frac{1}{\eta N_l + \gamma}(\gamma\alpha + \eta N_l \bar{p}_l) \equiv p_l^{\max}, \tag{5}$$

where $p_l^{\max} \leq \alpha$ represents the price at which demand for a variety in l is driven to zero.

3.1 Product Mix

Turning to modern firms, pricing to market arises from price discrimination on a geographical basis with firms setting c.i.f. prices in each market they sell to.[16] We use $q_{lh}(c)$ to denote the quantity sold in country h by a firm producing in country l at marginal cost c and $p_{lh}(c)$ to denote the corresponding c.i.f. price ($h = l$ refers to domestic transactions). Maximization of profits earned from sales to h are achieved for $q_{lh}(c)$ equal to

$$q_{lh}^m(c) = \begin{cases} \frac{L_h}{2\gamma}\tau_{lh}\left(c_{lh}^m - c\right) & \text{if } c \leq c_{lh}^m \equiv \frac{p_h^{\max}}{\tau_{lh}} = \frac{1}{\tau_{lh}}\left(\alpha - \eta\frac{Q_h^m}{L_h}\right) \\ 0 & \text{if } c > c_{lh}^m \end{cases} \tag{6}$$

where 'm' labels the free market equilibrium values of the variables and

$$Q_h^m \equiv \sum_{l=1}^{M}\left(N_{E,l}\int_0^{c_{lh}^m} q_{lh}^m(c)dG_l(c)\right)$$

is the total quantity of modern good sold in country h with $N_{E,l}$ denoting the number of entrants in country l. Expression (6) defines a cutoff rule as only entrants in country l with low enough marginal cost ($c \leq c_{lh}^m$) sell their variety to country h. For them, the profit-maximizing c.i.f. price is

[16] International price discrimination had been the traditional definition of 'dumping' before 1974, when the definition was extended to include sales below cost (see, e.g., Kolev and Prusa, 2002, for a discussion). Nowadays the legal definition of 'dumping' has little to do with any economic notion of dumping (Blonigen and Prusa, 2003).

$p^m_{lh}(c) = \tau_{lh}(c^m_{lh} + c)/2$, which implies markup $\mu^m_{lh}(c) = \tau_{lh}(c^m_{lh} - c)/2$ and maximized profit

$$\pi_{lh}(c) = \frac{L_h}{4\gamma}(\tau_{lh})^2 (c^m_{lh} - c)^2. \tag{7}$$

Equation (6) implies $\tau_{lh}c^m_{lh} = \tau_{hh}c^m_{hh} = p^{\max}_h$ and thus the cutoffs for domestic and foreign sellers in h are linked by the relation

$$c^m_{lh} = \frac{c^m_{hh}}{\tau_{lh}} \tag{8}$$

for $l, h = 1, ..., M$ given $\tau_{hh} = 1$.

These results show that, conditional on the country they produce in (l) and the country they sell to (h), firms with lower marginal cost c sell more output $q^m_{lh}(c)$ than higher cost firms as their price $p^m_{lh}(c)$ is lower despite higher markup $\mu^m_{lh}(c)$ – and the more so the lower the marginal cost cutoff c^m_{lh}. This leads to higher profit $\pi_{lh}(c)$. Considering two firms with different marginal costs c and c' with $c < c'$, their relative price $p^m_{lh}(c)/p^m_{lh}(c') = (c^m_{lh} + c)/(c^m_{lh} + c')$ is larger than their relative marginal cost c/c', the more so the larger the cutoff c^m_{lh}.[17]. Given $\tau_{lh} > 1$, $c^m_{lh} < c^m_{hh}$ implies that marginal sellers to h have lower marginal cost if they are foreign than if they are domestic.

3.2 Product Selection

Due to free entry, in equilibrium expected profit for an entrant in country l is exactly offset by the sunk cost f_l. Given (2), (7), (8) and $\tau_{hh} = 1$, this 'free entry condition' can be stated as a function of the cutoffs for domestic sellers only:

$$\sum_{h=1}^{M} L_h \rho_{lh} (c^m_{hh})^{k+2} = 2\gamma (k+2)(k+1) f_l (c_{M,l})^k. \tag{9}$$

where $\rho_{lh} \equiv (\tau_{lh})^{-k}$ is an inverse measure of trade frictions from l to h ('trade freeness') ranging between 0 for prohibitive international frictions and 1 for frictionless national trade ($\tau_{ll} = 1$). Together with analogous conditions for the other $M - 1$ countries, (9) yields a system of M equations that can be solved for the M equilibrium domestic cutoffs

$$c^m_{ll} = \left\{ \frac{2\gamma(k+1)(k+2)}{L_l} \frac{\sum_{h=1}^{M}\left[f_h (c_{M,h})^k |C_{hl}|\right]}{|P|} \right\}^{\frac{1}{k+2}} \tag{10}$$

for $l = 1, ..., M$, where $|P|$ is the determinant of the trade freeness matrix $P = [\rho_{hl}]_{(h=1,...,M; l=1,...,M)}$, $|C_{hl}|$ is the cofactor of its ρ_{hl} element and $f_h (c_{M,h})^k$ inversely measures the quality of the 'state of

[17] Following Hsieh and Klenow (2009), we can define firm 'TFP revenue' for sales from l to h in the free market outcome as $TFPR^m_{lh}(c) \equiv p^m_{lh}(c)/c = (\tau_{lh} + c^m_{hh}/c)/2$. Then, comparing two firms with marginal costs c and c' with $c < c'$, we have $TFPR^m_{lh}(c) > TFPR^m_{lh}(c')$. As $TFPR^m_{lh}(c)$ and $TFPR^m_{lh}(c')$ are not equalized, there is 'misallocation' of resources. In particular, $TFPR^m_{lh}(c) > TFPR^m_{lh}(c')$ implies that the low (high) cost firm is allocated too little (much) labor.

technology' in country h in terms of both innovation (f_h) and production ($c_{M,h}$).[18] The expression of the domestic cutoff (10) can be decomposed into two multiplicative components

$$c_{ll}^{ma} \equiv \left[\frac{2\gamma (k+1)(k+2) f_l (c_{M,l})^k}{L_l} \right]^{\frac{1}{k+2}} \tag{11}$$

and

$$\mathbb{C}_{ll} \equiv \left\{ \frac{1}{|P|} \sum_{h=1}^{M} \left[\frac{f_h}{f_l} \left(\frac{c_{M,h}}{c_{M,l}} \right)^k |C_{hl}| \right] \right\}^{\frac{1}{k+2}} \tag{12}$$

such $c_{ll}^m = c_{ll}^{ma} \mathbb{C}_{ll}$. The first component c_{ll}^{ma} corresponds to the cutoff that would materialize if trade frictions were prohibitive. This 'autarkic cutoff' is determined by own market size L_l and state of technology $f_l (c_{M,l})^k$, with lower c_{ll}^{ma} associated with larger market size (larger L_l), lower innovation cost (smaller f_l) and stronger comparative advantage (smaller $c_{M,h}$) in the modern good. The second component $\mathbb{C}_{ll}$ is, instead, trade-related and combines market access, ease of innovation and comparative advantage. Better accessibility to foreign markets (i.e. higher centrality in the trade network defined by P), lower innovation cost (larger f_h/f_l), and higher probability of low cost draws in production (larger $(c_{M,h}/c_{M,l})^k$) lead to a lower $\mathbb{C}_{ll}$. The second component equals 1 when country l is autarkic; it is positive but smaller than 1 otherwise as long as the trade freeness matrix satisfies the triangle inequality and there is some production of the modern good in all countries.[19] This implies $c_{ll}^m < c_{ll}^{ma}$: product selection is stronger with trade than in autarky.

To summarize, firm selection in the modern sector is stronger (c_{ll}^m is smaller) in countries that have larger market size (larger L_l) as well as better state of technology in terms of both innovation (smaller f_l) and production (smaller $c_{M,l}$), and that have better access to trade partners (as dictated by P). These are all factors that foster firm entry and make competition tougher. Henceforth, for conciseness, we will refer to such countries as 'advantaged' and to the others as 'disadvantaged'. For a given value of the cutoff, advantaged countries are more attractive to entrants as these anticipate higher profits in case of survival. The cutoff is, therefore, lower in such countries to reduce the probability of survival and make firms indifferent about which country to enter by equalizing their expected profits before entering to zero everywhere.

3.3 Product Range

To complete the characterization of the free market outcome, we need to pin down the equilibrium numbers of entrants ($N_{E,l}$), producers ($N_{P,l}$) and sellers (N_l) in each country. For the number of sellers (which determines the 'product range'), we can use $c_{lh}^m \equiv p_h^{\max}/\tau_{lh}$ from (6), $p_l^{\max} =$

[18] We focus on situations in which the modern sector is active in all countries. This is indeed the case as long as $\sum_{h=1}^{M} \left[f_h (c_{M,h})^k |C_{hl}| \right] / |P| > 0$ holds for all $l = 1, ..., M$. Given $\tau_{lh} > \tau_{ll} = 1$, that condition also implies $c_{ll}^m > c_{lh}^m$ so that marginal exporters have lower marginal cost than marginal producers.

[19] The trade freeness matrix P satifies the triangle inequality as long as $\rho_{hl} \geq \rho_{hn}\rho_{nl}$ for all $h, l, n = 0, ..., M$. When it does, modern production takes place everywhere ($\mathbb{C}_{ll} > 0$ for all $l = 0, ..., M$) as long as the cross-country variation of f_l, $c_{M,l}$ and ρ_{hl} is not too pronounced. The argument for $\mathbb{C}_{ll} < 1$ is detailed in Appendix A.

$(\gamma\alpha + \eta N_l \bar{p}_l) / (\eta N_l + \gamma)$ from (5), and $\bar{p}_l^m = \{(2k+1) / [2(k+1)]\} c_{ll}^m$ due to (2) to obtain

$$N_l^m = \frac{2\gamma (k+1)}{\eta} \frac{\alpha - c_{ll}^m}{c_{ll}^m} \tag{13}$$

for $l = 1, ..., M$. The key result here is that product variety is richer in countries with lower c_{ll}^m. Given (10), these are the advantaged countries. Hence, as in these countries consumers face not only lower prices (as already discussed) but also richer product variety, welfare is higher as captured by indirect utility

$$U_l^m = 1 + \overline{q}_{0l}^{\varepsilon} + \frac{1}{2\eta} (\alpha - c_{ll}^m) \left(\alpha - \frac{k+1}{k+2} c_{ll}^m \right). \tag{14}$$

Finally, to find the equilibrium number of entrants, it is useful to note that the number of sellers from country h to country l equals $N_{hl} = N_{E,h} G_h(c_{hl})$ (i.e. the share of entrants with marginal cost lower than the cutoff) so that, given (2), (8) and $N_l = \sum_{h=1}^M N_{hl}$, the equilibrium number of sellers in l also evaluates to $N_l^m = \sum_{h=1}^M \rho_{hl} N_{E,h} (c_{ll}^m / c_{M,h})^k$. This can be combined with (13) to obtain, for $l = 1, ..., M$, a system of M linear equations that can be solved for the equilibrium number of entrants

$$N_{E,l}^m = \frac{2\gamma (k+1) (c_{M,l})^k \sum_{h=1}^{M} \left[(\alpha - c_{hh}^m) (c_{hh}^m)^{-(k+1)} |C_{lh}| \right]}{\eta |P|}. \tag{15}$$

The corresponding equilibrium number of producers is then given by $N_{P,l}^m = N_{E,l}^m (c_{ll}^m / c_{M,l})^k$.

4 Globally Efficient Outcome

To evaluate the efficiency of the free market outcome we consider the problem faced by a benevolent social planner who maximizes *global* welfare taking as given, for each country l, the endowment of labor L_l, the endowment of the traditional good $\overline{q}_{0l} = \overline{q}_{0l}^{\varepsilon} L_l$, trade frictions and the production functions of the two goods. In the case of the modern good, this means that the planner takes as given the mechanism determining each variety's unit labor requirement c as a random draw from the distribution $G_l(c)$ after f_l units of labor have been allocated to the design of that variety. As the quasi-linearity of (1) implies transferable utility, global welfare W can be expressed as the sum of consumers' utilities across all countries: $W = \sum_{h=1}^M U_l L_l$. For each country l the planner's choice variables are then: the quantity of the traditional good ($q_{0l} = q_{0l}^{\varepsilon} L^l$); the number of varieties designed ($N_{E,l}$); and the quantity of each variety earmarked by country of production l and country of consumption h ($q_{lh}(c) = q_{lh}^{\varepsilon} L_h$).

