E. J. OURSO COLLEGE OF BUSINESS
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

A Simple Model of Quantity Heterogeneity and International Trade

Elias Dinopoulos
University of Florida

Bulent Unel
Louisiana State University

Working Paper 2009-04
http://www.bus.lsu.edu/economics/papers/pap09_04.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

A Simple Model of Quality Heterogeneity and International Trade*

Elias Dinopoulos
University of Florida

Bulent Unel
Louisiana State University

Current Version: February 13, 2009

Abstract

This paper develops a trade model with firm-specific quality heterogeneity, limit pricing, and an endogenous distribution of markups. Exposure to trade induces only the firms producing high-quality (high-price) products to enter the export markets, whereas firms producing low-quality (low-price) products serve the domestic market in accordance to the Alchian and Allen (1964) conjecture. Trade liberalization intensifies the competition; causes firms producing low-quality products to exit the market; increases the number of products consumed in each country; raises national and global welfare; and generates quality upgrading that results in higher and average domestic and export markups. Interestingly, the laissez-faire equilibrium is inefficient, and this leaves room for welfare-improving government intervention.

An electronic copy of the latest version of the paper is available at
http://bear.cba.ufl.edu/dinopoulos/Research.html
http://www.bus.lsu.edu/unel/research.html

JEL Codes: F10, F12, and F13

Keywords: International trade, monopolistic competition, firm heterogeneity, product quality, trade costs, and trade liberalization.

*Elias Dinopoulos, Department of Economics, University of Florida, Gainesville, FL 32611. E-mail: elias.dinopoulos@cba.ufl.edu; Tel: (352) 392-8150. Bulent Unel, Department of Economics, Louisiana State University, 2134 Patrick F. Taylor Hall, Baton Rouge, LA 70803. E-mail: bunel@lsu.edu, Tel: (225) 578-3792. We would like to thank Tetsu Haruyama and seminar participants at Louisiana State University for useful comments and suggestions. The authors are solely responsible for any possible remaining errors.

1 Introduction

Several empirical studies have documented the presence of substantial firm heterogeneity in narrowly defined product categories.[1] According to these studies, firm heterogeneity takes the form of productivity differences among establishments or quality differences among narrowly defined product categories. Relatively more productive firms are larger, charge lower prices and are more likely to export successfully. Firms producing higher-quality products charge higher prices and are more likely to engage in exporting. Firms face large sunk costs of exporting; there is a large turnover among establishments and reallocation of resources within industries; and the degree of turnover is correlated with exporting activities.

These empirical findings have served as building blocks in the development of a growing strand of theoretical literature that has highlighted the nexus between firm heterogeneity and international trade patterns. Melitz (2003) was the first to develop a tractable model of monopolistic competition and firm heterogeneity based on productivity differences that took into account and generated predictions that are consistent with several aforementioned stylized facts. The theoretical framework proposed by Melitz has been extended by other studies to address the role of firm heterogeneity in several areas of international economics such as foreign direct investment, intersectoral trade flows and the gravity equation.[2] In addition, several recent studies have analyzed the nexus between trade, growth, and firm heterogeneity.[3] It would not be an exaggeration to state that we are witnessing the genesis of a new theory of intraindustry trade in markets characterized by firm heterogeneity.

The present paper contributes to the new theory by proposing a tractable model of monopolistic competition and trade in markets with firm-specific quality heterogeneity.

[1]See Clerides et al. (1998), Bernard and Jensen (1999), Aw et al. (2000), Schott (2004), Hummels and Skiba (2004), Hummels and Klenow (2005), Hallak (2006), and Kugler and Verhoogen (2008) among many others. Tybout (2003) offers an excellent survey of the empirical literature on trade in the presence of heterogeneous firms.

[2]See Helpman (2006) for an insightful survey of the theoretical literature on trade with heterogeneous firms.

[3]Gustafsson and Segerstrom (2007), Baldwin and Robert-Nicoud (2008), Unel (2008), and Haruyama and Zhao (2008) introduce endogenous growth mechanisms in the Melitz model.

The production structure of the model follows closely Melitz's (2003) seminal work. Firms face uncertainty with respect to the future level of product quality when they engage in R&D to discover new products. This uncertainty is modeled by an unrestricted distribution of product quality levels. Production is characterized by constant marginal costs plus fixed production costs. Similarly, exporting involves fixed foreign-market entry costs plus per unit trade costs.

After a firm learns the quality of its product, it faces a unitary-elastic demand curve (which is derived from a Cobb-Douglas utility function) and a competitive fringe of potential imitators that can produce a generic, low-quality version of its product.[4] We assume that the firm competes against the fringe in a Bertrand fashion. Thus, the profit-maximizing strategy for the high-quality firm is to charge a limit price, which is proportional to the product quality level, and to drive the competitive fringe out of the market. Consequently, firms producing higher-quality products charge higher prices and markups, enjoy higher profits, and engage in exporting activities. In contrast, firms producing lower-quality products charge lower prices, earn lower profits, and serve only the domestic market.

Unlike the existing models with heterogeneous firms, trade in the present model affects the intensity of product market competition through the proliferation of varieties, which reduces consumer expenditure per variety and the quantity demanded for each brand. This income effect hurts firms producing low-quality products more than firms producing high-quality products and generates a reallocation of resources from low-quality towards high-quality goods. The trade-induced reallocation of resources puts upward pressure on the real wage which reinforces the aforementioned income effect. Thus trade increases the "cutoff" (zero-profit) quality levels in the domestic and foreign markets and generates quality upgrading, more varieties available for consumption, and higher domestic and export markups.

[4]Preferences are similar to those used routinely in quality-ladders growth models. Segerstrom et al. (1990) and especially Grossman and Helpman (1991), among many others, provide more details on this type of preference structure.

The simultaneous presence of an income-based mechanism, variable markups, and variable input prices (the real wage) within the same model is missing from the rest of the literature. For instance, Melitz's (2003) model that relies on Dixit and Stiglitz (1977) preferences generates endogenous input prices but exogenous markups; whereas Melitz and Ottaviano (2008) rely on quasi-linear preferences to generate endogenous markups, but their model yields exogenous input prices and excludes by assumption the factor-market based mechanism that transmits the effects of trade liberalization.

Having developed a highly-tractable model of quality-heterogeneity and limit-pricing strategies, we analyze the effects of trade liberalization in a global economy consisting of many structurally identical countries. Trade liberalization can take a variety of forms given the rich structure of our model: a move from autarky to restricted trade, a reduction in trade costs, a reduction in foreign-market entry costs, or an increase in the size of the global economy measured by the number of trading partners. All these different facets of trade liberalization generate the same effects: an increase in the intensity of product market competition captured by a reduction in expenditure per variety; a reallocation of resources from low-quality to high-quality products which results in exit of inefficient firms; quality upgrading in the domestic and export markets; an increase in the number of products available for consumption in each country; an increase in average markups; and an improvement in national and global welfare.

The trade-induced rise in average markups does not mean that the intensity of competition declines, as would be the case in models where a typical firm faces variable price demand elasticity which transmits the intensity of product market competition and lowers the markup (see Melitz and Ottaviano, 2008). In the present model, the elasticity of demand is equal to unity and each firm charges a limit price proportional to its quality level that yields a firm-specific constant markup. The trade-induced increase in average markups is a bi-product of the quality upgrading caused by the proliferation of varieties and the reduction in the quantity demanded for a typical variety. This novel aspect that relates quality upgrading and higher average markups is missing from the literature and complements the approach proposed by Melitz and Ottaviano (2008).

Despite the positive welfare impact of trade liberalization, the laissez-faire equilibrium is inefficient. This welfare property can be traced to the difference between the socially optimal and the market-equilibrium average markups. The social planner is interested in the welfare of the average consumer which depends on the average quality and the average consumer surplus, whereas the market is interested in the behavior of the marginal consumer which depends on the product with marginal quality. This difference yields the novel finding that the laissez faire "cutoff" quality levels of firms serving the domestic or foreign markets are socially suboptimal. Therefore, the combination of Cobb-Douglas preferences and limit-pricing strategies reveals the nature of welfare distortions and opens the door for welfare-improving policies in markets with heterogeneous firms. This welfare feature is consistent with the generalized theory of distortions and welfare and the quality-ladders growth theory. However, it is not present in models with Dixit and Stiglitz (1977) preferences and heterogeneous firms. For instance, Feenstra and Kee (2008) use a multi-sector version of Melitz's model to establish that the laissez faire "cutoff" productivity levels of firms producing either for the domestic of foreign markets are socially optimal. This property allows them to estimate a well defined GDP function using cross country data, but it also means that there is no room for desirable government intervention.

