Full length articles

International trade, product lines and welfare: The roles of firm and consumer heterogeneity☆

Phillip McCalman

Deaprtment of Economics, University of Melbourne, Australia

A B S T R A C T

A central prediction of international trade models is that increased integration leads to specialization. This mechanism has been used to gain insight into the location of industries across countries, the reallocation of output across firms as well as the variation of a firm's product range as countries liberalize. Nevertheless, the notion that international trade will lead firms to rationalize their product portfolios and concentrate on their "best" products doesn't always square with reality. In particular, firms in prominent industries have, on occasion, extended their offerings to include a lower quality version/option as international competition increases – expanding rather than contracting their product portfolio. This paper demonstrates that such behavior can be generated in a standard trade model if there is consumer heterogeneity within a country and firms leverage these differences to their advantage. In this setting, increased competition can be associated with either product line reductions or extensions. That is, both types of behavior can arise in equilibrium from ostensibly similar shocks. Since trade costs directly influence the intensity of competition, their variation has important implications for product line design and also the distribution of welfare gains. In particular, product line extensions due to trade liberalization have especially large welfare benefits for low income consumers.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The notion that international integration leads to specialization/concentration of production is ubiquitous in trade models. At the industry level, production is specialized/concentrated in countries according to comparative advantage. Within industries, international trade leads to a reallocation of production toward more efficient firms. And in a more recent literature that considers a firm's product range, international trade is a force that leads a firm to rationalize its product portfolio. While it is easy to find cases where firms have reduced their product range as foreign competition increases; nevertheless, there are a number of important examples that run counter to this wisdom.

Consider the "Quartz crisis" in the Swiss watch industry. The introduction of cheap reliable electronic watches in the 1970's by Japanese firms reduced the number of Swiss watchmakers from 1600 in 1970 to 600 by 1983. This reduction is attributed to a

continued focus on traditional high quality mechanical watches. 1983 proved to be a pivotal year for the Swiss watchmaking industry. While continuing to produce high end watches, the Swiss industry launched the Swatch aimed at the low end of the market. Contrary to the standard prediction, this cheap plastic watch was an extension of the product range rather than a contraction. Moreover, by the early 1990's, the Swatch was the world's best selling watch.[1]

The Swatch isn't an isolated example. It turns out that the pairing of low end product line extensions and increased competition is observed frequently enough for the business literature to give it a name/s; "fighter brands" or "flanker brands".[2] Take, for example, the introduction of GM's utilitarian Saturn range of small cars in response to the increased US market share of Japanese producer's in the 1980's. This move by GM confounded many industry commentators.[3] Especially when compared to the behavior of other incumbents; Peugeot, Renault, Alfa Romeo and Fiat all exited the US market, while Ford and Chrysler stuck with their traditional line-ups. Analysts have also puzzled over behavior in the desktop laser printer market. Why would firms already offering a high quality 10 pages per minute (ppm) printer introduce a lower quality 5 ppm at the same time the number of competitors increased from a handful to over 40? (see Deneckere and Preston McAfee (1996) and Teisberg and Clark (1994)). These examples raise a series of questions. Why do some firms expand their offerings and not others? Why do we observe these extensions only some of the time, while we observe product line rationalization at other times? Do these outcomes have different welfare implications? Are these differences big?

The common thread running through these examples is that entry/competition is associated with some incumbents extending their product range into the low end of the market, a segment they had not been serving. The emphasis on market segments points to a role for consumer heterogeneity; a dimension that has previously been overlooked in the international trade literature. The aim of this paper is to help fill this gap. Moreover, understanding why and when the length of product lines are varied allows for a welfare evaluation of their consequences.

The "versioning" strategy pursued by firms in the examples above is a form of second degree price discrimination, and is also the basis of the model developed below.[4] This sales technique is applied in settings where a firm is aware of the distribution of consumer types in the economy but does not observe the type of any given consumer. In this sense, the new margin being added to trade models is the ability of firms to design a menu of options for consumers.[5] The purpose of a properly designed menu is to motivate consumers to select items in a fashion consistent with their type.