Accordingly, the planner's program can be summarized as

$$\max_{\{q_{0l}, N_{E,l}, q_{lh}(c)\}|_{l=1}^M} W = \sum_{l=1}^{M} U_l L_l, \tag{16}$$

subject to the resource constraint

$$q_{0l} + f_l N_{E,l} + N_{E,l} \sum_{h=1}^{M} \left[\int_0^{c_{M,l}} \tau_{lh} c q_{lh}(c) dG_l(c) \right] = \overline{q}_{0l} + L_l \tag{17}$$

for $l = 1, ..., M$ and with $\tau_{ll} = 1$. In (17) the third term on the left hand side is overall labor employment in the production of the modern good in country l, taking the distribution of c and iceberg frictions τ_{lh} into account. Analogously, given (1), in $U_l L_l$ we can use $\sum_{h=1}^{M} N_{E,h} \int_0^{c_{M,h}} q_{hl}(c) dG_h(c)$ and $\sum_{h=1}^{M} N_{E,h} \int_0^{c_{M,h}} \left[q_{hl}(c) L^l\right]^2 dG_h(c)$ to substitute for $\int_{i\in\Omega_l} q_l(i)\, di$ and $\int_{i\in\Omega_l} (q_l(i))^2\, di$ respectively.

4.1 Product Mix

The first order condition with respect to $q_{lh}(c)$ gives

$$q_{lh}^o(c) = \begin{cases} \frac{L_h}{\gamma} \tau_{lh} \left(c_{lh}^o - c\right) & c \le c_{lh}^o \text{ with } c_{lh}^o \equiv \frac{1}{\tau_{lh}} \left(\alpha - \eta \frac{Q_h^o}{L_h}\right) \\ 0 & c > c_{lh}^o \end{cases} \tag{18}$$

where 'o' labels the efficient values of the variables and $Q_h^o \equiv \sum_{l=1}^{M} \left(N_{E,l} \int_0^{c_{lh}^o} q_{lh}^o(c) dG_l(c)\right)$. Hence, just like the free market, also the planner follows a cutoff rule: only varieties with low enough unit labor requirement ($c \le c_{lh}^o$) are produced in country l for consumption in country h. Analogously, conditional on the countries of production and consumption, varieties with lower unit labor requirement c are supplied in larger amounts than varieties with lower c, the more so the lower the cutoff in the country of consumption. Efficient quantity $q_{lh}^o(c)$ would clear the market in a decentralized scenario only if each producer in l priced the quantity sold in h at marginal delivered cost $p_{lh}^o(c) = \tau_{lh} c$.[20] Result (18) also implies that the relation of the optimal cutoff for marginal varieties consumed and produced in h with that for varieties consumed in h but produced in l is regulated by

$$c_{lh}^o = \frac{c_{hh}^o}{\tau_{lh}} \tag{19}$$

for $l, h = 1, ..., M$. The relation is, therefore, the same as for the free market equilibrium (see (8)), even though the cutoffs are different as we now show.

4.2 Product Selection

The cutoffs of the planner are derived from the first order conditions of the planner's problem with respect to $N_{E,l}$. These require

$$\sum_{h=1}^{M} L_h \rho_{lh} \left(c_{hh}^o\right)^{k+2} = \gamma (k+2)(k+1) f_l \left(c_{M,l}\right)^k \tag{20}$$

[20] This can be seen by substituting $p_{lh}(c) = p_{lh}^o(c) = \tau_{lh} c$ in the inverse demand function $p_{lh}(c) = \alpha - \gamma q_{lh}(c)/L_h - \eta Q_h^o/L_h$ and using the definition $\tau_{lh} c_{lh}^o = \alpha - \eta Q_h^o/L_h$ to obtain $q_{lh}(c) = q_{lh}^o(c) = \tau_{lh}\left(c_{lh}^o - c\right) L_h/\gamma$. Given $p_{lh}^o(c) = \tau_{lh} c$, we can define $TFPR_{lh}^o(c) \equiv p_{lh}^o(c)/c = \tau_{lh}$. Hence, $TFPR_{lh}^o(c)$ is the same for all firms selling from l to h and there is thus no 'misallocation' in the sense of Hsieh and Klenow (2009). See Footnote 17 on 'misallocation' at the free market outcome.

for $l = 1, ..., M$, generating a system of M equations that can be solved for the M optimal domestic cutoffs

$$c_{ll}^{o} = \left\{ \frac{\gamma (k+1)(k+2)}{L_l} \frac{\sum_{h=1}^{M} \left[f_h (c_{M,h})^k |C_{hl}| \right]}{|P|} \right\}^{\frac{1}{k+2}} \tag{21}$$

for $l = 1, ..., M$.[21] Accordingly, also the expression of the optimal cutoff can be decomposed into the product of an autarkic component

$$c_{ll}^{oa} \equiv \left[\frac{\gamma (k+1)(k+2) f_l (c_{M,l})^k}{L_l} \right]^{\frac{1}{k+2}} \tag{22}$$

and the trade-related component $\mathbb{C}_{ll}$ defined in (12) with $c_{ll}^{o} = c_{ll}^{oa} \mathbb{C}_{ll}$. As in the free market outcome, efficient product selection is stricter (c_{ll}^{o} is smaller) in advantaged countries. As the trade-related component is identical to the free market one and is therefore smaller than 1 with trade, we have $c_{ll}^{o} < c_{ll}^{oa}$: also in the efficient outcome product selection is stronger with trade than in autarky.

4.3 Product Range

As for the number of varieties sold (and thus the 'product range'), we can use the definition of c_{lh}^{o} from (18), the relation between c_{lh}^{o} and c_{hh}^{o} from (19), and the expression for Q_h^o obtained using (18) together with (2) in its definition to get

$$N_l^o = \frac{\gamma (k+1)}{\eta} \frac{\alpha - c_{ll}^{o}}{c_{ll}^{o}} \tag{23}$$

for $l = 1, ..., M$. As the efficient number of varieties consumed in l also evaluates to $N_l^o = \sum_{h=1}^{M} \rho_{hl} N_{E,h}^{o} \left(c_{ll}^{o} / c_{M,h} \right)^k$, this expression can be combined with (23) to obtain, for $l = 1, ..., M$, a system of M linear equations that can be solved for the efficient number of varieties designed

$$N_{E,l}^{o} = \frac{\gamma (k+1) (c_{M,l})^k \sum_{h=1}^{M} \left[(\alpha - c_{hh}^{o}) (c_{hh}^{o})^{-(k+1)} |C_{lh}| \right]}{\eta |P|} \tag{24}$$

with $l = 1, ..., M$. The corresponding efficient number of varieties produced in l is then given by $N_{P,l}^{o} = N_{E,l}^{o} \left(c_{ll}^{o} / c_{M,l} \right)^k$. Analogously to the free market outcome, product variety is richer (N_l^o is larger) in countries with lower c_{ll}^{o}. Given (21), these are again the advantaged countries. Since indirect utility can be written as

$$U_l^o = 1 + \overline{q}_{0l}^{\varepsilon} + \frac{1}{2\eta} \left(\alpha - c_{ll}^{o} \right)^2, \tag{25}$$

[21] For the market outcome we focused on situations in which the modern sector is active in all countries, which requires $\sum_{h=1}^{M} \left[f_h (c_{M,h})^k |C_{hl}| \right] / |P| > 0$ to hold for all $l = 1, ..., M$. This condition implies that the same happens also in the efficient outcome. Given $\tau_{lh} > \tau_{hh} = 1$, it also implies $c_{ll}^{o} > c_{lh}^{o}$ so that marginal exporters have lower marginal cost than marginal producers.

such countries enjoy higher welfare.

5 Market Failure

We are now ready to compare the free market and efficient outcomes in terms of product selection, product mix and product range.

5.1 Product Selection

Product selection is determined by cutoff (10) in the free market case and by cutoff (21) in the efficient case. Accordingly, the gap between the two evaluates to

$$c_{ll}^m - c_{ll}^o = \left(2^{\frac{1}{k+2}} - 1\right) c_{ll}^o. \tag{26}$$

As this shows that c_{ll}^m is larger than c_{ll}^o, the planner is more selective than the free market: the share of varieties designed but not produced by the planner is larger than the share of entrants that do not produce in equilibrium. In particular, varieties with $c \in (c_{ll}^o, c_{ll}^m]$ are supplied by the free market but should not be produced from an efficiency viewpoint. The length of the interval of inefficiency $c_{ll}^m - c_{ll}^o$ decreases as c_{ll}^o falls. It is, therefore, shorter in advantaged countries. Hence, these countries in the free market outcome not only enjoy higher welfare, but are also less inefficient in terms of product selection. Vice versa, disadvantaged countries face not only lower welfare but also more inefficient product selection at the free market outcome.

Inefficient selection also materializes in terms of exports. Using (10), (21), (8) and (19) allows us to write the gap between the export cutoffs as

$$c_{lh}^m - c_{lh}^o = \left(2^{\frac{1}{k+2}} - 1\right) c_{lh}^o = \left(2^{\frac{1}{k+2}} - 1\right) \frac{c_{hh}^o}{\tau_{lh}},$$

which implies $c_{lh}^o < c_{lh}^m$. Hence, the share of varieties produced but not exported is larger for the planner than for the free market. Specifically, varieties with $c \in (c_{lh}^o, c_{lh}^m]$ are exported in the free market outcome but they should not be exported on efficiency grounds. Conditional on bilateral friction τ_{lh}, this inefficiency is more severe in export countries with larger cutoff c_{hh}^o. These are disadvantaged countries that not only produce an inefficiently larger share of varieties they design, but also import an inefficiently large share of varieties produced elsewhere. On the other hand, conditional on the destination country's cutoff c_{hh}^o, the inefficiency is more pronounced for shipments to destinations associated with lower τ_{lh} and thus easier to reach.

5.2 Product Mix

Turning to output, comparing the free market outcome from (6) with the efficient outcome from (18) gives the quantity gap

$$q_{lh}^m(c) - q_{lh}^o(c) = \frac{L_h}{2\gamma} \tau_{lh} \left[c - \left(2 - 2^{\frac{1}{k+2}}\right) c_{lh}^o\right],$$

which implies that $q^m_{lh}(c)$ is larger than $q^o_{lh}(c)$ if and only if c is larger than $c^{m/o}_{lh} \equiv \left(2 - 2^{1/(k+2)}\right) c^o_{lh}$. As this threshold falls in the efficient selection interval $[0, c^o_{lh}]$, the free market provides inefficiently small quantities of varieties with unit labor requirement below $c^{m/o}_{lh}$, and inefficiently large quantities of varieties with unit input requirement above $c^{m/o}_{lh}$. All the rest given, inefficiency is larger the further away a variety's unit input requirement c is from the threshold. Clearly, $q^m_{lh}(c)$ is larger than $q^o_{lh}(c)$ also for $c \in (c^o_{lh}, c^m_{lh}]$. In this case, as discussed above, the free market quantity $q^m_{lh}(c)$ is positive whereas the planner's quantity $q^o_{lh}(c)$ is zero.