Several features of our model enjoy empirical support. Each surviving firm charges a price which is proportional to its product quality level. This feature is consistent with empirical studies that routinely use unit values to measure product quality (see, for instance, Schott (2004), Hammels and Skiba (2004), Hummels and Klenow (2005), and Hallak (2006)). The prediction that the quality (and price) of exports is higher than the quality (and price) of products sold only in the domestic market provides a novel general-equilibrium explanation of the Alchian and Allen (1964, 74-75) conjecture of "shipping the good apples out." This prediction is consistent with the findings of Hummels and Skiba (1994), Verhoogen (2008), and Baldwin and Harrigan (2007). Interestingly, models of productivity heterogeneity are inconsistent with the evidence on export unit values because they predict that more productive exporters charge lower prices than less productive non-exporters.[5] Finally, the prediction that trade-liberalization generates higher average markups in markets with vertical product differentiation raises a word of caution regarding empirical attempts to analyze the effects of trade liberalization on markups. Our result together with the main finding of Melitz and Ottaviano (2008), that trade liberalization generates lower markups, suggest that, without controlling for the nature of competition, one would expect to find an ambiguous effect of trade liberalization on industry markups.[6]

Our paper is also related to a few studies that focus on firm heterogeneity and industry markups. Bernard, Eaton, Jensen, and Kortum (2003) develop a model of heterogeneous firms, limit prices, and firm-specific Ricardian comparative advantage. Their model generates an exogenous distribution of markups and relies on an exogenous number of varieties. In contrast, our model delivers an endogenous distribution of markups, an endogenous number of varieties, and a different resource-reallocation mechanism. As mentioned earlier, Melitz and Ottaviano (2008) develop a model of productivity heterogeneity, quasi-linear preferences, and trade between two unequal-size countries. Their model generates endogenous markups and addresses similar questions to those addressed in our paper. However, in their model the wage is fixed by the presence of the outside good, production and exporting do not involve fixed costs, markets are segmented, and it is not clear how quality heterogeneity can be introduced without substantially complicating the analysis. Despite these differences, both models predict that trade liberalization intensifies the product market competition through variety proliferation in our model and through changes in the price

[5]See, for instance, Schott (2004, p. 676) who states that "unit-value patterns are inconsistent with new trade theory models that have producer price varying inversely with producer productivity."

[6]The study by Harrison (1994) illustrates this point. Using plant-level data from Cote d'Ivoire, she explores changes in productivity and markups following a 1985 trade liberalization episode. She finds that productivity increased, but price-cost markups fell only in few sectors and increased in others following the reform. According to Harrison (1994, Table 5) markups, measured as profits over sales, increased in five out of nine sectors and fell in the rest.

elasticity of demand in theirs. Our findings, therefore, complement their analysis by revealing the positive effects of trade liberalization on markups in markets where firms adopt limit-price strategies and compete with each other in a Bertrand fashion. Finally, Baldwin and Harrigan (2007) propose a model of quality heterogeneity based on Dixit and Stiglitz (1977) preferences where each firm's marginal costs increase in the quality of the firm's product. Their model predicts that high-quality and high-price products will be exported, as in our model, but delivers constant markups.

In summary, our paper makes three novel contributions to the literature on firm heterogeneity and trade. First, it builds a simple model of quality-based firm heterogeneity and trade with limit-pricing strategies, variable markups, and a variable real wage. This allows the study of the impact of trade that is transmitted through both the intensity of product-market competition and the real wage. This general-equilibrium mechanism is missing from other related studies. Second, the model offers several empirically relevant predictions on the pattern and impact of trade in markets with quality-heterogeneous firms. These predictions are very similar to those of other models of trade with heterogeneous firms with the exception of the positive correlation between trade liberalization and average domestic and export markups. Third, consistent with the generalized theory of distortions and welfare, our model demonstrates the sub-optimality of the laissez faire equilibrium and opens the door for welfare improving policies. This important welfare result is also missing from the rest of the relevant literature.

Section 2 of the paper presents the basic elements of the model and the steady-state equilibrium. Section 3 analyzes the impact of trade liberalization and the effects of a move from autarky to trade. Section 4 describes the model's welfare properties, and Section 5 concludes.

2 The Model

In this section, we present the basic elements of the model regarding consumer preferences, structure of production, and firm entry decisions. We consider a global economy consisting of $n + 1$ structurally identical countries with $n \geqslant 1$. Each economy has a single industry populated by heterogeneous firms, and labor is the only factor of production. In each country, the aggregate supply of labor, L, is fixed and remains constant over time.

2.1 Consumer Preferences

Consumer preferences are identical across all countries and modeled by the following Cobb-Douglas utility function defined over a continuum of products indexed by ω

$$U = \int_{\omega \in \Omega} \ln \left[\beta \lambda(\omega) \frac{q(\omega)}{L} \right] d\omega, \tag{1}$$

where $\beta > 0$ is a constant, $\lambda(\omega)$ denotes the *time-invariant* product quality, $q(\omega)$ is the aggregate consumption of brand ω, and Ω is the set of varieties available for consumption in a typical country. We focus our analysis on the case where each consumer buys all available varieties, that is, we assume that the non-satiation principle holds. This case arises if parameter β is sufficiently high to ensure that the utility increases monotonically in the mass of varieties consumed.[7]

Maximizing (1) subject to the budget constraint yields the standard Cobb-Douglas demand for a typical variety

$$q(\omega) = \frac{EL}{p(\omega) M_c}, \tag{2}$$

where E is *per-capita* consumer expenditure, L is the number of consumers in a typical (home or foreign) market, $p(\omega)$ is the corresponding price of brand ω, and M_c is the measure of Ω (i.e., the mass of varieties available for consumption). The market demand for a

[7]The condition $\beta > eL/f_x$, where e is the natural logarithm base and f_x is the fixed foreign market entry cost, guarantees the validity of the non-satiation principle. Section 2.4 provides more details on its derivation. We would like to thank Tetsu Haruyama for pointing this out.

product increases in aggregate consumer expenditure EL; and decreases in price $p(\omega)$ and the number of available products M_c.

2.2 Production

There is a continuum of firms, each choosing to produce a different product variety. Labor is the only factor of production, with each worker supplying one unit of labor. Production involves both fixed and variable costs: in order to produce q units of output, $\ell = f_p + q$ units of labor are required independently of the level of quality, where f_p denotes the fixed overhead cost of production measured in units of labor. Without any loss of generality, this formulation assumes that the marginal cost of production is equal to the wage of labor.

Firms wishing to export must incur per-unit trade costs and fixed costs as in Melitz (2003). Iceberg trade costs (such as transport costs and tariffs) are modeled in the standard fashion: $\tau > 1$ units of output must be produced at home in order for one unit to arrive at its destination. In addition, exporting involves a fixed foreign-market-entry cost of $F_x > 0$ that does not depend on the firm's quality level or the geographic location of production. This cost covers the costs of setting a distribution system, collecting information about the foreign market demand, product modifications and adjustments to local tastes, and costs based on regulations imposed by governments.[8] The decision to export occurs after the product's quality is revealed.

Each incumbent firm faces a constant probability of death δ in each period. In the present context, this stochastic shock can be interpreted as adverse changes in tastes that eliminate the demand for a particular variety. Consequently, in the steady-state, each firm is indifferent in principle between paying $f_x = \delta F_x$ in each period and the one-time fixed cost F_x in the first period of its existence. Hereafter, we assume that in each period exporters face an overhead fixed cost f_x in addition to the overhead production cost f_p. Firms that serve only the domestic market face just the overhead production cost f_p.

[8] Existence of such market costs of exporting have been well documented by several studies (see, for example, Bernard and Jensen (1999); and Tybout (2003)).

Next, consider the optimal pricing decision of a firm selling a brand of quality λ in its home market. Because each brand is associated with a unique quality level, in what follows we label products based on their quality levels. The aggregate quantity demanded is given by equation (2), which implies that that expenditure per variety, $p(\lambda)q(\lambda)$, is independent of the brand's quality level. Because the elasticity of demand for each variety is unity, a typical firm has an incentive to charge an infinite price and produce an infinitesimally small quantity independently of the product's quality level. To prevent this from happening and to create an endogenous distribution of markups, we follow the spirit of Schumpeterian growth theory and assume that once a product is introduced in a market (domestic or foreign), a generic, lower-quality version of the product can be produced instantaneously by a competitive fringe of firms. The production of each generic product exhibits constant returns to scale with one unit of labor producing one unit of output. We suppose that the generic version of a product cannot be produced in a country unless the original product is sold there. In other words, the technology to produce generics diffuses internationally through imports.[9] We normalize the quality level of each generic good to one independently of the quality level of the copied product and the location of production.

Denote with $p_d(\lambda)$ and $p_x(\lambda)$ the consumer price prevailing in the domestic and foreign markets respectively, and assume that competition within each product occurs in a Bertrand fashion. The possibility of costless imitation forces firms to maximize profits by charging a (limit) price no higher than $p_d(\lambda) = p_x(\lambda) = \lambda w$, where w is the common wage rate across all countries, hereafter normalized to unity. This optimal pricing rule drives domestic and foreign imitators out of the market and implies that firms with higher-quality products

[9] Alternatively, one can assume that once a product is developed, its low-quality generic version can be produced by a competitive fringe in all countries, i.e., technology diffuses instantly across all countries. Analysis based on this assumption yields qualitatively the same results, and is available upon request. Moreover, one can also assume that there is no international transfer of technology. This assumption would allow exporters to charge a higher price abroad that would be proportional to the product's quality level adjusted by per-unit trade costs. More precisely, in the absence of international technology transfer, the quality leader charges two limit prices: $p_d(\lambda) = \lambda$ in the domestic market to get rid of the competitive fringe; and (in the absence of a competitive foreign fringe) it charges an export limit price $p_x(\lambda) = \tau\lambda$ which prevents the domestic fringe from exporting. In this case, all the results go through with the exception that trade costs do not affect the export cutoff quality level.

charge higher prices.[10]

The limit-pricing rule $p_d(\lambda) = p_x(\lambda) = \lambda$ and (2) yield

$$\frac{q(\lambda_2)}{q(\lambda_1)} = \frac{p(\lambda_1)}{p(\lambda_2)} = \frac{\lambda_1}{\lambda_2}, \tag{3}$$

which means that firms with higher-quality products charge higher prices and sell lower quantities. In addition, it is obvious from (2) that all firms earn the same revenue $p(\lambda)q(\lambda) = EL/M_c$, thanks to Cobb-Douglas preferences.

The per-period profits of exporting firms can be decomposed into two parts: profits earned from domestic sales $\pi_d(\lambda)$, and profits earned from sales in each of n export markets $\pi_x(\lambda)$.

$$\pi_d(\lambda) = [p_d(\lambda) - 1]q_d(\lambda) - f_p = (1 - \lambda^{-1})\frac{EL}{M_c} - f_p. \tag{4}$$

$$\pi_x(\lambda) = [p_x(\lambda) - \tau]q_x(\lambda) - f_x = (1 - \tau\lambda^{-1})\frac{EL}{M_c} - f_x, \tag{5}$$

where the quantities demanded by domestic and foreign consumers $q_d(\lambda)$ and $q_x(\lambda)$ are given by (2) and $p_d(\lambda) = p_x(\lambda) = \lambda$.

It is important to emphasize that, as the Melitz model, our approach is isomorphic to a model of process innovations, where firm heterogeneity is derived from productivity differences. The Appendix establishes formally that the latter yields the profit functions described by (4) and (5), an endogenous distribution of markups, and the property that more productive firms charge lower prices and produce more output. In summary, the combination of Cobb-Douglas preferences and limit-pricing strategies generates an endogenous distribution of markups that is missing from models that rely on Dixit and Stiglitz (1977) preferences.[11]

[10] The Appendix shows formally how an augmented version of the consumer utility function (1) can generate the aforementioned optimal limit-pricing rule.

[11] This feature of our model complements the productivity-heterogeneity model proposed by Melitz and Ottaviano (2008) which generates variable average markups based on quasi-linear (as opposed to homothetic) preferences.

We present the main results of our analysis using the quality (as opposed to productivity) version of the model in order to directly relate and compare our findings to the theoretical and empirical literature that focuses on product quality as a determinant of the pattern of intraindustry trade.[12] One caveat of our model is that, under Cobb-Douglas preferences, the revenue per variety is independent of the firm's quality level and price. Another possible caveat (subject to the qualification in footnote 8) is that under limit-pricing each firm charges the same price in all markets independently of per-unit trade costs. These implications are clearly unrealistic. However, as in quality-ladders growth models, Cobb-Douglas preferences deliver surprising tractability and preserve several desirable aggregate properties.

Notice that symmetry of foreign markets implies that the global profit flow generated by exporting equals $n\pi_x(\lambda)$. Because only a fraction of incumbent firms export, a firm producing a good with quality λ earns a per-period profit $\pi(\lambda) = \pi_d(\lambda) + max\{0, n\pi_x(\lambda)\}$. Since each firm faces a constant probability of death δ in each period, the market value of a typical firm is given by

$$\nu(\lambda) = max\left\{0, \sum_{t=0}^{\infty}(1-\delta)^t \pi(\lambda)\right\} = max\left\{0, \frac{1}{\delta}\pi(\lambda)\right\}, \tag{6}$$

where the second equality follows from the fact that each firm's product quality remains constant during its lifetime.

A product of quality λ is produced only if $\pi(\lambda) \geqslant 0$. Therefore, the production cutoff quality level λ_p is determined as follows. Equation (4) implies that $\pi_d(\lambda)$ increases in λ, and that $\pi_d(1) = -f_d < 0$. Thus, for any given values of $f_p > 0$, E, and M_c (which are common to all firms), there exists a production cutoff quality level λ_p, such that all firms producing varieties with quality $\lambda \geqslant \lambda_p$ earn non-negative profits and stay in the market. Firms that know how to produce varieties with quality below the production cutoff quality level λ_p exit the market. Setting (4) equal to zero generates the following *production cutoff*

[12] See, for instance, Schott (2004), Hummels and Skiba (2004), Hummels and Klenow (2005), Hallak (2006), and Verhoogen (2008).

quality level

$$\lambda_p = \frac{1}{1 - M_c f_p / EL} > 1. \tag{7}$$

The production cutoff quality level λ_p depends positively on production fixed costs f_p, and negatively on its market size EL/M_c. This makes sense: in markets with high fixed production costs or low expenditure per variety, only high-quality (high-price) products can earn non-negative profits.

A product with quality λ is exported to all foreign markets, only if $\pi_x(\lambda) \geqslant 0$. Equation (5) implies that, as long as $\lambda > \tau > 1$ holds, $\pi_x(\lambda)$ increases monotonically in λ, and that $\pi_x(\lambda) = -f_x < 0$. Therefore, for any $f_x \geqslant 0$, E, and M_c there exists a cutoff quality level $\lambda_x \geqslant \tau$, such that all firms producing products with quality $\lambda \geqslant \lambda_x$ earn non-negative profits from exporting to any foreign market. However, firms that produce varieties with quality $\lambda < \lambda_x$ face strictly negative foreign profits and do not export. Setting (5) equal to zero yields the following *export cutoff quality level*

$$\lambda_x = \frac{\tau}{1 - M_c f_x / EL} \geqslant \tau > 1. \tag{8}$$

The export cutoff quality level λ_x depends positively on factors (such as variable trade costs τ and foreign-market entry costs f_x) that adversely affect profits from exporting; and negatively on factors (such as foreign market size EL/M_c) that have a positive impact on export profits.

Solving (7) for EL/M_c and substituting the resulting expression in (8) yields

$$\lambda_x = \frac{\tau}{1 - (1 - \lambda_p^{-1})(f_x / f_p)}. \tag{9}$$

Equation (9), together with the restrictions imposed by (7) and (8), establishes the dependence of the exporting cutoff quality level λ_x on the production cutoff quality level λ_p and the model's parameters. The requirement that the denominator of (9) must be non-negative implies that $\lambda_p \in (1, k)$, where $k = 1/[1 - f_p/f_x]$. Inspection of (9) also indicates that λ_x is a monotonically increasing function of the production cutoff quality level $\lambda_p \in (1, k)$, the level of per-unit trade costs τ, and the ratio of overhead fixed costs f_x/f_p. In addition, if $f_x = f_p$, then (9) implies that $\lambda_x = \tau\lambda_p > \lambda_p$ for all $\lambda_p \in (1, \infty)$. Therefore, under the parameter restrictions $f_x \geqslant f_p$ and $\tau > 1$, the exporting cutoff quality level λ_x is strictly greater than the production cutoff quality level λ_p. In this case, firms whose product quality level is less than λ_p exit; firms whose product quality level is $\lambda \in [\lambda_p, \lambda_x)$ produce exclusively for the domestic market because they earn non-negative profits from the domestic operations only; and firms producing high-quality products ($\lambda \geqslant \lambda_x$) sell their products in both domestic and all foreign markets.

A sufficient condition for the partition of firms by export status is that the overhead costs of operating in the domestic market f_p must not exceed the overhead costs of entering a foreign market f_x. A similar condition has been derived by Melitz (2003, p. 1709) for the case of Dixit and Stiglitz (1977) preferences. However, in Melitz's model, no level of trade costs τ can generate the aforementioned partitioning in the absence of exporting fixed costs ($f_x = 0$). In contrast, the present model generates this partitioning due to the limit-pricing behavior of firms. Notice that, in the absence of fixed exporting costs, equation (9) yields $\lambda_x = \tau$. In addition, observe that as the level of production fixed costs approaches zero ($f_p \to 0$) equation (7) yields $\lambda_p \to 1$. By continuity, the present model can deliver the partition of firms by export status under sufficiently high trade costs combined with low production fixed costs. This property leads to a novel prediction: a firm facing foreign markets with sufficiently different trade costs will not export to markets with high trade costs.[13]

This prediction is consistent with the empirical findings of Schott (2004, figure 2) and

[13] This possibility is illustrated with the following simple example. Consider the case of a firm producing a product with quality λ and facing two foreign markets, one with a low per-unit trade costs τ_0 and one with high per-unit trade costs τ_1 such that $1 < \tau_0 < \lambda < \tau_1$. In addition, assume that production and exporting fixed costs are equal to zero ($f_p = f_x = 0$). It is obvious from equations (4) and (5) that this firm earns positive profits in the domestic market and in the low-trade-costs market, but strictly negative profits in the high-trade-costs market and will not export to the latter. Similar considerations apply to the model of productivity heterogeneity and quasi-linear preferences developed by Melitz and Ottaviano (2008).

Baldwin and Harrigan (2007). The former study uses U.S. product level data and reports that in 1994 about 90 percent of ten-digit HS product categories and almost 80 percent of 4-digit SITC categories exhibited zero imports! The second study argues that trade costs are positively correlated with the absence of U.S. exports in narrowly defined product categories. Models of firm heterogeneity that rely on Dixit and Stiglitz (1977) preferences cannot generate this prediction, thanks to constant markup pricing across all varieties.

The following proposition summarizes the aforementioned analysis.

Proposition 1. *Let λ_p and λ_x denote the production and export cutoff quality levels, respectively; and let $k = 1/[1 - f_p/f_x]$ denote the upper bound of the production cutoff quality level λ_p. Then,*

a. *The export cutoff quality level λ_x is an increasing function of the production cutoff quality λ_p.*

b. *If the production fixed cost does not exceed the foreign-market entry cost (i.e., $f_p \leqslant f_x$), then the export cutoff quality level is strictly greater than the production cutoff quality level (i.e., $\lambda_x > \lambda_p > 1$).*

c. *In the absence of foreign market entry cost (i.e., $f_x = 0$), there exists a level of trade cost such that the exporting cutoff quality level is strictly greater than the production cutoff quality level (i.e., $\lambda_x = \tau > \lambda_p > 1$).*

Proposition 1 implies that firms with high-quality products charge higher prices, enjoy higher profits and ship these products abroad in accordance to the Alchian and Allen (1964) conjecture, which has been confirmed empirically by Hammels and Skiba (2004). Our paper develops a general equilibrium model in which trade costs and self-selection among heterogeneous firms lead to the desired result.