Given (19), in the free market outcome the shares of inefficiently under-supplied and over-supplied varieties ($c^{m/o}_{lh}/c^m_{lh} = 2^{\frac{k+1}{k+2}} - 1$ and $1 - c^{m/o}_{lh}/c^m_{lh}$ respectively) do not depend on country characteristics and are thus the same for advantaged and disadvantaged destinations. What differs across destinations is, instead, the extent of the inefficiency in the distribution of quantities across varieties supplied. To see this, consider two varieties that are supplied both in the free market and in the planner's outcomes with unit labor requirements c and c' in $[0, c^o_{lh}]$ and such that $c < c'$ holds. Using (6) and (18) to compare their relative quantities in the two outcomes yields

$$\frac{q^m_{lh}(c)}{q^m_{lh}(c')} - \frac{q^o_{lh}(c)}{q^o_{lh}(c')} = -\frac{(c^m_{lh} - c^o_{lh})(c' - c)}{(c^m_{lh} - c')(c^o_{lh} - c')}. \tag{27}$$

As this is negative and holds for any c and c' in $[0, c^o_{lh}]$, the distribution of quantities supplied by the planner is always more skewed towards varieties with low unit labor requirement than the distribution at the free market outcome. However, using (8), (19) as well as $c^m_{hh} = 2^{\frac{1}{k+2}} c^o_{hh}$ to substitute for c^m_{lh} in the right hand side of (27) and differentiating the resulting expression with respect to c^o_{hh} shows that the inefficiency gap in the quantity ratio $[q^m_{lh}(c)/q^m_{lh}(c') - q^o_{lh}(c)/q^o_{lh}(c')]$ is an increasing function of the cutoff c^o_{hh}. This implies that the inefficiency in the distribution of quantities is more severe in disadvantaged than in advantaged countries. Disadvantaged countries, therefore, not only produce inefficiently larger shares of the varieties they design and import inefficiently larger shares of varieties produced elsewhere, but they also feature a more inefficient product mix biased towards varieties with higher unit labor requirements.

5.3 Product Range

The range of products consumed in country l are given by (13) and (23) for the free market and the planner respectively. Given $c^m_{ll} = 2^{\frac{1}{k+2}} c^o_{ll}$, the resulting product range gap evaluates to

$$N^m_l - N^o_l = \frac{\gamma (k+1)}{\eta} \left[\left(2^{\frac{k+1}{k+2}} - 1\right) \frac{\alpha}{c^o_{ll}} - 1 \right].$$

As this is generally different from zero, in all countries the free market offers an inefficient product range. Moreover, N^m_l is smaller (larger) than N^o_l for $c^o_{ll} > (<) \left(2^{\frac{k+1}{k+2}} - 1\right) \alpha$. The free market product range is, therefore, inefficiently narrow (wide) for large (small) c^o_{ll}. Accordingly, the free market makes too few varieties available in disadvantaged countries, and too many varieties available in advantaged countries. This does not imply however that consumption of the modern good is inefficiently low in the former and inefficiently high in the latter. Using (2), (6), (18), (19) and (21) to compute country l's average quantities $\overline{q}^m_h$ and $\overline{q}^o_l$ consumed in the free market and efficient outcomes respectively, the gap in per-capita consumption of the modern good turns out to be

$$\frac{N_l^m \overline{q}_l^m}{L_l} - \frac{N_l^o \overline{q}_l^o}{L_l} = -\frac{2^{\frac{1}{k+2}} - 1}{\eta} c_{ll}^o.$$

As this is negative, in the free market outcome all countries consume an inefficiently low per-capita amount of the modern good, the more so the larger c_{ll}^o. Hence, the under-consumption is more severe in disadvantaged countries. The same holds for the average per-capita consumption of modern varieties as the corresponding gap evaluates to

$$\frac{\overline{q}_l^m}{L_l} - \frac{\overline{q}_l^o}{L_l} = -\frac{1 - 2^{-\frac{k+1}{k+2}}}{\gamma(k+1)} c_{ll}^o. \tag{28}$$

The fact that all individuals in all countries consume inefficiently little of the modern good implies that the global supply of that good must be inefficiently small. Given $\overline{q}_l^m = \overline{q}_{hl}^m$ and $\overline{q}_l^o = \overline{q}_{hl}^o$ for all h, (28) also implies that in the free market outcome trade per capita is inefficiently low at the intensive margin, especially for disadvantaged countries.[22] Given $N_{hl} = N_{E,h}\rho_{hl} \left(c_{ll}/c_{M,h}\right)^k$, (15) and (24) imply that also the extensive margin of trade is inefficiently low towards these countries. It is, however, inefficiently high towards advantaged countries.

5.4 International Inequality

There are finally implications in terms of welfare inequality between advantaged and disadvantaged countries. Given (14), (25) and $c_{ll}^m = 2^{\frac{1}{k+2}} c_{ll}^o$, all countries enjoy higher welfare in the efficient outcome than in the free market one ($U_l^o > U_l^m$). However, the welfare gap $U_l^o - U_l^m$ is a concave quadratic function of c_{ll}^o (or equivalently c_{ll}^m) since its derivative with respect to c_{ll}^o is positive (negative) for small (large) c_{ll}^o.[23] This implies that even though in the free market outcome both disadvantaged and advantaged countries suffer from inefficiently low welfare levels, the welfare gap is smaller (larger) in advantaged (disadvantaged) countries if their free market cutoffs are different enough, that is, if their differences in market size, state of technology and geography are large enough. In this case, the free market provides an inefficiently high degree of welfare inequality between the two types of countries.

To see this, denote the efficient levels of welfare in advantaged and disadvantaged countries by U_A^o and U_D^o respectively so that $U_A^o - U_D^o > 0$. Analogously, denote the free market levels of welfare by U_A^m and U_D^m so that $U_A^m - U_D^m > 0$. The corresponding cutoffs are related by $c_A^m = 2^{\frac{1}{k+2}} c_A^o < c_D^m = 2^{\frac{1}{k+2}} c_D^o$. Given that the welfare gap $U_l^o - U_l^m$ is a concave quadratic function of the efficient cutoff, for any country l we can find a threshold unit labor requirement $c_{ll}^{om} \leq c_{ll}^o$ such that the corresponding welfare gap equals the one attained at c_{ll}^o. Accordingly, when the efficient cutoffs for the two types of countries are such that $c_A^o < c_{ll}^{om} \leq c_D^o$, the welfare gap is larger (smaller) in disadvantaged (advantaged) countries: $U_D^o - U_D^m > U_A^o - U_A^m$. This is more likely to happen when

[22] The market average quantity produced in l and consumed in h is $\overline{q}_{lh}^m = \int_0^{c_{lh}^m} q_{lh}^m(c) dG_l(c)/G_l(c_{lh}^m)$. Under the distributional assumption (2), it evaluates to $\overline{q}_{lh}^m = L_h \tau_{lh} c_{lh}^m / [2\gamma(k+1)] = L_h c_{hh}^m / [2\gamma(k+1)]$ with the second equality granted by (8). As this does not depend on the country of production l, we then have that it is also the average quantity consumed in h : $\overline{q}_h^m = \overline{q}_{lh}^m$. Analogously, for the efficient outcome we get: $\overline{q}_h^o = \overline{q}_{lh}^o = L_h c_{hh}^o / [\gamma(k+1)]$.

[23] The same applies with respect to c_{ll}^m as $c_{ll}^m = 2^{\frac{1}{k+2}} c_{ll}^o$.

c_A^o and c_D^o are farther apart, that is, for larger differences in market size, state of technology and geography. As $U_D^o - U_D^m > U_A^o - U_A^m$ can be equivalently written as $U_A^m - U_D^m > U_A^o - U_D^o$, in that case the free market provides an inefficiently high degree of welfare inequality between advantaged and disadvantaged countries: improved efficiency goes hand in hand with reduced international inequality. Vice versa, when differences between countries are small, $c_{ll}^{om} < c_A^o < c_D^o$ holds and improved efficiency comes with increased international inequality.

Hence, when differences between advantaged and disadvantaged countries in terms of market size, state of technology and geography are large, the free market provides inefficiently high international inequality. When such differences are small, it provides inefficiently low international inequality.

5.5 Distortions and Externalities

The comparison between free market and efficient outcomes in terms of product selection, product mix and product range shows that the free market errs in all three dimensions. First, as the share of entrants that produce in equilibrium is larger than the share of varieties designed but not produced by the planner, the free market is less selective than the planner. As the share of varieties produced but not exported is smaller for the free market than the planner, inefficiently weak selection in equilibrium affects also exports. This inefficiency is more pronounced in disadvantaged countries.

Second, as the supplied quantity of varieties with lower (higher) unit labor requirement is smaller (larger) for the free market than the planner, the free market offers a sub-optimal product mix that is not skewed enough towards lower cost varieties. This holds for both locally produced and imported varieties. A corollary is that, for given unit labor requirement, the free market product basket gives inefficiently small weight to locally produced vs. imported varieties and, among these, to varieties coming from close vs. distant countries (as, due to iceberg frictions, imported varieties have higher delivered cost than locally produced ones, and imported varieties have higher delivered cost from distant than close countries). Also this inefficiency is more pronounced in disadvantaged countries.

Third, the free market provides an inefficiently narrow (wide) range of varieties to disadvantaged (advantaged) countries. Nonetheless, all countries consume an inefficiently low per-capita amount of the modern good due to the dominant impact of inefficiently low average per-capita consumption of varieties. These inefficiencies are again more severe in disadvantaged countries. As a corollary, the fact that in all countries individual consumption of the modern good is inefficiently low implies that also the global supply of that good is inefficiently low.

The source of inefficiency lies in four types of externalities (Nocco, Ottaviano and Salto, 2014; Behrens et al, 2016; Dhingra and Morrow, 2019). Three of them are at work even in the absence of firm heterogeneity and do not require MSLD. These are the ones highlighted in early models of monopolistic competition (Spence, 1976; Dixit and Stilglitz, 1977). On the one hand, 'love of variety' for the modern good implies that the product range enters utility as a direct argument on top of the quantities consumed. This first type of externality acts as a force tending towards too few varieties since firms do not take into account their positive impact on the product range when deciding to enter and serve any given market. On the other hand, there are two types of externalities that act as forces tending towards too many varieties. By keeping price above marginal cost, firms leave more room for entry in the free market equilibrium than it would happen under (shadow) marginal cost pricing associated with the planner's outcome. Moreover, when firms enter the market, they do not consider their negative impact on rivals' profits.

In general, the net effect on product range is ambiguous as it depends on the cross elasticities of demand. A special case arises with CES demand. Without the traditional good, the opposite externalities exactly offset each other so that the free market and efficient outcomes coincide. With the traditional good ('outside good'), the free market still provides the efficient amount of each variety but, due to markup pricing, an inefficiently small number of varieties. The modern good is, therefore, under-supplied relative to the traditional one. The fact that CES implies the same markup for all firms determines the efficiency of the product mix between locally produced and imported varieties at the free market outcome. The introduction of firm heterogeneity does not alter these properties (Melitz and Redding, 2015).

The fourth type of externality materializes, instead, in the presence of firm heterogeneity and, crucially, MSLD. The fact that demand becomes more inelastic with consumption is reflected in larger markups for firms with lower marginal cost so that these firms do not fully transmit their cost advantage to prices.[24] This generates a positive externality in favor of firms with higher marginal cost. The externality works at the intensive margin: higher marginal cost firms are inefficiently large relative to lower marginal costs firms. It also works at the extensive margin: by keeping price above marginal cost more than higher marginal cost rivals, lower marginal cost firms leave inefficiently larger room for entry. This applies both to domestic and foreign rivals. Hence, with MSLD but not with CES, firm heterogeneity becomes an additional driver of inefficiency.

A lower cutoff reduces these distortions. By reducing the prices of all firms but disproportionately those of firms with lower marginal cost and larger markup, it forces these firms to trasmit more of their cost advantage to prices.[25] This explains why welfare is higher in advantaged countries where sellers face lower cutoffs.

6 Optimal Multilateral Policy

The analysis in the previous section has drawn a complex map of market failures. There are several ways in which the free market outcome departs from the efficient outcome, in terms of product selection, product mix and product range. Moreover, the extent (and sometimes also the direction) of the departures is country or firm specific. We will now characterize the tools that national policy makers can use cooperatively to make the market achieve the efficient outcome in a decentralized fashion. For this to happen, we will need to give the policy makers an unconstrained choice of tools ('first best'), including country-specific and variety/firm-specific consumption/production subsidies/taxes as well as lump-sum transfers for consumers and firms. We will then comment on what policy makers could achieve when deprived of the use of variety/firm-specific consumption/production subsidies/taxes ('second best') and also of lump-sum transfers for firms ('third best').