2.3 Entry Decision

The determination of the production cutoff quality level depends on entry and exit considerations. We assume that there is a large number of prospective and ex-ante identical entrants. Each entrant faces a fixed entry cost $f_e > 0$, which is measured in units of labor and interpreted as the number of R&D researchers employed by the entrant to discover a new variety. After a firm incurs the fixed entry cost, it draws its quality parameter λ from a common and known distribution $g(\lambda)$ with positive support over $(0, \infty)$ and with continuous cumulative distribution $G(\lambda)$. The properties of $g(\lambda)$ determine the benefits of entry measured by the relevant expected discounted profits.

The ex-ante probability of drawing a quality level λ is governed by the density function $g(\lambda)$ and the ex-ante probability of successful entry $1 - G(\lambda_p)$. As in Melitz (2003), the ex-post distribution of product quality levels μ is the conditional distribution of $g(\lambda)$ on the interval $[\lambda_p, \infty)$:

$$\mu(\lambda) = \begin{cases} \frac{g(\lambda)}{1-G(\lambda_p)} & \text{if } \lambda > \lambda_p \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The ex-ante probability that an incumbent firm will export is given by

$$\zeta_x = \frac{1 - G(\lambda_x)}{1 - G(\lambda_p)}. \tag{11}$$

In addition, the law of large numbers implies that ζ_x equals the ex-post fraction of incumbent firms that export.

Let M_p denote the mass of varieties (and firms) produced in any country and M_x be the number of varieties that each country exports. Then, we have $M_x = \zeta_x M_p$, which further ensures that $M_c = (1 + n\zeta_x)M_p$ is the the mass of products available for consumption in any country. Armed with the aforementioned probability distributions, one can calculate the aggregate quantities demanded:

$$Q_c = \int_{\lambda_p}^{\infty} q(\lambda) M_p \mu(\lambda) d\lambda + n \int_{\lambda_x}^{\infty} q(\lambda) M_p \mu(\lambda) d\lambda = \frac{M_p}{M_c} \frac{EL}{\tilde{\lambda}_p} + \frac{nM_x}{M_c} \frac{EL}{\tilde{\lambda}_x}, \tag{12}$$

where $\lambda_i \quad (i = p, x)$ is given by

$$\tilde{\lambda}_i \equiv \tilde{\lambda}(\lambda_i) = \left[\frac{1}{1 - G(\lambda_i)} \int_{\lambda_i}^{\infty} \lambda^{-1} g(\lambda) d\lambda \right]^{-1}. \tag{13}$$

That is, $\tilde{\lambda}_p$ is the weighted harmonic mean of the quality levels (and prices) of all produced goods and can be interpreted as the average (expected) quality level. Similarly, $\tilde{\lambda}_x$ is the weighted harmonic mean of the quality levels of a country's exports and can be interpreted as the average export quality.

Because the production cutoff quality level is the minimum quality level of all surviving products, it must be lower than the average quality level. Moreover, an increase in λ_p forces producers with low-quality products to exit the market, which in turn increases the average quality level of all produced varieties. The same intuition applies to the relationship between the export quality cutoff level and the average quality of exports. The following lemma summarizes these properties (see Appendix for proof).

Lemma 1. *The average quality level of all products produced in a typical market is strictly greater and increases in the production cutoff quality level, i.e.,* $\tilde{\lambda}_p > \lambda_p$ *and* $\partial \tilde{\lambda}_p / \partial \lambda_p > 0$. *The average quality level of exports is strictly greater and increases in the export cutoff quality level, i.e.,* $\tilde{\lambda}_x > \lambda_x$ *and* $\partial \tilde{\lambda}_x / \partial \lambda_x > 0$.

The ex-ante quantity demanded for each variety is obtained by dividing (12) by M_p and is given by

$$\bar{q} = \tilde{\lambda}_p^{-1} EL/M_c + n\zeta_x \tilde{\lambda}_x^{-1} EL/M_c, \tag{14}$$

where first and second terms correspond to the domestic and foreign demand for a typical variety, respectively. Consequently, the ex-ante per-period profit of a typical prospective entrant is given by

$$\bar{\pi} = \pi_d(\tilde{\lambda}_p) + n\zeta_x \pi_x(\tilde{\lambda}_x) = (1 - \tilde{\lambda}_p^{-1}) \frac{EL}{M_c} - f_p + n\zeta_x \left[(1 - \tau \tilde{\lambda}_x^{-1}) \frac{EL}{M_c} - f_x \right]. \tag{15}$$

Because the probability of successful entry is $1 - G(\lambda_p)$, the net benefits of entering the domestic market are equal to the expected value of a firm $[1 - G(\lambda_p)]\bar{\nu}$, where $\bar{\nu} = \bar{\pi}/\delta$ is the ex-ante value of a prospective entrant. Setting the benefits of entry equal to the fixed R&D costs yields the free-entry condition

$$[1 - G(\lambda_p)]\frac{\bar{\pi}}{\delta} = f_e, \tag{16}$$

where $\bar{\pi}$ is defined by (15). This concludes the description of the model.

2.4 Steady-State Equilibrium

This section determines the market cutoff quality levels, number of varieties, the average markups, and the welfare level in a typical country. Substituting (9) and (11) in (15) yields an expression for ex-ante profits that depends only on the two cutoff quality levels. Further substitution of $\bar{\pi}$ into the free-entry condition (16) yields the basic steady-state equilibrium condition

$$H(\lambda_p, 1) + n\left(\frac{f_x}{f_p}\right)H(\lambda_x, \tau) = \delta\frac{f_e}{f_p}, \tag{17}$$

where H is defined as

$$H(\lambda_i, \alpha) \equiv [1 - G(\lambda_i)]\left[\frac{1 - \alpha\tilde{\lambda}_i^{-1}}{1 - \alpha\lambda_i^{-1}} - 1\right], \tag{18}$$

where $(i, \alpha) \in \{(p, 1), (x, \tau)\}$.

The export cutoff quality level λ_x is an increasing function of λ_p (see equation (9)). In addition, the Appendix proves that H is strictly decreasing in cutoff levels and that the left-hand-side of equation (17) is also decreasing in $\lambda_p \in (1, k)$. The above considerations imply the following result (see Appendix for proof).

Proposition 2. *Let $k = 1/(1 - f_p/f_x)$ and assume that $f_p < f_x$. There exist a unique production and a unique export cutoff quality levels $\lambda_p \in (1, k)$ and $\lambda_x \in (\tau, \infty)$ which satisfy equations (9) and (17) such that $\lambda_p > 1$, $\lambda_x > \tau$, and $\lambda_x > \lambda_p$.*

Once the two cutoff quality levels are determined, one can solve for the values of the remaining endogenous variables. We start with the determination of the mass of products produced in each market. In the steady-state equilibrium, the per-period flow of successful entrants must be equal to flow of incumbents who exit the market because they are hit by a bad shock, i.e., $[1 - G(\lambda_p)]M_e = \delta M_p$, where M_e is the mass of all (as opposed to successful) entrants. Then the aggregate amount of labor employed by prospective entrants is $L_e = M_e f_e = \delta M_p f_e/[1 - G(\lambda_p)] = M_p\bar{\pi}$, where the last equality follows from the free-entry condition (16). Thus the aggregate amount of labor devoted to R&D equals the level of aggregate profits earned by all producers in a typical market.

The aggregate demand for labor in a typical market equals the aggregate supply of labor $L_p + L_e = L_p + \Pi = L$, where L_p denotes the total amount of labor employed in the production of surviving goods and Π is the level of aggregate profits earned by all producers. In addition, the standard GDP identity implies that the total wage bill must be equal to the aggregate expenditure on all goods produced $wL_p + wL_e = EL$. Therefore, per-capita expenditure equals unity due to the choice of labor as the numeraire, i.e., $E = w = 1$. Substituting $E = 1$ in (7) yields the mass of products available for consumption M_c

$$M_c = (1 - \lambda_p^{-1})\frac{L}{f_p}. \tag{19}$$

Substituting[14] the relationship $M_c = (1 + n\zeta_x)M_p$ into (19) yields the mass of varieties produced in each country

$$M_p = \left[\frac{1 - \lambda_p^{-1}}{1 + n\zeta_x}\right]\frac{L}{f_p}. \tag{20}$$

[14]We can now derive a sufficient condition which guarantees that the consumer consumes all available varieties. In principle, each consumer chooses $q(\omega)$ (the quantity of each variety) and M (the number of varieties) to maximize (1) subject to the budget constraint. The first order condition with respect to M yields $\ln[\beta\lambda(M)q(M)/L] \geqslant 1$, which holds with equality if $M < M_c$. Inserting (2) into this condition yields $\ln[\beta\lambda(M)E/[p(M)M]] \geqslant 1$. Thus to ensure that the principle of non-satiation holds (i.e., $M = M_c$), we must have that $\ln[\beta\lambda(M_c)E/[p(M_c)M_c]] > 1$. Since $p(M_c) = \lambda(M_c)$ and $E = 1$, we must have $\ln(\beta/M_c) > 1$, that is $\beta > eM_c$, where e is the base of the natural logarithm. Using (19) implies that $\beta f_p/eL > 1 - \lambda_p^{-1}$. Moreover, because $\lambda_p \in (1, k)$, the inequality condition holds if $\beta f_p/eL > 1 - k^{-1}$, which in turn yields $\beta > eL/f_x$.

Notice that in the absence of trade ($n = 0$) the mass of varieties produced equals the mass of varieties consumed, i.e., $M_c = M_p$.

As we mentioned in the introduction, the model generates an endogenous distribution of markups. Each incumbent firm charges a price equal to its quality level $p(\lambda) = \lambda$, and therefore its markup measured by the price marginal-cost margin as given by $(p - 1)/p = 1 - \lambda^{-1}$. Subsequently, the aggregate markup over all incumbents equals $PCM_p = \int_{\lambda_p}^{\infty}(1 - \lambda^{-1})M_p\mu(\lambda)d\lambda = M_p(1 - \tilde{\lambda}_p^{-1})$. Similarly, an exporter charges a price $p(\lambda) = \lambda$, incurs a marginal cost τ, and earns a price marginal-cost margin $(p - \tau)/p = 1 - \tau\lambda^{-1}$. Thus the aggregate export markup is $PCM_x = \int_{\lambda_x}^{\infty}(1 - \tau\lambda^{-1})M_p\mu(\lambda)d\lambda = M_x(1 - \tau\tilde{\lambda}_x^{-1})$. Consequently, the average production (domestic) and export markups are given by

$$pcm_p = 1 - \tilde{\lambda}_p^{-1} \qquad \text{and} \qquad pcm_x = 1 - \tau\tilde{\lambda}_x^{-1}, \tag{21}$$

where $\tilde{\lambda}_p$ and $\tilde{\lambda}_x$ are defined by (13). Both average markups increase in the production cutoff quality level λ_p (see Proposition 1), but the average export markup decreases in the level of trade costs τ.

One can obtain an expression for per capita welfare as follows. Substituting the per-capita quantity demanded $q(\lambda)/L = E/\lambda M_c$ for each variety into the utility function of a typical consumer (1) and performing the integration yields

$$U = M_c \ln\left(\frac{\beta E}{M_c}\right) = M_c \ln\left(\frac{\beta}{M_c}\right), \tag{22}$$

where per-capita expenditure E is set equal to unity due to the choice of labor as the numeraire. Since $\ln(\beta/M_c) > 1$ (see footnote 14), U is always positive. Observe that per-capita welfare depends positively on the mass of varieties consumed M_c and on the expenditure per variety E/M_c.

To unveil the intuition for the welfare expression notice that an increase in M_c has two conflicting welfare effects. First, each consumer becomes better off because she consumes more products. Second, she becomes worse off as her expenditure, which is equal to her wage, spreads among more varieties. Limit pricing renders the consumer indifferent between receiving q units of a generic product with quality 1 and q/λ units of a good with quality λ at a higher price $p(\lambda) = \lambda$, and thus the average quality does not appear directly as an argument in the welfare function. Higher average quality allows firms to reduce the demand for manufacturing labor by charging a higher price and producing less quantity per variety. This, in turn, means that the economy can afford the production of more varieties and enjoy a higher welfare level.

3 The Impact of International Trade

The model is well suited to analyze the general equilibrium effects of trade liberalization measured by an increase in the number of trading partners n, a reduction in per-unit trade costs τ, and a reduction in exporting overhead costs f_x. These parameters capture a variety of forces including reductions in transportation and communication costs, reductions in trade barriers, and the formation of trading blocks (albeit in a highly stylized fashion given the assumption of structurally identical countries). The impact of trade liberalization is channeled through two interacting general-equilibrium channels: changes in the demand for labor, which are captured by changes in the real wage; and changes in the intensity of product-market competition, which are captured by changes in the average markup.

Formally, the effects of trade liberalization are transmitted through changes in the production cutoff quality level λ_p as are described in Lemma 2 (see Appendix for proof).

Lemma 2. *Trade liberalization, captured by an increase in the number of trading partners ($n \uparrow$), a reduction in per-unit transport costs ($\tau \downarrow$), or a reduction in foreign-market entry costs ($f_x \downarrow$), increases the production cutoff quality level λ_p (i.e., $\partial\lambda_p/\partial n > 0$, $\partial\lambda_p/\partial\tau < 0$, and $\partial\lambda_p/\partial f_x < 0$).*

The economic intuition behind Lemma 2 is as follows. For any initial value of the production cutoff quality level λ_p, the export cutoff quality level λ_x and the mass of varieties consumed M_c are fixed (see equations (9) and (19)). Equation (9) and Lemma 1 imply that a decline in τ or f_x increases the average quality of exports $\tilde{\lambda}_x$. Therefore, any form of trade liberalization ($n \uparrow$, $\tau \downarrow$, or $f_x \downarrow$) increases the ex-ante profits $\bar{\pi}$ (see equation (15) and raises the demand for labor for any wage level (see equation (16) and, in particular, equation (17)). The excess demand for labor induces a reallocation of resources from low-quality products towards high-quality products that translates into a larger mass of products available for consumption M_c. To see this, recall that for any level of expenditure, a firm with a higher quality product charges a higher price, produces less output, and employs less labor than a firm with a lower quality product (see equation (3)). Thus any given aggregate supply of labor can sustain more higher-quality products.

The reallocation of resources from lower to higher-quality products increases the demand for varieties and intensifies the product market competition by reducing the demand for each product as the aggregate expenditure EL is spread among more varieties. Consequently, the flow of profits of the marginal firm, which produces a product with the cutoff quality level λ_p, become negative, and induce an increase in the production cutoff quality level to restore the zero-profit condition (7) that determines λ_p. An increase in the production cutoff quality level λ_p, caused by trade liberalization, generates quality upgrading (measured by an increase in the average quality) and increases the average markup by shifting resources from low to high-quality products. Consequently, in the present model, trade liberalization operates primarily through the increased intensity of product-market competition and the associated increase in the mass of sustainable products (as opposed to a reduction in markups).

The next step of the analysis is to establish the impact of trade liberalization on markups, on the number of varieties consumed, and on welfare. First, consider the impact of trade liberalization on the intensity of product-market competition captured by the domestic and export markups defined by (21). Any type of trade liberalization increases λ_p and λ_x (from Lemmas 1 and 2) and increases both domestic and export markups (see equation (21)). In addition, inspection of equation (19) and Lemma 2 establish that any form of trade liberalization increases the production cutoff quality level λ_p and the mass of varieties available for consumption M_c in every country. Finally, differentiating equation (22) with respect to varieties consumed yields $\partial U/\partial M_c = \ln(\beta/M_c) - 1 > 0$, where the inequality directly follows from our assumption on β (see footnote 14). Thus, the effect of trade liberalization on welfare is positive. We have established:

Proposition 3. *Trade liberalization, captured by an increase in the number of trading partners ($n \uparrow$), a reduction in per-unit transport costs ($\tau \downarrow$), or a reduction in foreign-market entry costs ($f_x \downarrow$):*

a. increases the average domestic and export markups ($pcm_p \uparrow$, $pcm_x \uparrow$);

b. raises the mass of products available for consumption ($M_c \uparrow$);

c. and has a positive effect on national and global welfare ($U \uparrow$).

Next consider an extreme form of trade liberalization: the move from autarky to (restricted) trade. The closed-economy steady-state equilibrium corresponds to the case of no trading partners (i.e., $n = 0$). Equation (17) then implies that, under autarky, the production cutoff quality level λ_p^A is determined by $H(\lambda_p, 1) = \delta f_e/f_p$, and is strictly less than the open-economy cutoff quality level λ_p : the absence of export markets reduces the benefits of entry and shifts labor from the production of higher-quality products towards the production of lower-quality products. Observe that equations (19), (21), and (22) determine the closed-economy values the number of varieties consumed ($M_p^A = M_c^A = [1 - (\lambda_p^A)^{-1}]L/f_p$), the average markups ($pcm_p^A = 1 - (\tilde{\lambda}_p^A)^{-1}$), and the level of welfare ($U^A = M_c^A \ln(\beta/M_c^A)$) which are all functions of the production cutoff quality level. Therefore, a move from autarky to trade has the same qualitative impact as an increase in the number of trading partners n. These effects are summarized in the following proposition.

Proposition 4. *A move from autarky to trade*

a. increases the production cutoff quality level ($\lambda_p > \lambda_p^A$);

b. generates higher markups ($pcm_p > pcm_p^A$);

c. raises the mass of products available for consumption ($M_c > M_c^A$);

d. and has a positive effect on national and global welfare ($U > U^A$).

As in Melitz's (2003) analysis, trade liberalization induces entry of better firms into foreign markets, forces firms with low-quality products to exit, and expands the number of varieties consumed. The model's prediction that trade liberalization increases industry markups in markets with Bertrand competition and vertical product differentiation is based on a novel mechanism that complements the work of Melitz and Ottaviano (2008). In their model, trade liberalization intensifies the product market competition via a change in the price elasticity of demand, whereas in the present model trade expands the mass of varieties and reduces the income spent on each product without affecting the price elasticity of demand. In Melitz and Ottaviano (2008) trade reduces average markups, whereas in the present model, trade generates quality upgrading and higher average markups. The next section explores formally this important issue by analyzing the model's welfare properties.

4 Welfare Properties

The presence of quality-heterogeneous firms raises the following welfare question: does the market provide the socially optimal quality cutoff levels, markups, and the mass of available varieties? The constrained optimality (in the Dixit and Stiglitz (1977) sense, where firms make non negative profits) of the cutoff productivity levels has been demonstrated in a multi-sector version of the Melitz model by Feenstra and Kee (2008). This is a surprising result considering the presence of potential welfare distortions caused by imperfect competition, trade costs, and heterogeneous productivity levels.