[24] Recall that the markup on sales from h to l of a firm with marginal cost c is $\mu_{hl}(c) = \left(c_{ll}^m - \tau_{hl}c\right)/2$.

[25] Consider two firms selling from h to l with marginal costs c and c' such that $c < c'$. Their relative price is $p_{hl}(c)/p_{hl}(c') = \left(c_{ll}^m + \tau_{hl}c\right) / \left(c_{ll}^m + \tau_{hl}c'\right)$, which is smaller than 1 and increasing in c_{ll}^m.

6.1 First Best Policies

The efficient outcome can be decentralized through country-pair variety specific per-unit transfers $s_{hl}(c)$ subsidizing (taxing) trade of low (high) marginal cost varieties from country h to country l ($l = 1, ..., M$), complemented by country-specific lump-sum taxes on firms' profits in h and consumers' incomes. For international trade from h to $l \neq h$, per-unit subsidies can indifferently take the form of export subsidies in the production country or import subsidies in the consumption country. Analogously, per-unit taxes can indifferently be export taxes in the production country or import tariffs in the consumption country. For domestic trade within country h, per-unit transfers can indifferently take the form of production subsidies (taxes) for local firms or consumption subsidies (taxes) for local consumers. In any case, due to the externalities discussed in the previous section, free trade is not efficient and restoring efficiency requires policy tools that vary across countries and firms. One size does not fit all.

6.1.1 Per-Unit Transfers

Per-unit transfers are needed to remove the product mix distortion. Consider quantity $q_{hl}(c)$ supplied to country l by a firm producing in country h at marginal cost c. Let $s^{\varsigma}_{hl}(c)$ and $s^{\upsilon}_{hl}(c)$ denote per-unit consumption and production transfers earmarked to that quantity. Given (3), (6), (4) and (5), the revenue the firm earns on $q_{hl}(c)$ becomes $\left(p^{\max}_l + s^{\varsigma}_{hl}(c) - \frac{\gamma}{L_l} q_{hl}(c)\right) q_{hl}(c)$ while the corresponding total production cost becomes $\left(\tau_{hl}c - s^{\upsilon}_{hl}(c)\right) q_{hl}(c)$. The firm's profit then becomes

$$\pi^s_{hl}(c) = \left(p^{\max}_l + s^{\varsigma}_{hl}(c) - \frac{\gamma}{L_l} q_{hl}(c) - \tau_{hl}c + s^{\upsilon}_{hl}(c)\right) q_{hl}(c),$$

which shows that the distinction between per-unit consumption and production transfers is immaterial from the perspective of profit maximization. Accordingly, for parsimony we introduce the bundling notation $s_{hl}(c) \equiv s^{\varsigma}_{hl}(c) + s^{\upsilon}_{hl}(c)$. The profit-maximizing quantity then evaluates to

$$q^s_{hl}(c) = \frac{L_l}{2\gamma}\left(p^{\max}_l + s_{hl}(c) - \tau_{hl}c\right), \tag{29}$$

which is non-negative as long as c is not larger than the cutoff c^s_{hl}, i.e. the value of the marginal cost such that

$$p^{\max}_l = \tau_{hl}c^s_{hl} - s_{hl}(c^s_{hl}). \tag{30}$$

The efficient outcome is achieved when the corresponding price

$$p^s_{hl}(c) = \frac{1}{2}\left(p^{\max}_l - s_{hl}(c) + \tau_{hl}c\right)$$

equals the delivered marginal cost $\tau_{hl}c$ for $c^s_{hl} = c^o_{hl}$ so that the after-transfer marginal delivered cost is equalized across all sellers to l ('level playing field'). This is the case for

$$s_{hl}(c) = -s_{hl}(c^o_{hl}) + \tau_{hl}\left(c^o_{hl} - c\right). \tag{31}$$

As, given (30), $s_{hl}(c^o_{hl}) = 0$ identifies the unique transfer schedule that allows for $p^s_{hl}(c^o_{hl}) = \tau_{hl} c^o_{hl}$, (31) can be rewritten as

$$s^o_{hl}(c) = \tau_{hl} \left(c^o_{hl} - c \right). \tag{32}$$

The efficient per-unit transfer is decreasing in marginal cost, being zero for firms with $c = c^o_{hl}$, negative ('tax') for high marginal cost firms with $c \in (c^o_{hl}, c_{M,h}]$ and positive ('subsidy') for low marginal cost firms with $c \in [0, c^o_{hl})$. Hence, trade by low cost firms is subsidized whereas trade by high cost firms is taxed. Equivalently, low cost varieties enjoy export or import subsidies whereas high cost varieties face export taxes or import tariffs. In both cases transfers are bigger for more distant shipments (larger τ_{hl}) to disadvantaged countries (larger c^o_{hl}).

For shipments to any given country l, the average per-unit transfer across all firms and countries is

$$\bar{s}^o_l = \frac{2k+1}{k+1} c^o_{ll},$$

which is larger for disadvantaged countries.[26]

6.1.2 Lump-Sum Transfers

Lump-sum transfers are needed to deal with the distortions in product selection and product range. Let S_h be a lump-sum transfer for firms in country h. It is a subsidy if positive and a tax if negative. Marginal cost pricing implies that producers make no profits so that all they eventually earn in excess of marginal cost comes from per-unit subsidies. Accordingly, given (29), (30) and (32), the earnings on quantity supplied to l by a firm producing in h at marginal cost c evaluate to

$$s^o_{hl}(c) q^o_{hl}(c) = \frac{L_l}{\gamma} \left(\tau_{hl} \right)^2 \left(c^o_{hl} - c \right)^2.$$

The 'free entry condition' in country h can then be stated as

$$\sum_{l=1}^{M} \left[\int_0^{c^o_{hl}} \frac{L_l}{\gamma} \left(\tau_{hl} \right)^2 \left(c^o_{hl} - c \right)^2 dG_h(c) \right] + S_h = f_h, \tag{33}$$

which by (2) and (19) can be rewritten as

$$\sum_{l=1}^{M} \left[\left(\tau_{hl} \right)^{-k} L_l \left(c^o_{ll} \right)^{k+2} \right] = \frac{\gamma (k+1)(k+2)(f_h - S_h)(c_{M,h})^k}{2}.$$

[26] The average subsidy received by firms producing in h for sales to l is defined as $\bar{s}^o_{hl} = \int_0^{c^o_{hl}} s^o_{hl}(c) dG_l(c) / G_l(c_{hl})$, which by (2) evaluates to $\bar{s}^o_{hl} = [(2k+1)/(k+1)] \tau_{hl} c^o_{hl}$. Given (19), this can be rewritten as $\bar{s}^o_{hl} = [(2k+1)/(k+1)] c^o_{ll}$ and thus it does not depend on the country of production. Averaging across countries of production then obviously gives $\bar{s}^o_l = \bar{s}^o_{hl}$.

For $l = 1, ..., M$ this yields a system of M equations that can be solved for the M equilibrium cutoffs

$$c_{ll}^{o} = \left\{ \frac{\gamma\left(k+2\right)\left(k+1\right)}{L_{l}} \frac{\sum_{h=1}^{M}\left[\frac{f_{h}-S_{h}}{2}\left(c_{M,h}\right)^{k}\left|C_{hl}\right|\right]}{\left|P\right|} \right\}^{\frac{1}{k+2}}. \quad (34)$$

Comparing (34) with (21) reveals that decentralization of the efficient outcome requires to set $S_h = -f_h$. Being negative, this amounts to a country-specific lump-sum tax on firm profit ($T_h^{\upsilon} = f_h$), which is higher in disadvantaged countries as these face higher innovation costs. Without the profit tax, efficient per-unit transfers would generate expected earnings that are higher than the innovation cost so that free entry would lead to a decentralized cutoff smaller than the efficient one. The lump-sum transfers also implement the efficient numbers of entrants, producers and sellers in each country. Given (30) and (29), $s_{hl}(c_{hl}^{o}) = 0$ implies $p_l^{\max} = c_{ll}^{o}$. This, together with definition (5) and average price $\bar{p}_l^{o} = [k/(k+1)]c_{ll}^{o}$, yields the efficient number of entrants (23).[27] Efficiency can also be gauged from the fact that, as discussed in Section 4.3, the efficient number of sellers is alternatively given by $N_l^{o} = \sum_{h=1}^{M} \rho_{hl} N_{E,h} \left(c_{ll}^{o}/c_{M,h}\right)^{k}$. Then, the derivation we followed for the free market outcome implies that also the number of entrants is the efficient one $N_{E,l}^{o}$. The same holds for the number of producers as this is given by $N_{P,l}^{o} = N_{E,l}^{o}\left(c_{ll}^{o}/c_{M,l}\right)^{k}$.

To close the characterization of efficient decentralization, we need to check whether the revenues from the lump-sum taxes on firms' profits together with those from the per-unit taxes on high marginal cost firms are enough to finance the per-unit subsidies to low marginal cost firms. This can be done by computing the aggregate net per-unit transfers across all firms and countries. These aggregate transfers total

$$S^{o} \equiv \sum_{l=1}^{M}\sum_{h=1}^{M} N_{lh}^{o}\left[\int_{0}^{c_{lh}^{o}} s_{lh}(c) q_{lh}^{o}(c) dG_l(c)/G_l(c_{lh}^{o})\right] = 2\sum_{l=1}^{M} f_l N_{E,l}^{o},$$

where the second equality is granted by the free entry condition (33) and the term between brackets is the average transfer for quantities produced in l and sold in h.[28] As S^o is positive, per-unit taxes do not generate enough revenues to cover per-unit subsidies. Moreover, given $T_l^{\upsilon} = f_l$, $S^o = 2\sum_{l=1}^{M} f_l N_{E,l}^{o}$ implies that aggregate net per-unit transfers $\sum_{l=1}^{M} T_l^{\upsilon} = \sum_{l=1}^{M} f_l$ are twice as large as aggregate tax revenues from lump-sum taxes on firm profits. This deficit can be financed through an additional lump-sum tax on consumers equal to $T^{\varsigma} = \sum_{l=1}^{M} f_l N_{E,l}^{o}$. Hence, the deficit generated by per-unit transfers is equally shared between producers and consumers. Note, however, that the payments of lump-sum profit taxes are earmarked by production country whereas the distribution of the burden of lump-sum consumption taxes is immaterial due to the absence of income effects. The reason for this difference is that the former are used to correct distortions while the latter only for budget balance.[29]

[27] The average delivered price quoted by firms producing in h for sales to l is defined as $\bar{p}_{hl}^{o} = \int_0^{c_{hl}^{o}} p_{hl}^{o}(c) dG_l(c)/G_l(c_{hl})$. By (2) this evaluates to $\bar{p}_{hl}^{o} = [k/(k+1)]\tau_{hl} c_{hl}^{o}$, which in turn by (19) can be rewritten as $\bar{p}_{hl}^{o} = [k/(k+1)]c_{ll}^{o}$. As this does not depend on the country of production, averaging across countries of production gives $\bar{p}_l^{o} = \bar{p}_{hl}^{o}$.

[28] See footnote (26).

[29] Which countries run the bigger deficit per capita before consumer taxation is not a well defined question as it is

6.2 Second and Third Best Policies

The decentralization of the efficient outcome requires to set trade subsidies and taxes that differ not only across countries but also across varieties produced at different marginal cost. We now analyze what is achievable when subsidies and taxes can vary across countries but not across shipments made by firms between the same country pair. When this is the case, policy makers do not have enough tools to remove all distortions. In particular, as they have to use the same per-unit transfer s_{lh} for all shipments from country l to country h, they lack the specific tools needed to target the product mix distortion.

In this second best scenario, policy makers implement the 'constrained' optimal allocation of a planner who cannot affect the relation between quantity and cutoff dictated by (6). This planner thus maximizes welfare (16), subject not only to the resource constraint (17) but also to the product mix constraint (6), with respect to the choice variables q_{0l}, $N_{E,l}$ and c_{lh} (instead of $q_{lh}(c)$) for $l, h = 1, ..., M$. Solving this maximization problem shows that 'constrained' efficient product selection is ruled by the cutoff

$$c_{ll}^{co} = \left\{ \frac{4\gamma (k+2)(k+1)^2}{(2k+1) L_l} \frac{\sum_{h=1}^{M} \left[f_h (c_{M,h})^k |C_{hl}| \right]}{|P|} \right\}^{\frac{1}{k+2}} \tag{35}$$

for $h, l = 1, ..., M$, with the relation between domestic and foreign cutoffs given once more by $c_{hl}^{co} = c_{ll}^{co} / \tau_{hl}$. Profit maximization also determines the 'constrained' efficient number of entrants as

$$N_{E,l}^{co} = \frac{2\gamma (k+1)(c_{M,l})^k \sum_{h=1}^{M} \left[\left(\alpha - \frac{2k+1}{2k+2} c_{hh}^{co} \right) (c_{hh}^{co})^{-(k+1)} |C_{lh}| \right]}{\eta |P|}, \tag{36}$$

with associated number of producers $N_{P,l}^{co} = N_{E,l}^{co} (c_{ll}^{co} / c_{M,l})^k$ and product range

$$N_l^{co} = \sum_{h=1}^{M} \rho_{hl} N_{E,h}^{co} (c_{ll}^{co} / c_{M,h})^k$$

for $h, l = 1, ..., M$.

The 'constrained' efficient outcome exhibits similar properties as the free market and ('unconstrained') efficient outcomes. In particular, also the 'constrained' planner follows a cutoff rule: only varieties with low enough marginal cost ($c \leq c_{lh}^{co}$) are produced in country h for consumption in country l. The cutoff marginal cost c_{lh}^{co} is lower in advantaged countries. Moreover, conditional on the countries of production and consumption, varieties with lower unit labor requirement c are supplied in larger amounts, the more so the lower the cutoff in the country of consumption.

The cutoff is, however, larger for the 'constrained' planner than for the free market outcome and even larger than for the 'unconstrained' planner: $c_{ll}^{co} = [2(k+1)/(2k+1)]^{\frac{1}{k+2}} c_{ll}^{m} = [4(k+1)/(2k+1)]^{\frac{1}{k+2}} c_{ll}^{o}$. This way the 'constrained' planner partially compensates the product mix distortion with larger consumption of the modern good. Accordingly, as shown in Appendix B,

indeterminate whether subsidies (taxes) should come in the form of export subsidies (taxes) in the country of origin or import subsidies (taxes) in the country of destination.

the 'constrained' efficient outcome can be decentralized through a per-unit trade subsidy common to all firms selling to the same given country l

$$s_l^{co} = \frac{1}{2(k+1)} c_{ll}^{co}, \tag{37}$$

matched by a lump-sum profit tax common to all firms producing in the same given country h equal to $T_h^{co} = f_h/(2k+1)$ for all $h, l = 1, ..., M$. The per-unit trade subsidy is thus larger for supplies to disadvantaged countries, which however face also higher lump-sum profit taxes due to higher innovation costs.

Comparing the first and second best policy tools reveals that the 'constrained' efficient per-unit trade subsidy s_l^{co} is smaller than the average 'unconstrained' efficient per-unit trade subsidy $\bar{s}_l^{o}$. The aggregate 'constrained' efficient trade subsidy corresponding to (37) amounts to

$$S^{co} \equiv \sum_{l=1}^{M} \sum_{h=1}^{M} N_{lh}^{co} \int_0^{c_{lh}^{co}} s_h^{co} q_{lh}^{co}(c) dG_l(c)/G_l(c_{lh}^{co}) = \frac{k+2}{2k+1} \sum_{l=1}^{M} f_l N_{E,l}^{co},$$

which is $k+2$ times larger than aggregate revenues from lump-sum profit taxation as these are equal to $\sum_{l=1}^{M} f_l N_{E,l}^{co}/(2k+1)$. This implies that firms bear less than half of the subsidy burden with the rest financed by lump-sum taxes on consumers.

It is also interesting to analyze the situation that corresponds to the traditional 'second-best problem' in entry models without firm heterogeneity (Mankiw and Whinston, 1986). In this case policy makers have the tools needed to manipulate the number of entrants but not those that would allow them to affect firm behavior after entry due to the unavailability of lump sum transfers for firms. They are therefore forced to take that behavior as given. For concreteness, we call this the 'third best scenario', in which policy makers implement the optimal allocation of a planner who is not only unable to affect the relation between quantity and cutoff dictated by (6) and thus cannot remove the product mix distortion, but is also unable to choose the cutoff to deal with the product selection distortion. Specifically, this 'third best planner' maximizes welfare (16) with respect to q_{0l} and $N_{E,l}$ for $l = 1, ..., M$, subject not only to the resource constraint (17) and the product mix constraint (6) but also to the free entry condition (9) as this condition, together with the relation between domestic and foreign cutoffs given once more by $c_{hl}^{m} = c_{ll}^{m}/\tau_{hl}$, imposes the free market cutoff (10) on the planner. Solving this maximization problem yields the third best number of entrants

$$N_{E,l}^{cco} = \frac{2\gamma (k+1) (c_{M,l})^k \sum_{h=1}^{M} \left[\left(\alpha - \frac{1}{2} \frac{2k+3}{k+2} c_{hh}^{m} \right) (c_{hh}^{m})^{-(k+1)} |C_{lh}| \right]}{\eta |P|}, \tag{38}$$

with associated number of producers $N_{P,l}^{cco} = N_{E,l}^{cco} (c_{ll}^{m}/c_{M,l})^k$ and product range

$$N_l^{cco} = \sum_{h=1}^{M} \rho_{hl} N_{E,h}^{cco} (c_{ll}^{m}/c_{M,h})^k$$

for $h, l = 1, ..., M$.

As the third best outcome entails the same cutoff as the free market outcome, it shares the same properties of the free market in terms of selection. However, comparing the two outcomes reveals that in each country the number of entrants, the number of producers, the range of products sold and the range of products exported are richer in the third best allocation than in the free market equilibrium, whereas not only the cutoff but also individual and average quantities supplied by firms are the same in the two outcomes. This way the third best planner partially compensates the product mix and product selection distortions with richer product range of the modern good.

Just like the second best outcome, also the third best outcome can be decentralized through a per-unit trade subsidy common to all firms selling to any given country l

$$s_l^{cco} = \frac{1}{2(k+2)} c_{ll}^m \quad \forall l = 1, ..., M \tag{39}$$

with no associated lump-sum profit tax levied in this case as this tool is not available.[30] The third best subsidy is larger for supplies to disadvantaged countries as these have larger cutoff c_{ll}^m. Moreover, given $c_{ll}^{co} > c_{ll}^m$, comparing (37) with (39) reveals that the per-unit trade subsidy is smaller in the third best than in the second best outcomes: $s_l^{co} > s_l^{cco}$. Together with $\bar{s}_l^o > s_l^{co}$, that implies $\bar{s}_l^o > s_l^{co} > s_l^{cco}$: the third best subsidy s_l^{cco} is smaller than the second best subsidy s_l^{co} and even smaller than the average first best subsidy $\bar{s}_l^o$.

Finally, the aggregate third best trade subsidy corresponding to (39) is given by

$$S^{cco} \equiv \sum_{l=1}^{M} \sum_{h=1}^{M} N_{lh}^{cco} \int_0^{c_{lh}^m} s_h^{cco} q_{lh}^m(c) dG_l(c)/G_l(c_{lh}^m) = \frac{1}{2} \sum_{l=1}^{M} N_{E,l}^{cco} f_l,$$

which is totally financed by lump-sum taxes on consumers as there are no lump-sum tools for firms.

7 Unilateral Deviation

We conclude our analysis by studying whether countries have any incentive to deviate unilaterally from the (globally) efficient outcome and, if that were the case, how deviations would take place depending on countries being advantaged or disadvantaged.

Specifically, we consider the problem faced by a benevolent social planner in country l who maximizes *local* welfare $W_l = U_l L_l$ with U_l as in (1). This local planner takes as exogenously given the endowment of labor, the endowment of the traditional good, the production technologies of the two goods, the trade frictions and the innovation technology of the modern good. The local planner also considers as exogenously given at their globally efficient values all foreign-related variables, including the prices of imports but excluding the prices of modern exports and the bilateral trade flows between country l and all the other countries $h \neq l = 1, ..., M$.

The local planner then solves

$$\max_{\{q_{0l}, N_{E,l}, q_{ll}(c), q_{lh}(c), q_{hl}(c)\}|_{h \neq l=1}^{M}} W = \sum_{l=1}^{M} U_l L_l \tag{40}$$

[30] See Appendix B for additional details.

subject to the country's resource and technology constraints as well as the trade balance condition. The resource and technology constraints together imply

$$q_{0l} = \overline{q}_{0l} + L_l - f_l N_{E,l} - N_{E,l} \sum_{h=1}^{M} \left[\int_0^{c_{M,l}} \tau_{lh} c q_{lh}(c) dG_l(c) \right] + X_{0l}, \tag{41}$$

whereby the consumption of the traditional good equals the sum of its endowment, its local production and its net imports X_{0l}, which due to balanced trade must be matched in value by the net exports of the modern good

$$X_{0l} = \sum_{h=1,h\neq l}^{M} \left(N_{E,l} \int_0^{c_{M,l}} p_{lh}(c) q_{lh}(c) dG_l(c) - N_{E,h}^o \int_0^{c_{M,h}} p_{hl}(c) q_{hl}(c) dG_h(c) \right), \tag{42}$$

where the (shadow) price of the modern good is normalized to 1 by choice of numeraire and the (shadow) prices of exported and imported modern varieties are denoted by $p_{lh}(c)$ and $p_{hl}(c)$ respectively. The export (shadow) price $p_{lh}(c)$ is related to the corresponding quantity by the inverse demand curve

$$p_{lh}(c) = c_{hh}^o - \frac{\gamma}{L_h} q_{lh}(c), \tag{43}$$

where the domestic cutoff c_{hh}^o of foreign country $h \neq l$ is taken as exogenously given at the globally optimal level due the small country assumption. This holds also the number $N_{E,h}^o$ of varieties designed abroad as well as for the export (shadow) price $p_{hl}(c) = p_{hl}^o(c) = \tau_{hl} c$ and the export cutoff c_{lh}^o from h to l with $c_{hh}^o = \tau_{lh} c_{lh}^o$ in light of (19).

The local planner's problem can be solved by first using (41), (42) and (43) to substitute q_{0l} out of (40) and then maximizing the resulting expression with respect to country l's domestic quantities $q_{ll}(c)$, imported quantities $q_{hl}(c)$, exported quantities $q_{lh}(c)$ and number of locally designed varieties $N_{E,l}$.

7.1 Product Mix

The first order condition with respect to $q_{hl}(c)$ gives

$$q_{hl}^u(c) = \begin{cases} \frac{L_l}{\gamma} \left(c_{ll}^u - \tau_{hl} c \right) & c \leq c_{ll}^u / \tau_{hl} \text{ with } c_{ll}^u \equiv \alpha - \eta \frac{Q_l^u}{L_l} \\ 0 & c > c_{ll}^u \end{cases} \tag{44}$$

where 'u' labels the local welfare maximizing values of the variables and

$$Q_l^u \equiv \sum_{h=1}^{M} \left(N_{E,h} \int_0^{c_{ll}^u / \tau_{hl}} q_{hl}^u(c) dG_h(c) \right).$$

The local planner thus follows a cutoff rule for imports: only varieties with low enough unit labor requirement ($c \leq c_{ll}^u / \tau_{hl}$) are allowed into country l and, conditional on the country of origin, varieties with lower unit labor requirement c are supplied in larger amounts than those with higher c, the more so the lower the cutoff in country l. Quantity $q_{hl}^u(c)$ clears the market in a decentralized scenario only if the quantity of each variety produced in h and sold in l is priced at marginal delivered cost $p_{hl}^{u}(c) = \tau_{hl}c$.[31] Accordingly, $p_{hl}^{u}(c)$ does not deviate from $p_{hl}^{o}(c)$ whereas $q_{hl}^{u}(c)$ may deviate from $q_{hl}^{o}(c)$ in (18) to the extent that the locally efficient cutoff c_{ll}^{u} deviates from the globally efficient cutoff c_{ll}^{o} in (10).