Contrary to the existing literature on trade with heterogeneous firms, the present model generates the possibility of divergence between the socially optimal and laissez-faire equilibrium due to the endogenous distribution of markups. In order to minimize the algebra, we illustrate this possibility in a closed-economy setting noting that similar considerations apply to an open economy, where trade costs create additional welfare distortions. The social planner maximizes per-capita utility function $U = M_c \ln(\beta E/M_c)$ subject to the resource constraint. Since labor is the only factor of production, the resource constraint requires that at each instant in time labor is fully employed. The aggregate supply of labor L is fixed, whereas the aggregate demand for labor consists of two components: labor employed by potential entrants $L_e = \delta M_c f_e/[1 - G(\lambda_p)]$, where the mass of varieties produced equals the mass of varieties consumed ($M_p = M_c$); and labor employed in production. The latter is derived as follows. Since one unit of labor produces one unit of output, each surviving firm hires $f_p + q(\lambda)$ workers, where f_p is the production fixed cost expressed in units of labor and $q(\lambda)$ is the demand for a product of quality λ given by (2). Thus the labor employed in production is $L_p = \int_{\lambda_p}^{\infty} [f_p + q(\lambda)] M_c \mu(\lambda) d\lambda = f_p M_c + \tilde{\lambda}_p^{-1} EL$. Consequently the resource constraint is

$$\frac{\delta M_c f_e}{1 - G(\lambda_p)} + f_p M_c + \tilde{\lambda}_p^{-1} EL = L. \tag{23}$$

The social planner maximizes $U = M_c \ln(\beta E/M_c)$ subject to (23) with respect to the production cutoff quality level λ_p, per-capita expenditure E, and the mass of varieties available for consumption M_c. Denoting with ψ the Lagrangian multiplier, one can write the corresponding first-order conditions as follows.

$$(\lambda_p^{-1} - \tilde{\lambda}_p^{-1})(E/M_c)L = \delta f_e/[1 - G(\lambda_p)], \tag{24}$$

$$1/\psi = \tilde{\lambda}_p^{-1}(E/M_c)L, \tag{25}$$

$$(1/\psi)[\ln(\beta E/M_c) - 1] - f_p = \delta f_e/[1 - G(\lambda_p)]. \tag{26}$$

These conditions determine the socially optimum values of λ_p, E, and M_c. Equation (25) implies that the Lagrangian multiplier must be positive ($\psi > 0$).

We can obtain further insights on the solution to the social planner's problem by substituting (25) into (26) and equating the resulting expression to (24) to obtain

$$\left[\tilde{\lambda}_p^{-1} \ln(\beta E/M_c) - \lambda_p^{-1}\right](E/M_c)L = f_p. \tag{27}$$

Since the average quality level $\tilde{\lambda}_p$ is a function of the cutoff quality level λ_p, equations (24) and (27) determine the socially optimal values of λ_p^S and E^S/M_c^S. Once these two endogenous variables are determined, the resource constraint (23), or the assumption $E^s = 1$, can be used to obtain individual values for E^S and M_c^S.

How does the socially optimum solution compare to the closed-economy market equilibrium? Set the number of trading partners n equal to zero in (15) and substitute the resulting expression for $\bar{\pi}$ and f_p from condition (7) into (16) to obtain the closed-economy free-entry condition which is identical to (24). In other words, for any level of E/M_c the laissez-faire cutoff quality level λ_p^A is socially efficient. This result, which is identical to the ones obtain by Feenstra and Kee (2008), holds because the cutoff quality level does not appear as an argument in the social planner's objective function, but only in the resource constraint. Thus, the social planner chooses the cutoff quality level based on efficient resource allocation considerations. In the case of Dixit and Stiglitz (1977) preferences and exogenous markups, the welfare distortions associated with a marginal increase in the mass of varieties just happen to cancel each other out. However, this is not the case in the present model. To see this, note that the market equilibrium values λ_p^A and E^A/M_c^A are simultaneously determined by the entry condition (24) and the zero-profit condition (7). The latter is reproduced below for illustrative purposes

$$[1 - \lambda_p^{-1}](E/M_c)L = f_p. \tag{28}$$

Comparing the zero-profit condition (28) to the socially-optimal condition (27) reveals the following. The market cutoff quality level depends on profitability considerations that are captured by the marginal markup $1 - \lambda_p^{-1}$. This differs from the efficient "markup" $\tilde{\lambda}_p^{-1} \ln(\beta E/M_c) - \lambda_p^{-1}$ in (27) which depends on the ratio between the marginal utility derived from an additional variety and the average quality of available products. Observe that the social planner and the market assign the same total cost to the product with the cutoff quality level, namely $f_p + \lambda_p^{-1}(E/M_c)L$. However, there is a divergence between the benefits associated with the introduction of a new variety. The market cares about the total revenue earned $\lambda_p[EL/(M_c\lambda_p)] = EL/M_c$, where the term in square brackets equals the quantity demanded and λ_p is the corresponding limit price. In contrast, equation (27) reveals that the social benefits of introducing an additional product equal $\{\ln(\beta E/M_c)\}[EL/M_c\tilde{\lambda}_p]$, where the term in curly brackets is the change in consumer surplus (welfare) associated with the introduction of a new variety (for any given expenditure per variety E/M_c), and the term in square brackets represents the quantity demanded for a product with average (as opposed to marginal) quality.

In other words, the planner cares about the average (infra-marginal) consumer, whereas the market cares about the marginal consumer. On one hand, because the marginal quality level is less than the average quality level ($\lambda_p < \tilde{\lambda}_p$), the market has a tendency to overstate the benefits of introducing a new variety by charging a lower "price" and producing a higher quantity than the social planner. On the other hand, because $\ln(\beta E/M_c) > 1$ and $\lambda_p > 1$ the ranking between the social and market prices is in general ambiguous. Consequently, there is a divergence between the market and social valuation of the marginal benefits associated with an introduction of an additional variety, and inspection of equations (23), (24), (27) and (28) yields the following proposition.

Proposition 5. *The laissez-faire cutoff quality level λ_p^A, mass of varieties M_c^A, and per-capita expenditure E^A are socially sub-optimal.*

How does the market solution differ from the socially optimum solution? Figure 1 illustrates the market and efficient values of λ_p and E/M_c by plotting the graphs of equations (24), (27), and (28) under the assumption that quality levels are drawn from a Pareto distribution with scale parameter b and shape parameter $\kappa > 0$,

$$G(\lambda) = 1 - \left(\frac{b}{\lambda}\right)^{\kappa} \qquad \text{for } \lambda \geqslant b > 0. \tag{29}$$

Assuming that quality levels follow a Pareto distribution is not essential in our analysis. However, it makes the analysis analytically more tractable. For example, using this distribution function in (13) yields $\tilde{\lambda}_i/\lambda_i = (\kappa+1)/\kappa$, for $i = p, x$.

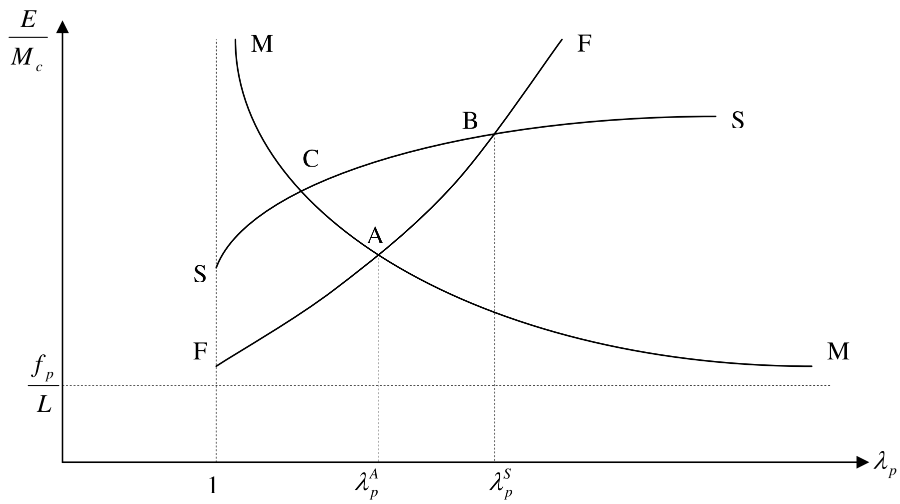

FIGURE 1. Efficient and Market Cutoff Quality Levels

Equation (24) defines a direct relationship between E/M_c and λ_p which is illustrated with the convex FF curve in Figure 1.[15] Equation (28) defines an inverse relationship between E/M_c and λ_p, which is illustrated with the negatively-sloped MM curve.[16] The unique intersection between the FF and MM curves at point A determines the market cutoff quality level λ_p^A. The graph of equation (27) is illustrated with the concave SS curve.[17] More interestingly, at any intersection point of the FF and SS curves, the slope of the FF curve is greater than that of the SS curve.[18] Thus, the most the FF and SS curves can intersect each other is once.[19] We assume that the parameters of the model are such that the FF

[15]With the above cumulative distribution function, equation (24) becomes $E/M_c = (\kappa+1)\delta f_e \lambda_p^{\kappa+1}/(Lb^{\kappa})$.

[16]Equation (28) can be written as $E/M_c = f_p/[(1-\lambda_p^{-1})L]$, which implies that E/M_c is decreasing in λ_p, $E/M_c \to \infty$ as $\lambda_p \to 1$, and $E/M_c \to f_p/L$ as $\lambda_p \to \infty$.

[17]With the Pareto distribution defined by (29), equation (27) becomes $[\kappa \ln(\beta E/M_c) - (\kappa+1)](E/M_c) = (\kappa+1)\lambda_p f_p/L$. Totally differentiating this equation yields $d(E/M_c)/d\lambda_p = (\kappa+1) f_p [L\kappa \ln(\beta E/M_c) - L]^{-1} > 0$, where the last inequality follows from the facts that $\kappa \ln(\beta E/M_c) - (\kappa+1) > 0$ and $\kappa > 0$. Differentiation of this slope with respect to λ_p implies that $d^2(E/M_c)/d\lambda_p^2 = -\kappa\,[d(E/M_c)d\lambda_p]^2\,\{(E/M_c)[\kappa \ln(\beta E/M_c) - 1)]\}^{-1} < 0$.