The first order condition with respect to $q_{ll}(c)$ gives

$$q_{ll}^{u}(c) = \begin{cases} \frac{L_l}{\gamma}\left(c_{ll}^{u} - c\right) & c \leq c_{ll}^{u} \\ 0 & c > c_{ll}^{u} \end{cases} \tag{45}$$

so that only varieties with low enough unit labor requirement ($c \leq c_{ll}^{u}$) are produced in country l for domestic consumption: varieties with lower unit labor requirement c are supplied in larger amounts than those with larger c, the more so the lower the cutoff c_{ll}^{u}. Quantity $q_{ll}^{u}(c)$ clears the market in a decentralized scenario only if each producer in l prices the quantity sold domestically at marginal cost $p_{ll}^{u}(c) = c$.[32] Accordingly, also $p_{ll}^{u}(c)$ does not deviate from $p_{ll}^{o}(c)$ whereas $q_{ll}^{u}(c)$ may deviate from $q_{hll}^{o}(c)$ to the extent that c_{ll}^{u} deviates from c_{ll}^{o}.

Results (44) and (45) together imply that the relation of the locally efficient cutoff for marginal varieties consumed and produced in l with that for marginal varieties consumed in l but produced in h is given by

$$c_{hl}^{u} = \frac{c_{ll}^{u}}{\tau_{hl}}. \tag{46}$$

Hence, (44), (45) and (46) mirror the corresponding results (18) and (19) in the globally efficient outcome, but with different cutoffs.

The first order condition with respect to $q_{lh}(c)$ gives

$$q_{lh}^{u}(c) = \begin{cases} \frac{L_h}{2\gamma}\left(c_{hh}^{o} - \tau_{lh}c\right) & c \leq c_{hh}^{o}/\tau_{lh} \text{ with } c_{hh}^{o} \equiv \alpha - \eta\frac{Q_h^o}{L_h} \\ 0 & c > c_{lh}^{o} \end{cases} \tag{47}$$

which implies that only varieties with low enough unit labor requirement ($c \leq c_{hh}^{o}/\tau_{lh}$) are exported from country l to country h and varieties with lower unit labor requirement c are exported in larger amounts, the more so the lower the cutoff in country h. Result (47) also implies that the relation of the (globally efficient) cutoff for marginal varieties consumed and produced in h with the locally efficient one for varieties consumed in h but produced in l is regulated by

$$c_{lh}^{u} = \frac{c_{hh}^{o}}{\tau_{lh}}, \tag{48}$$

which in turn implies $c_{lh}^{u} = c_{lh}^{o}$ given (19). Substituting (47) into inverse demand (43) then yields

$$p_{lh}^{u}(c) = \frac{1}{2}\left(c_{hh}^{o} + \tau_{lh}c\right) \tag{49}$$

so that the locally efficient exported quantity $q_{lh}^{u}(c)$ clears the market in a decentralized scenario only

[31] This can be seen by substituting $p_{hl}(c) = p_{hl}^{u}(c) = \tau_{hl}c$ in the inverse demand function – analogous to (43) – $p_{hl}(c) = \alpha - \gamma q_{hl}(c)/L_l - \eta Q_l^u/L_l$ and using the definition $c_{ll}^{u} = \alpha - \eta Q_l^u/L_l$ to obtain $q_{hl}(c) = q_{hl}^{u}(c) = \left(c_{ll}^{u} - \tau_{hl}c\right)L_l/\gamma$.

[32] This can be seen by substituting $p_{ll}(c) = p_{ll}^{u}(c) = c$ in the inverse demand function – analogous to (43) – $p_{ll}(c) = \alpha - \gamma q_{ll}(c)/L_l - \eta Q_l^u/L_l$ and using the definition $c_{ll}^{u} = \alpha - \eta Q_l^u/L_l$ to obtain $q_{ll}(c) = q_{ll}^{u}(c) = \left(c_{ll}^{u} - c\right)L_l/\gamma$.

if exports are priced above marginal cost with markup $\mu^u_{lh}(c) = p^u_{lh}(c) - \tau_{lh}c = (c^o_{hh} - \tau_{lh}c)/2$. While this mirrors the free market result with markup $\mu^m_{lh}(c) = (c^m_{hh} - \tau_{lh}c)/2$ (see Section 3.1), the local planner's markup is smaller than the market one due to $c^o_{hh} < c^m_{hh}$. Hence, $p^u_{lh}(c)$ deviates from the globally efficient price $p^o_{lh}(c) = \tau_{lh}c$ by a smaller extent than the free market price $p^m_{lh}(c)$.

To recap, the local planner's prices for modern exports to all destinations are above delivered marginal costs and are thus higher than the global planner's ones; locally efficient quantities exported are, instead, below the globally efficient ones. Differently, the local planner's prices for domestic sales and imports are equal to delivered marginal costs and thus coincide with the global planner's ones. Whether the corresponding quantities also coincide depends on whether the two planners' domestic cutoffs in l are the same. Markup pricing on exports implies that a country that unilaterally deviates from the globally efficient outcome exploits foreign love of variety in the modern good to extract rents from its trading partners or, equivalently, to improve its terms of trade.

7.2 Product Selection

To find the local planner's optimal cutoff we have to look at the first order condition with respect to $N_{E,l}$, which requires

$$L_l \left(c^u_{ll}\right)^{k+2} + \frac{1}{2}\sum_{h\neq l=1}^{M} \rho_{lh} L_h \left(c^o_{hh}\right)^{k+2} = \gamma\left(k+2\right)\left(k+1\right)\left(c_{M,l}\right)^k f_l. \tag{50}$$

This can be compared with the analogous expression for the global planner (20) to express the local planner's optimal cutoff as

$$c^u_{ll} = \left[\frac{\left(c^o_{ll}\right)^{k+2} + \left(c^{oa}_{ll}\right)^{k+2}}{2}\right]^{\frac{1}{k+2}}, \tag{51}$$

where c^o_{ll} is the globally efficient cutoff (21) and c^{oa}_{ll} is the value it takes when country l is autarkic. Country l's locally efficient cutoff c^u_{ll} is thus equal to a (geometric) average of its globally efficient cutoff c^o_{ll} when the country is inserted in the trade network and its globally efficient cutoff c^{oa}_{ll} when the country is excluded from the trade network. Then, given $c^o_{ll} < c^{oa}_{ll}$, we have $c^o_{ll} < c^u_{ll} < c^{oa}_{ll}$: the locally efficient domestic cutoff lies above the globally efficient one so that varieties with marginal cost $c \in (c^o_{ll}, c^u_{ll}]$ are supplied by the local planner but not by the global planner. This implies that, while smaller quantities of varieties produced by country l are exported, larger quantities are sold domestically ($q^u_{ll}(c) > q^o_{ll}(c)$). It also implies that larger quantities are imported ($q^u_{hl}(c) > q^o_{hl}(c)$). The unilateral deviations for what globally efficiency would require are larger for advantaged countries. Given $c^{oa}_{ll} = c^o_{ll}/\mathbb{C}_{ll}$, the gap

$$c^u_{ll} - c^o_{ll} = c^o_{ll}\left\{\left[\frac{1 + (1/\mathbb{C})^{k+2}}{2}\right]^{\frac{1}{k+2}} - 1\right\} \tag{52}$$

is larger for advantaged countries as their $\mathbb{C}_{ll}$ is smaller.

7.3 Product Range

As for the number of varieties sold (and thus the 'product range'), we can exploit the definition of c_{ll}^u from (44), the relation between c_{hl}^u and c_{ll}^u from (48), and the expression for Q_l^u obtained using (44) and (45) together with (2) in its definition to get

$$N_l^u = \frac{\gamma(k+1)}{\eta}\frac{\alpha - c_{ll}^u}{c_{ll}^u}. \tag{53}$$

The locally efficient number of varieties sold in l also evaluates to $N_l^u = N_{E,l}^u (c_{ll}^u/c_{M,h})^k + \sum_{h\neq l=1}^M \rho_{hl} N_{E,h}^o (c_{ll}^u/c_{M,h})^k$ while the globally efficient number also evaluates to $N_l^o = N_{E,l}^o (c_{ll}^o/c_{M,h})^k + \sum_{h\neq l=1}^M \rho_{hl} N_{E,h}^o (c_{ll}^o/c_{M,h})^k$. These two expressions can be combined with (53) and (23) to obtain the locally efficient number of varieties designed

$$N_{E,l}^u = N_{E,l}^o + \frac{\gamma(k+1)(c_{M,l})^k}{\eta}\left[\frac{\alpha - c_{ll}^u}{(c_{ll}^u)^{k+1}} - \frac{\alpha - c_{ll}^o}{(c_{ll}^o)^{k+1}}\right]$$

with the corresponding locally efficient number of varieties produced in l for local or export sales given by $N_{P,l}^u = N_{E,l}^u (c_{ll}^u/c_{M,l})^k$ and $N_{lh}^u = N_{E,l}^u (c_{lh}^u/c_{M,l})^k$ respectively. As in the globally efficient outcome, also in the locally efficient outcome more varieties are sold (N_l^u is larger) in advantaged countries as these have lower c_{ll}^u. However, given $c_{ll}^u > c_{ll}^o$, the locally efficient numbers of varieties designed, produced and sold are smaller than the globally efficient ones: $N_{E,l}^u < N_{E,l}^o$, $N_{P,l}^u < N_{P,l}^o$ and $N_l^u < N_l^o$.[33] Based on (52), the gaps in these numbers between the local and global planners are larger for advantaged countries.

Finally, indirect utility in the deviating country can be written as

$$U_l^u = 1 + \overline{q}_{0l}^{\varepsilon} + \frac{1}{2\eta}(\alpha - c_{ll}^u)^2 + \frac{1}{\gamma(k+2)(k+1)}(c_{ll}^u)^2 \sum_{h\neq l=1}^M N_{lh}^u, \tag{54}$$

which is higher that U_l^o by the revealed preference of the local planner. Comparing (54) with (25) shows that the local planner more than compensate the loss in allocative efficiency captured by the second term on the right hand side of (54) with the gain in foreign rent extraction captured by the fourth term.

In summary, compared with the globally efficient outcome, fewer varieties are designed in the deviating country l ($N_{E,l}^u < N_{E,l}^o$) and fewer varieties are exported ($N_{lh}^u < N_{lh}^o$) in smaller quantities ($q_{lh}^u(c) < q_{lh}^o(c)$) from l to any other country $h \neq l$. By contrast, more labor is allocated to expand supply to domestic consumers ($q_{ll}^u(c) > q_{ll}^o(c)$), activating also the production of less productive varieties (with $c \in (c_{ll}^o, c_{ll}^u]$), even though the overall mass of varieties produced decreases ($N_{P,l}^u < N_{P,l}^o$). In addition, more varieties are imported ($N_{hl}^u > N_{hl}^o$) in larger quantities ($q_{hl}^u(c) > q_{hl}^o(c)$), activating also imports of less productive varieties (those with $c \in (c_{hl}^o, c_{hl}^u]$). Due to $c_{ll}^u > c_{ll}^o$ labor is disproportionately reallocated to the production of varieties with relatively high unit requirements ($q_{ll}^u(c)/q_{ll}^u(c') < q_{ll}^o(c)/q_{ll}^o(c')$ and $q_{hl}^u(c)/q_{hl}^u(c') < q_{hl}^o(c)/q_{hl}^o(c')$ for $c < c'$).

[33] By definition, we have $N_l = \sum_{h=1,h\neq l}^M N_{hl} + N_{P,l}$. The number of imported varieties is $N_{hl}^u = N_{E,h}^o \left(c_{hl}^u/c_M^h\right)^k > N_{hl}^o = N_{E,h}^o \left(c_{hl}^o/c_M^h\right)^k$, which implies $\sum_{h=1,h\neq l}^M N_{hl}^u > \sum_{h=1,h\neq l}^M N_{hl}^o$. Then, given that $N_l^u < N_l^o$ holds, it follows that also $N_{P,l}^u < N_{P,l}^o$ must hold.

7.4 Deviant Policies

The policy tools needed to decentralize the locally efficient outcome can be determined through the same logic followed to decentralize the globally efficient outcome in Section (6.1). Accordingly, the locally efficient quantities of imports and domestic sales can be implemented at marginal cost pricing by introducing per-unit transfers equal to

$$s_{hl}^{u}(c) = \tau_{hl}\left(c_{hl}^{u} - c\right) = c_{ll}^{u} - \tau_{hl} c$$

and

$$s_{ll}^{u}(c) = c_{ll}^{u} - c \tag{55}$$

respectively. For both domestic sales and shipments from any foreign country h, the average per-unit transfer across all firms and countries is

$$\bar{s}_{l}^{u} = \frac{2k+1}{k+1} c_{ll}^{u},$$

which is larger than in the globally efficient outcome for all countries and larger for disadvantaged than advantaged countries. Differently, the implementation of the locally efficient quantities of exports at markup pricing does not require any active policy intervention as, for given c_{hh}^{o}, firms by themselves sell $q_{lh}^{u}(c)$ at price set price $p_{lh}^{u}(c)$ under laissez-fair.

Implementing the locally efficient cutoff c_{ll}^{u} still requires active policy intervention in terms of a lump-sum transfer S_f for firms. To see this, consider the free entry condition

$$\int_{0}^{c_{ll}^{u}} s_{ll}^{u}(h) q_{ll}^{u}(h) dG_l(c) + \sum_{h \neq l=1}^{M} \left[\int_{0}^{c_{lh}^{o}} \mu_{lh}^{u}(c) q_{lh}^{u}(c) dG_l(c) \right] + S_f = f_l.$$

The first term on the left hand side refers to the earnings that prospective entrants can expect on domestic sales. As firms price at marginal cost, all earnings come in the form of per-unit subsidies. The second term refers to profits gained on export sales thanks to markup pricing. Using (45), (47), (49) and (55) together with (2) and $\mu_{lh}^{u}(c) = \left(c_{hh}^{o} - \tau_{lh} c\right)/2$, a firm's free entry condition can be restated as

$$2L_l \left(c_{ll}^{u}\right)^{k+2} + \frac{1}{2} \sum_{h \neq l=1}^{M} \rho_{lh} L_h \left(c_{hh}^{o}\right)^{k+2} = \gamma \left(k+2\right)\left(k+1\right)\left(c_{M,l}\right)^{k} \left(f_l - S_f\right)$$

Comparison with (50) reveals that for c_{ll}^{u} to solve the free entry condition the lump-sum transfer must be equal to

$$S_f = -\frac{L_l \left(c_{ll}^{u}\right)^{k+2}}{\gamma \left(k+2\right)\left(k+1\right)\left(c_{M,l}\right)^{k}}.$$

This is a lump-sum tax on profit that is larger in advantaged than disadvantaged countries.

Hence, the deviating country's policies consist of a lump-sum profit tax, per-unit subsidies to domestic sales and imports, and laissez-faire for exports.

8 Conclusion

We have addressed the question how multilateral trade policy should be designed in a world in which countries differ in terms of market access and technology, and firms with market power differ in terms of productivity. We have argued that, in general, the answer depends on market structure, demand characteristics and technological constraints. In the 'canonical' models of monopolistic competition with CES demand, iceberg trade friction, sunk entry costs, fixed production and constant marginal costs that are (inverse) Pareto distributed across firms, the free market equilibrium is efficient. Accordingly, free trade is the best multilateral trade policy and there is no room for welfare improving policy intervention.

This property of the free market equilibrium does not carry on to monopolistic competitive models in which demand is not CES. We have argued that an important departure from CES materializes when demand satisfies 'Marshall's Second Law of Demand' (MSLD), according to which demand becomes more inelastic with consumption. We have shown that, in a model with linear demand satisfying MSLD, the free trade allocation of resources is inefficient in terms of product range, product selection and product mix, and that the extent of inefficiency varies across countries depending on market size, state of technology and geography.

We have used the term 'disadvantaged' to refer to countries with smaller market size, worse state of technology (in terms of higher innovation and average production costs), and worse geography (in terms of more pronounced remoteness from countries with better state of technology). We have found that, from a global welfare viewpoint, optimal multilateral trade policy should act as follows. On the one hand, to remove the product mix inefficiency, it should promote the sales of low cost firms to all countries, but especially to disadvantaged ones. It should also trim the sales of high cost firms to all countries, but especially to disadvantaged ones. On the other hand, to simultaneously remove the product range and product selection inefficiencies, it should reduce firm entry in all countries, but especially in disadvantaged ones. This would not only restore efficiency but also reduce welfare inequality between advantaged and disadvantaged countries provided that their differences in market size, state of technology and geography are large enough.

Such an optimal trade policy requires to set trade subsidies and taxes that differ not only across countries but also across products supplied at different marginal cost. We have also analyzed what is achievable in a restricted scenario in which subsidies and taxes vary across countries but not across firms. In this case, the product mix inefficiency cannot be targeted specifically and the resulting 'constrained' optimal trade policy should (partially) compensate the welfare loss due to the product mix distortion with larger consumption of all products, especially in disadvantaged countries. When the additional unavailability of lump sum tools for firms makes it impossible to target not only the product mix distortion but also the product selection distortion the (even more) 'constrained' optimal trade policy should (partially) compensate the corresponding welfare losses with richer product variety, especially in disadvantaged countries. Finally, we have shown that all countries have a unilateral incentive to deviate from the optimal multilateral outcome in order to extract rents from their trading partners or, equivalently, to improve their terms of trade. In doing so, advantaged countries impose bigger allocative distortions to their own economies than disadvantaged countries do as the former are able to extract more rents than the latter.

References

[1] Arkolakis C., A. Costinot and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review* 102. 94-130.

[2] Arkolakis, C., A. Costinot, D. Donaldson and A. Rodríguez-Clare (2015). The elusive pro-competitive effects of trade, NBER Working Paper n.21370. Forthcoming: *Review of Economic Studies.*

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

[4] Bagwell, K., and S. H. Lee (2015). Trade policy under monopolistic competition with firm selection, Stanford University, mimeo.

[5] Bagwell, K., and R. W. Staiger (2016). *Handbook of Commercial Policy* (Amsterdam: Elsevier).

[6] Bagwell, K., C.P. Bown and R.W. Staiger (2016). Is the WTO passé?, *Journal of Economic Literature* 54, 1125-1231.

[7] Behrens, K., G. Mion, Y. Murata and J. Suedekum (2016). Distorted monopolistic competition, CEPR Discussion Paper n.11642.

[8] Berman, N., P. Martin, and T. Mayer (2012). How do different exporters react to exchange rate changes? *Quarterly Journal of Economics* 127, 437-492.

[9] Bernard, A.B., J. Eaton, J.B. Jensen and S. Kortum (2003). Plants and productivity in international trade. *American Economic Review* 93, 1268-1290.

[10] Bertoletti, P. and P. Epifani (2014). Monopolistic competition: CES redux?, *Journal of International Economics* 93, 227-238.

[11] Bertoletti, P., F. Etro and I. Simonovska (2016). International trade with indirect additivity. NBER Working Paper n.21989. Forthcoming: American Economic Journal: Microeconomics.

[12] Bhagwati, J., and V.K. Ramaswami (1963). Domestic distortions, tariffs and the theory of optimum subsidy, *Journal of Political Economy* 71, 44-50.

[13] Blonigen, B.A. and T.J. Prusa (2003). Antidumping, in Choi E.K. and J. Harrigan (eds.) *Handbook of International Trade* (Oxford: Blackwell).

[14] Brander, J., and P. Krugman (1983). A 'reciprocal dumping' model of international trade, *Journal of International Economics* 15, 313-321.

[15] Campolmi, A., H. Fadinger and C. Forlati (2014). Trade policy: Home market effect versus terms-of-trade externality, *Journal of International Economics* 93, 92-107.

[16] Campolmi, A., H. Fadinger and C. Forlati (2018). Trade and domestic policies in models with monopolistic competition. CEPR Discussion Paper n.13219.

[17] Costinot A., A. Rodriguez-Clare and I. Werning (2016) Micro to macro: Optimal trade policy with firm heterogeneity. NBER Working Paper n.21989.

[18] Costinot, A., D. Donaldson, J. Vogel and I. Werning (2015). Comparative advantage and optimal trade policy. *Quarterly Journal of Economics* 130, 659-702.

[19] De Loecker, J., and P. Goldberg (2014). Firm performance in a global market. *Annual Review of Economics* 6, 201-227.

[20] De Loecker, J., P. Goldberg, A. Khandelwal and N. Pavcnik (2016). Prices, Markups and Trade Reform. *Econometrica* 84, 445-510.

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

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

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

[24] Dixit, A., and J. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67, 297-308.

[25] Dornbusch, R. (1987). Exchange rates and prices. *American Economic Review* 77, 93-106.

[26] Edmond, C., V. Midrigan and D.Y. Xu (2015). Competition, markups and the gains from international trade. *American Economic Review* 105, 3183-3221.

[27] Epifani, P., and G. Gancia (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics* 83, 1-13.

[28] Fabinger, Michal, and E. Glen Weyl (2014). A tractable approach to pass-through patterns with applications to international trade, University of Chicago, mimeo.

[29] Felbermayr, G., B. Jung and M. Larch (2013). Optimal tariffs, retaliation and the welfare loss from tariff wars in the Melitz model. *Journal of International Economics* 89, 13-25.

[30] Flam, H., and E. Helpman (1987). Industrial policy under monopolistic competition. *Journal of International Economics* 22, 79-102.

[31] Gros, D., (1987). A note on the optimal tariff, retaliation and the welfare loss from tariff wars in a framework with intra-industry trade. *Journal of International Economics* 23, 357-367.

[32] Grossman G.M. (ed.) (1992) *Imperfect Competition and International Trade* (Cambride MA: MIT Press).

[33] Grossman, G.M. and E. Helpman (2001) *Special interest politics* (Cambridge MA: MIT Press).

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

[35] Helpman E. and P. Krugman (1989) *Trade Policy and Market Structure* (Cambridge MA: MIT Press).

[36] Holmes, T.J., W.-T. Hsu and S. Lee (2014). Allocative efficiency, mark-ups, and the welfare gains from trade. *Journal of International Economics* 94, 195-206.

[37] Hopenhayn, H. (2014). Firms, Misallocation, and Aggregate Productivity: A Review. *Annual Review of Economic*s 6, 735-770.

[38] Hsieh, C.-T., and P.J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Econonomics* 124, 1403-48.

[39] Jung, J.-W., I. Simonovska and A. Weinberger (2015). Exporter heterogeneity and price discrimination: A quantitative view. NBER Working paper n.21408. Forthcoming: *Journal of International Economics.*

[40] Kolev, D.R., and T.J. Prusa (2002). Dumping and double crossing: The (in) effectiveness of cost-based trade policy under incomplete information. *International Economic Review* 43, 895-918.

[41] Krugman, P.R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics* 9, 469-479.

[42] Krugman, P.R. (1987). Pricing to market when the exchange rate changes, in Arndt S.W. and J.D. Richardson (eds.) *Real-Financial Linkages among Open Economies* (Cambridge MA: MIT Press).

[43] Mankiw, N.G., and M.D. Whinston (1986). Free entry and social inefficiency. *RAND Journal of Economic*s 17, 48-58.

[44] Marshall, A. (1920). *Principles of Economics, An Introductory Volume, eighth edn.* (London: Macmillan).

[45] Mayer, T., M. Melitz and G.I.P. Ottaviano (2014). Market size, competition, and the product mix of exporters. *American Economic Review* 104, 495-536.

[46] Mayer, T, M. Melitz and G.I.P. Ottaviano (2016). Product mix and firm productivity responses to trade competition. NBER Working paper n.22433.

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

[48] Melitz M.J. and G.I.P. Ottaviano (2008). Market size, trade, and productivity, *Review of Economic Studies* 75, 295-316.

[49] Melitz M.J. and S. Redding (2014). Heterogeneous firms and trade, in Gopinath G., E. Helpman and K. Rogoff (eds.) *Handbook of International Economics* (Amsterdam: North Holland).