[18]To see this, notice that the slope of the FF curve is given by $d(E/M_c)/d\lambda_p = (\kappa+1)(E/M_c)/\lambda_p$. The slope of the SS curve, on the other hand, can be rewritten as $d(E/M_c)/d\lambda_p = (\kappa+1)(f_p/L)(E/M_c)[(\kappa+1)(f_p/L)\lambda_p + \kappa(E/M_c)]^{-1}$. It then easily follows that at any intersection $(\lambda_p, E/M_c)$, $(\kappa+1)(E/M_c)/\lambda_p > (\kappa+1)(E/M_c)(f_p/L)/[(\kappa+1)(f_p/L)\lambda_p + \kappa(E/M_c)]$.

[19]The slope of the SS curve is generally indeterminate under other distribution functions. In this case, the SS curve may intersect the FF curve more than once, and hence, might give rise to *multiple* socially optimum cutoff quality levels.

and SS curves intersect at point B, at which the socially optimum cutoff level is greater than the market cutoff level, i.e., $\lambda_p^S > \lambda_p^A$.[20]

Figure 1 can be used to perform more comparative statics exercises. An increase in production fixed cost f_p does not affect the position of the FF curve, shifts the MM curve to the right raising λ_p^A and E^A/M_c^A; while it shifts the SS curve to the left raising λ_p^S and E^S/M_c^S. Intuitively, an economy facing higher production fixed costs allocates a larger share of resources to the creation of more varieties than to the production of low-quality products. An increase in market size, measured by the labor endowment L, shifts the FF and SS curves to the right and the MM curve to the left (not shown in Figure 1). The magnitude of the FF curve shift equals the magnitude of the MM shift and exceeds the corresponding magnitude of the SS curve shift.[21] Therefore, an increase in market size L lowers E/M_c, raises the efficient cutoff quality level λ_p^S, and does not affect the market cutoff quality level λ_p^A.

5 Conclusion

The present paper developed a highly-tractable model of quality heterogeneity and international trade. Firms in our model face Cobb-Douglas preferences and charge prices that are proportional to the quality level of their products as in Schumpeterian growth models. The production side of the model mimics the model of productivity heterogeneity proposed by Melitz (2003). In our model, firms with high-quality products export, firms with intermediate-quality products produce for the domestic market, and firms with low-quality

[20] Equation (27) implies that $\ln(\beta E/M_c) > (\kappa+1)/\kappa$, i.e., $E/M_c > \exp(1+1/\kappa)/\beta$. At $\lambda_p = 1$, equation (24) yields $E/M_c = (\kappa+1)\delta(f_e/L)b^{-\kappa}$. Given that $\delta < 1$ and $f_e/L < 1$, unless b is too low and κ is too high, as λ_p approaches 1 the SS curve intercept is higher than the FF curve intercept. In depicting Figure 1, we assume that these conditions are satisfied. Also notice that when k increases the FF curve becomes steeper (assuming that $b \leqslant 1$). In this case, the FF curve may intersect the SS curve at a point to the left of point C in Figure 1, which implies that the socially optimum cutoff level is smaller than the market cutoff level λ_p^A.

[21] Totally differentiating equations (24), (28) and (27) yields the marginal shifts of the FF curve $d(E/M_c)/dL = -(E/M_c)/L$, the MM curve $d(E/M_c)/dL = -(E/M_c)/L$, and the SS curve $d(E/M_c)/dL = -[(E/M_c)/L]\left\{[\tilde{\lambda}_p^{-1}\ln(\beta E/M_c) - \lambda_p^{-1}]/[\tilde{\lambda}_p^{-1} + \tilde{\lambda}_p^{-1}\ln(\beta E/M_c) - \lambda_p^{-1}]\right\}$. The term in curly brackets is positive, less than one, and implies that the magnitude of the SS curve shift is less than that of the FF curve.

products exit the market. This trade pattern is consistent with the findings of several empirical studies and the Alchian-Allen conjecture. Firms producing high-quality products charge high prices and enjoy high markups, whereas firms producing low-quality products charge low prices and charge low markups.

The model generates several novel results that complement those of the existing theory of monopolistic competition with heterogeneous firms. First, the distribution of markups, the mass of varieties consumed and produced, the real wage, and cutoff quality levels are all endogenous and determined by general equilibrium forces. Therefore, the model delivers a general equilibrium mechanism, based on endogenous markups and endogenous factor prices, which governs the impact of trade on intra-industry reallocation of resources. Second, trade liberalization (or a move from autarky to trade) raises the intensity of product market competition, increases the cutoff quality levels, increases the average markups and forces inefficient firms producing lower-quality products to exit the market. The welfare impact of trade liberalization (or a move from autarky to trade) is positive. However, the laissez-faire cutoff quality levels are socially suboptimal. The main reason behind this inefficiency is the ambiguous welfare ranking between the market and socially optimal markups. This, in turn, leaves room for welfare-improving government intervention.

References

Alchian, Armen A. and William R. Allen, *University Economics*, Belmont, CA: Wadsworth Pub. Co., 1964.

Aw, Bee Yan, Sukkyun Chung, and Mark J. Roberts, "Productivity and Turnover in the Export Market: Micro Evidence from the Republic of Korea and Taiwan (China)," *The World Bank Economic Review*, 2000, *14*, 65–90.

Baldwin, Richard E. and Frederic Robert-Nicoud, "Trade and Growth with Heterogeneous Firms," *Journal of International Economics*, 2008, *74*, 21–34.

——— and James Harrigan, "Zeros, Quality and Space: Trade Theory and Trade Evidence," 2007. NBER Working paper, No:13214.

Bernard, Andrew B. and J. Bradford Jensen, "Exceptional Exporter Performance: Cause, Effect, or Both," *Journal of International Economics*, 1999, *47*, 1–25.

———, Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in international Trade," *American Economic Review*, September 2003, *93*, 1268–90.

Clerides, Sofronis K., Saul Lach, and James R. Tybout, "Is Learning by Exporting Important? Micro-Dynamic Evidence from Columbia, Mexico, and Morocco," *Quarterly Journal of Economics*, 1998, *113*, 903–47.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 1977, *67*, 297–308.

Feenstra, Robert C. and Hiau L. Kee, "Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity," *Journal of International Economics*, 2008, *74*, 500–18.

Grossman, Gene M. and Elhanan Helpman, "Quality Leaders in the Theory of Growth," *Review of Economic Studies*, January 1991, *58*, 43–61.

Gustafsson, Peter and Paul Segerstrom, "Trade Liberalization and Productivity Growth," 2007. Mimeo, Stockholm School of Economics.

Hallak, Juan C., "Product Quality and the Direction of Trade," *Journal of International Economics*, 2006, *68*, 238–65.

Harrison, Ann, "Productivity, Imperfect Competition, and Trade Reform: Theory and Evidence," *Journal of International Economics*, 1994, *36*, 53–73.

Haruyama, Tetsugen and Laixun Zhao, "Trade and Firm Heterogeneity in A Quality-Ladders Model of Growth," 2008. Mimeo, Kobe University.

Helpman, Elhanan, "Trade, FDI, and the Organization of Firms," *Journal of Economic Literature*, 2006, *44*, 589–630.

Hummels, David and Alexandre Skiba, "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 2004, *112*, 1384–1402.

___ and Peter J. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, 2005, *95*, 704–23.

Maurice, Kugler and Eric Verhoogen, "The Quality Complementarity Hypothesis: Theory and Evidence from Columbia," 2008. NBER Working Paper, 14418.

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

___ and Gianmarco I. Ottaviano, "Market Size, Trade, and Productivity ," *Review of Economic Studies*, 2008, *75*, 295–316.

Schott, Peter K., "Across-Product Versus Within Product Specialization in International Trade," *Quarterly Journal of Economics*, 2004, *119*, 647–78.

Segerstrom, Paul, T. C. A. Anant, and Elias Dinopoulos, "A Schumpeterian Model of the Product Life Cycle," *American Economic Review*, 1990, *80*, 1077–92.

Taylor, Scott M., "Quality Ladders and Ricardian Trade," *Journal of International Economics*, 1993, *34*, 225–43.

Tybout, James R., "Plant- and Firm-level Evidence on the 'New' Trade Theories," in E. Kwan Choi and James Harrigan, eds., *Handbook of International Trade*, Oxford: Basil-Blackwell, 2003.

Unel, Bulent, "Technology Diffusion Through Trade with Heterogeneous Firms," 2008. *Review of International Economics*, Forthcoming.

Verhoogen, Eric A., "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 2008, *123*, 489–530.

Appendix

A. Limit Pricing

Consider the following utility function, which is a generic version of the preferences used routinely in quality-ladders growth models, that describes the tastes of a typical consumer

$$U = \int_{\omega \in \Omega} \ln \left[\beta \frac{q_0}{L} + \beta \lambda(\omega) \frac{q(\omega)}{L} \right] d\omega,$$

where Ω denotes the set of *potential* varieties. Each variety is associated with two quality levels: $q(\lambda(\omega))$ that can be produced by only one firm as in our model, and the low-quality (generic) version q_0 that can be produced by a competitive fringe after the original product has been produced. We assume that the low-quality version can be produced under constant marginal costs equal to the wage and perfect competition. This implies that the generic product commands a price $p_0 = w = 1$. In addition, since both goods adjusted for quality are identical, a consumer spends her income only on the good with the lower quality adjusted price. Thus, if a consumer buys only the generic good q_0, she obtains a utility level equal to $\ln(\beta E/[p_0 M_c L]) = \ln(\beta E/[M_c L])$, whereas if she buys the high-quality good she obtains a utility level equal to $\ln(\beta \lambda E/[p(\lambda) M_c L])$.

Therefore the consumer is indifferent between the two quality versions of the product if and only if $p(\lambda) = \lambda$. If the high-quality producer sets a limit price $p - \varepsilon$, where ε is infinitesimally small, she maximizes profits and drives the competitive fringe out of the market. Following the standard practice of quality-ladders growth models, we assume that even if $\varepsilon = 0$, all consumers buy the high-quality version of each product even if in principle they are indifferent between the two quality versions of each good.

B. Productivity Heterogeneity

Following the spirit of Taylor (1993), suppose that the utility function of the representative consumer is given by

$$U = \int_{\omega \in \Omega} \ln \left[\beta q(\omega) \right] d\omega.$$

The demand for a typical variety is $q(\omega) = EL/(p(\omega)M_c)$ as in the main text. Assume that firms discover new varieties associated with a productivity level $\lambda(\omega)$, and once the product is developed and sold in the market it can be produced by a competitive fringe with marginal and average costs equal to the wage $w = 1$. The marginal cost of a firm with productivity level $\lambda(\omega)$ is $1/\lambda(\omega)$. Consider the profit flow of a firm with productivity λ which charges the limit price $w - \varepsilon = 1 - \varepsilon$, where $\varepsilon \to 0$, to drive the competitive fringe out of the market. The profit flow is given by

$$\pi(\omega) = p(\omega)q(\omega) - (1/\lambda)q(\omega) - f_p = (1 - \lambda^{-1})\frac{EL}{M_c} - f_p,$$

which is identical to (4) in the main text. Limit-pricing strategies and Cobb-Douglas tastes generate a model with heterogeneous productivity and markups: More productive firms charge lower prices and earn higher profits as in Melitz and Ottaviano (2008).

C. Proof of Lemma 1

Substitute $\lambda_p^{-1} > \lambda^{-1}$ in the integral expression of equation (13) to obtain

$$\tilde{\lambda}_p^{-1} < \int_{\lambda_p}^{\infty} \lambda_p^{-1} g(\lambda) d\lambda / [1 - G(\lambda_p)] = \lambda_p^{-1},$$

which yields $\tilde{\lambda}_p > \lambda_p$. Differentiation of (13) yields $\partial\tilde{\lambda}_p/\partial\lambda_p = \tilde{\lambda}_p \mu(\lambda_p)(\lambda_p^{-1} - \tilde{\lambda}_p^{-1}) > 0$. The same logic and calculations apply to $\tilde{\lambda}_x$ defined by (13). *Q.E.D.*

D. Proof of Proposition 2

We will prove that $H(\lambda_i, \alpha)$ defined in (18) is strictly decreasing in λ_i. Let $J(\lambda_i, \alpha) = [1 - G(\lambda_i)](1 - \alpha\tilde{\lambda}_i^{-1})$. Since $\lambda_i > \alpha$, it easily follows that $J(\lambda_i, \alpha) > 0$. Applying the Leibnitz rule in calculus, we have

$$dJ(\lambda_i, \alpha)/d\lambda_i = -(1 - \alpha\lambda_i^{-1})g(\lambda_i).$$

Differentiating $H(\lambda_i, \alpha)$ with respect to λ_i yields the desired result:

$$\frac{\partial H}{\partial \lambda_i} = \frac{(1 - \alpha\lambda_i)dJ/d\lambda_i - \alpha\lambda_i^{-2}J}{(1 - \alpha\lambda_i^{-1})^2} + g(\lambda_i) = -\alpha\lambda_i^{-2}J/(1 - \alpha\lambda_i^{-1})^2 < 0.$$

The property $\partial H/\partial \lambda_i < 0$ in conjunction with equation (9) imply that the left hand side of (17) is strictly decreasing in the production cutoff quality level λ_i.

The full characterization of the solution requires the evaluation of the left-hand-side of (17) when λ_p approaches its boundaries. Equation (9) implies that at the lower bound of the production cutoff quality level ($\lambda_p = 1$) the export cutoff quality level equals the per-unit trade costs ($\lambda_x = \tau$), whereas as λ_p approaches its upper bound ($\lambda_p \to k$), λ_x approaches infinity ($\lambda_x \to \infty$). The following conditions characterize the limiting behavior of and $H(\lambda_p, 1)$ and $H(\lambda_x, \tau)$:

$$\lim_{\lambda_p \to 1} H(\lambda_p, 1) = \infty, \qquad \lim_{\lambda_x \to \tau} H(\lambda_x, \tau) = \infty, \tag{30}$$

$$\lim_{\lambda_p \to k} H(\lambda_p, 1) > 0, \qquad \lim_{\lambda_x \to \infty} H(\lambda_x, \tau) = 0. \tag{31}$$

To prove the claims in (30), first note that $J(\lambda_i, \alpha) = \int_{\lambda_i}^{\infty} [1 - \alpha\lambda^{-1}] g(\lambda) d\lambda$. Inserting this expression into $H(\lambda_i, \alpha)$ yields

$$H(\lambda_i, \alpha) = \frac{\int_{\lambda_i}^{\infty} [1 - \alpha\lambda^{-1}] g(\lambda) d\lambda}{1 - \alpha\lambda_i^{-1}} - [1 - G(\lambda_i)].$$

The numerator of the first term on the right hand side is always positive as long as $g(\cdot)$ is continuous at λ_i. The claims in (30) easily follow as we take the limits of both sides as $\lambda_p \to 1$ and $\lambda_x \to \tau$. Also, note that as $\lambda_x \to \infty$, the right hand side of the above equation approaches to 0, i.e., $\lim_{\lambda_x \to \infty} H(\lambda_x, \tau) = 0$. The first claim in (31) immediately follows from $H(k, 1) = [1 - G(k)](k^{-1} - \tilde{k}^{-1})/(1 - k^{-1}) > 0$ since $k^{-1} > \tilde{k}^{-1}$, where $\tilde{k} \equiv \left[\frac{1}{1-G(k)} \int_k^{\infty} \lambda^{-1} g(\lambda) d\lambda\right]^{-1}$.

In summary, the left-hand-side of (17) is always positive and defines a negatively sloped curve starting at infinity as $\lambda_p \to 1$ and reaching the value $H(k, 1)$ as $\lambda_p \to k$. The right-hand-side of (17) equals to $\delta f_e / f_p$ and it is independent of λ_p. Therefore, the existence of a unique cutoff quality level $\lambda_p > 1$ is guaranteed for a sufficiently high value of $\delta f_e / f_p$. Formally, a sufficient but hardly necessary condition for the existence of the unique equilibrium is that $H(k, 1) < \delta f_e / f_p$. It is then obvious from equation (9) that the unique production cutoff quality level λ_p determines the unique export cutoff quality level λ_x such that $\lambda_x > \tau$

and $\lambda_x > \lambda_p$. *Q.E.D.*

E. Proof of Lemma 2

Totally differentiating equations (9) and (17) yields

$$\frac{\partial \lambda_p}{\partial n} = -\frac{f_x}{f_p} H(\lambda_x, \tau) \left[\frac{\partial H(\lambda_p, 1)}{\partial \lambda_p} + \frac{n}{\tau} \left(\frac{f_x \lambda_x}{f_p \lambda_p} \right)^2 \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} \right]^{-1} > 0, \quad (32)$$

$$\frac{\partial \lambda_p}{\partial \tau} = -\frac{f_x}{f_p} \left[\frac{\lambda_x}{\tau} \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} + \frac{\partial H(\lambda_x, \tau)}{\partial \tau} \right] \left[\frac{\partial H(\lambda_p, 1)}{\partial \lambda_p} + \frac{1}{\tau} \left(\frac{f_x \lambda_x}{f_p \lambda_p} \right)^2 \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} \right]^{-1} < 0, \quad (33)$$

$$\frac{\partial \lambda_p}{\partial f_x} = [1 - G(\lambda_x)] \left[f_p \frac{\partial H(\lambda_p, 1)}{\partial \lambda_p} + \left(\frac{\lambda_x}{\lambda_p} \right)^2 \left(\frac{f_x^2}{\tau f_p} \right) \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} \right]^{-1} < 0. \quad (34)$$

The sign of the last bracket in each expression is negative due to Proposition 2 which also implies that the sign of the first bracket in (33) is negative. To see this note that the first term of that bracket can be written as

$$\frac{\lambda_x}{\tau} \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} = -\frac{[1 - G(\lambda_x)](1 - \tau \tilde{\lambda}_x^{-1})}{\lambda_x (1 - \tau \lambda_x^{-1})^2},$$

and that the second term in the same bracket is

$$\frac{\partial H(\lambda_x, \tau)}{\partial \tau} = [1 - G(\lambda_x)] \left[\frac{\lambda_x^{-1} - \tilde{\lambda}_x^{-1}}{(1 - \tau \lambda_x^{-1})^2} \right].$$

Adding these two expressions yields

$$\frac{\lambda_x}{\tau} \frac{\partial H(\lambda_x, \tau)}{\partial \lambda_x} + \frac{\partial H(\lambda_x, \tau)}{\partial \tau} = \frac{[1 - G(\lambda_x)](\tau - \lambda_x) \tilde{\lambda}_x^{-1}}{\lambda_x (1 - \tau \lambda_x^{-1})^2} < 0,$$

which implies that the sign of (33) is negative. *Q.E.D.*