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

[51] Mrazova, M., and J.P. Neary (2013). Selection effects with heterogeneous firms. CEPR Discussion Paper n.9288. Forthcoming: *Journal of the European Economic Association.*

[52] Nocco, A., G.I.P. Ottaviano and M. Salto (2014). Monopolistic competition and optimum product selection, *American Economic Review* 104, 304-309.

[53] Nocco, A., G.I.P. Ottaviano and M. Salto (2017). Monopolistic competition and optimum product selection: why and how heterogeneity matters. *Research in Economics* 71, 704-717.

[54] Ossa, R. (2011). A "new trade" theory of GATT/WTO negotiations, *Journal of Political Economy* 119, 122-152.

[55] Ossa, R. (2014) Trade wars and trade talks with data, *American Economic Review* 104, 4104-4146.

[56] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002). Agglomeration and trade revisited. *International Economic Review* 43, 409-436.

[57] Peters, M. (2018). Heterogeneous markups, growth and endogenous misallocation. Yale University, mimeo.

[58] Spence, M. (1976). Product selection, fixed costs and monopolistic competition, *Review of Economic Studies* 43, 217-235.

[59] Srinivasan, T.N. (1996). The generalized theory of distortions and welfare two decades later, in Bhagwati, J.N., R.C. Feenstra, G.M. Grossman and D.A. Irwin (eds.) *The Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati* (Cambridge: MIT Press).

[60] Venables, A., (1987). Trade and trade policy with differentiated products: A Chamberlinian-Ricardian model. *Economic Journal* 97, 700-717.

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

9 Appendix A: Trade-Related Selection

To show that $\mathbb{C}_{ll} < 1$ holds with trade as long as there is some production of the modern good in all countries, consider (12) and rearrange it as

$$\frac{1}{|P|}\sum_{h=1}^{M}\left[f_h\left(c_{M,h}\right)^k\left|C_{hl}\right|\right] = f_l\left(c_{M,l}\right)^k\left(\mathbb{C}_{ll}\right)^{k+2} = \frac{L_l\left(c_{ll}^m\right)^{k+2}}{2\gamma\left(k+1\right)\left(k+2\right)}, \tag{56}$$

where the second equality is granted by (10). By construction, the cutoff c_{ll}^m solves the system of M free entry conditions (9), which can be restated as

$$\sum_{h=1}^{M} \rho_{lh} \frac{L_h \left(c_{hh}^m\right)^{k+2}}{2\gamma \left(k+2\right)\left(k+1\right)} = f_l \left(c_{M,l}\right)^k$$

for $l = 1, ..., M$. Hence, by (56), $f_l \left(c_{M,l}\right)^k \left(\mathbb{C}_{ll}\right)^{k+2}$ solves the system

$$\sum_{h=1}^{M} \rho_{lh} f_h \left(c_{M,h}\right)^k \left(\mathbb{C}_{hh}\right)^{k+2} = f_l \left(c_{M,l}\right)^k \tag{57}$$

for $l = 1, ..., M$. Rewrite (57) as

$$f_l \left(c_{M,l}\right)^k \left(\mathbb{C}_{ll}\right)^{k+2} + \sum_{h \neq l=1}^{M} \rho_{lh} f_h \left(c_{M,h}\right)^k \left(\mathbb{C}_{hh}\right)^{k+2} = f_l \left(c_{M,l}\right)^k$$

and then as

$$\left(\mathbb{C}_{ll}\right)^{k+2} = 1 - \sum_{h \neq l=1}^{M} \rho_{lh} \frac{f_h}{f_l} \left(\frac{c_{M,h}}{c_{M,l}}\right)^k \left(\mathbb{C}_{hh}\right)^{k+2},$$

which shows that $\mathbb{C}_{hh} > 0 \; \forall l = 1, ..., M$ implies $\mathbb{C}_{ll} < 1 \; \forall l = 1, ..., M$.

10 Appendix B: Derivation of Second and Third Best Policies

In the second and third best scenariosdiscussed in Section 6.2, policy makers implement the optimal allocations of planners 'constrained' to use a per-unit production subsidy common to all shipments from any given country h to any other country l. If we use s_{hl}^r with $r = co$ and $r = cco$ to denote such subsidy in the second and the third best scenarios respectively, the subsidized profit of a firm with marginal cost c supplying quantity $q_{hl}^r(c)$ from h to l is

$$\pi_{hl}^r(c) = \left(p_l^{\max} + s_{hl}^r - \frac{\gamma}{L_l} q_{hl}^r(c) - \tau_{hl} c\right) q_{hl}^r(c),$$

with profit-maximizing subsidized quantity

$$q_{hl}^r(c) = \frac{L_l}{2\gamma} \left(p_l^{\max} + s_{hl}^r - \tau_{hl} c\right). \tag{58}$$

For given subsidy, the choke price $p_l^{\max}$ pins down the highest marginal cost c_{hl}^r such that $q_{hl}^r(c)$ is non-negative:

$$p_l^{\max} = \tau_{hl} c_{hl}^r - s_{hl}^r. \tag{59}$$

Accordingly, (58) can be restated as

$$q_{hl}^r(c) = \begin{cases} \frac{L_l}{2\gamma} \tau_{hl} \left(c_{hl}^r - c\right) & c \leq c_{hl}^r \\ 0 & c > c_{hl}^r \end{cases} \tag{60}$$

with corresponding profit-maximizing subsidized price

$$p_{hl}^{r}(c) = \frac{1}{2}\left(p_{l}^{\max} - s_{hl}^{r} + \tau_{hl}c\right), \tag{61}$$

and maximized profit

$$\pi_{hl}^{r}(c) = \frac{L_l}{4\gamma}\left(\tau_{hl}\right)^2\left(c_{hl}^{r} - c\right)^2.^{34} \tag{62}$$

Expression (60) implies that country l's domestic cutoff c_{ll}^{r} (such that $q_{ll}^{r}(c_{ll}^{r}) = 0$) and any other country h's export cutoff c_{hl}^{r} to l (such that $q_{hl}^{r}(c_{hl}^{r}) = 0$) are related by $c_{hl}^{r} = c_{ll}^{r}/\tau_{hl}$. Then, (59) in turn implies $p_{l}^{\max} = c_{ll}^{r} - s_{hl}^{r}$ and $p_{l}^{\max} = c_{ll}^{r} - s_{ll}^{r}$ so that we have

$$s_{hl}^{r} = s_{ll}^{r} \tag{63}$$

and all firms selling to a given country receive the same subsidy for local sales, which we henceforth denote by s_{l}^{r}.

The number N_{l}^{r} of firms selling to h can be found by noting that, by (5), we have

$$p_{l}^{\max} = \frac{1}{\eta N_{l}^{r} + \gamma}\left(\gamma\alpha + \eta N_{l}^{r}\bar{p}_{l}^{r}\right), \tag{64}$$

with average price

$$\bar{p}_{l}^{r} = \frac{2k+1}{2(k+1)}c_{ll}^{r} - s_{l}^{r}, \tag{65}$$

as (2), (61), $c_{hl}^{r} = c_{ll}^{r}/\tau_{hl}$ and $s_{hl}^{r} = s_{l}^{r}$ together imply that the average price set in l by firms producing in h is the same for all $h = 1, ..., M$. Substituting $p_{l}^{\max} = c_{ll}^{r} - s_{l}^{r}$ and $\bar{p}_{l}^{r}$ from (65) into (64) yields

$$N_{l}^{r} = \frac{2\gamma(k+1)}{\eta}\frac{\left(\alpha - c_{ll}^{r} + s_{l}^{r}\right)}{c_{ll}^{r}}. \tag{66}$$

Finally, the number of entrants can be determined as follows. By definition, given (2), $c_{hl}^{r} = c_{ll}^{r}/\tau_{hl}$, $N_{hl}^{r} = N_{E,h}^{r}\left(c_{hl}^{r}/c_{M,h}\right)^k$ and $N_l = \sum_{h=1}^{M} N_{hl}$, the equilibrium number of sellers to l is $N_{l}^{r} = \sum_{h=1}^{M}\rho_{hl}N_{E,h}^{r}\left(c_{ll}^{r}/c_{M,h}\right)^k$. This expression can be combined with (66) to obtain a system of M linear equations for $l = 1, ..., M$ that can be solved for the equilibrium number of entrants

$$N_{E,l}^{r} = \frac{2\gamma(k+1)\left(c_{M,l}\right)^k \sum\limits_{h=1}^{M}\left[\frac{\left(\alpha - c_{hh}^{r} + s_{h}^{r}\right)}{\left(c_{hh}^{r}\right)^{k+1}}\left|C_{lh}\right|\right]}{\eta\left|P\right|}. \tag{67}$$

10.1 Second Best

The implementation of $N_{E,l}^{co}$ requires setting a per-unit subsidy $s_{h}^{r}|_{r=co}$ in all countries such that $N_{E,l}^{co}$ in (36) coincides with $N_{E,l}^{r}\Big|_{r=co}$ in (67):

$$\sum_{h=1}^{M}\left[\frac{\left(\alpha - c_{hh}^{co} + s_{h}^{co}\right)}{\left(c_{hh}^{co}\right)^{k+1}}\left|C_{lh}\right|\right] = \sum_{h=1}^{M}\left[\frac{\left(\alpha - \frac{2k+1}{2k+2}c_{hh}^{co}\right)}{\left(c_{hh}^{co}\right)^{k+1}}\left|C_{lh}\right|\right].$$

This holds for

$$s_h^{co} = \frac{1}{2(k+1)} c_{hh}^{co},$$

which corresponds to expression (37) in Section 6.2.

The implementation of c_{hh}^{co} requires setting a lump-sum profit tax T_h^{co} in all countries such that c_{hh}^{co} solves the 'free entry condition'

$$\sum_{l=1}^{M} \left[\int_0^{c_{hl}^{co}} \frac{L_l}{4\gamma} (\tau_{hl})^2 (c_{hl}^{co} - c)^2 \, dG_h(c) \right] = f_h + T_h^{co},$$

which, by (2) and $c_{hl}^{co} = c_{ll}^{co}/\tau_{hl}$, can be rewritten as

$$\sum_{l=1}^{M} \left[\rho_{hl} L_l (c_{ll}^{co})^{k+2} \right] = 2\gamma (c_{M,h})^k (k+2)(k+1)(f_h + T_h^{co})$$

for $h = 1, ..., M$. This yields a system of M equations that can be solved to find the M equilibrium cutoffs

$$c_{ll}^{co} = \left\{ \frac{2\gamma (k+2)(k+1)}{L_l} \frac{\sum_{h=1}^{M} \left[(c_{M,h})^k (f_h + T_h^{co}) |C_{hl}| \right]}{|P|} \right\}^{\frac{1}{k+2}}. \tag{68}$$

Accordingly, the profit tax T_h^{co} that makes (68) coincide with (35) solves

$$\sum_{h=1}^{M} \left[(c_{M,h})^k (f_h + T_h^{co}) |C_{hl}| \right] = \frac{2(k+1)}{2k+1} \sum_{h=1}^{M} \left[f_h (c_{M,h})^k |C_{hl}| \right]$$

and it is therefore equal to

$$T_h^{co} = \frac{f_h}{2k+1}$$

as stated in Section 6.2.

10.2 Third Best

The implementation of $N_{E,l}^{cco}$ requires setting a per-unit subsidy $s_h^r|_{r=cco}$ in all countries such that $N_{E,l}^{cco}$ in (38) coincides with $N_{E,l}^r\Big|_{r=cco}$ in (67) given $c_{hl}^{cco} = c_{hl}^m$:

$$\sum_{h=1}^{M} \left[\frac{(\alpha - c_{hh}^m + s_h^{cco})}{(c_{hh}^m)^{k+1}} |C_{lh}| \right] = \sum_{h=1}^{M} \left\{ \frac{\left[\alpha - \frac{2k+3}{2(k+2)} c_{hh}^m \right]}{(c_{hh}^m)^{k+1}} |C_{lh}| \right\},$$

This is the case for

$$s_h^{cco} = \frac{1}{2(k+2)} c_{hh}^m,$$

which corresponds to expression (39) in Section 6.2.