This behavior has been studied previously in a monopoly setting.[6] In contrast, our model considers firms that compete in a monopolistically competitive manner, and additionally, these firms differ in their productivities.[7] Together these features form a tractable model which generates a rich set of equilibrium possibilities, including either the expansion or contraction of product lines from seemingly similar shocks that increase competition.[8]

The three dimensions of the model (versioning, monopolistic competition and firm heterogeneity) all retain characteristics familiar from other applications, and it is their interaction that generate the rich results. A key prediction of "versioning" is that offering a low end product comes at the expense of rents foregone from high value customers. For example, offering a low end version (5 ppm printer) provides a viable outside option for the high end customer (who the 10 ppm printer is aimed at) and the price of the high end version needs to reflect this alternative. The foregone rents (or information rents captured by the high types) are an implicit cost of serving the low end, with this implicit cost driven by the difference in valuation between the high and low type for the low end product.

In a monopoly setting this difference in valuations is exogenous, while under monopolistic competition this difference depends on the intensity of competition in each segment; hence, it is endogenous. To accommodate this feature, a translated CES structure is adopted, partially for analytic convenience, but mostly for the type specific choke price it generates.[9] The analytic convenience is due to a symmetric demand system across varieties within a type. Combined with the type specific choke prices, this implies that the implicit cost of serving the low type, while endogenous, is the same for all firms.[10]

The cost that differs across firms is the marginal cost of production. Without an implicit cost, the decision to serve a market segment is based on the comparison of the marginal cost (explicit cost) and the segment's choke price. This is the familiar "selection" mechanism. With an implicit cost, the decision to serve the low end deviates from marginal cost. Moreover, the impact

---

[1] See Tushman and Radov (2000) and Moon (2004).

[2] See Porter (1980) and Ritson (2009). The term "fighting brand" comes from the strategies employed by American Tobacco in the 1890's for "plug" or chewing tobacco.

[3] The Saturn range was an instant success, selling over a million cars within a few years of its launch in 1989.

[4] See Shapiro and Varian (1998) for a discussion of versioning.

[5] Models of international trade that consider vertical differentiation tend to focus on firms that produce a single version – see Flam and Helpman (1987) and Fajgelbaum et al. (2011).

[6] See Maskin and Riley (1984).

[7] McCalman (2018) considers second degree price discrimination and monopolistic competition in the context of a CES setting with homogeneous firms. Importantly, the toughness of competition doesn't vary in a monop comp/CES setting so selection is based on the interaction of productivity draws and fixed costs. If fixed costs occur at the variety level, then all firms will provide full product lines in all markets in which they are active. For fixed costs to generate variation in product line length requires them to be version specific. While possible, such an assumption is extremely close to assuming the outcome.

[8] The IO literature has also considered what factors could lead to the extension of product lines. Johnson and Myatt (2003) is the most prominent paper in this literature. They adopt an upgrades approach and examine an asymmetric Cournot duopoly. In their setting, consumer heterogeneity can generate a marginal revenue function that has upward sloping portions. Moving from a monopoly to a duopoly, they show that the new equilibrium can involve jumping over this upward sloping segment (i.e. reaction functions are non-monotonic), resulting in the previous monopolist extending their product lines at the lower end. In contrast, if the marginal revenue function is always downward sloping product lines are pruned when entry occurs.

[9] For other recent applications of this demand system see Arkolakis et al. (2018) and Jung et al. (2019)

[10] For a set-up that gives similar results based on quadratic preferences/linear demand system see McCalman (2019).

of the implicit cost is most pronounced for the higher productivity firms; low productivity firms do not possess an marginal cost that would enable them to serve the low end of the market (their marginal cost is greater than the low end choke price).

The resulting equilibrium is characterized by firms that offer product lines of differing length. Furthermore, the extent to which any firm chooses to (under-) serve the low end of the market is conditional on surplus available in the high end of the market; which itself depends on the degree and nature of competition. This setting provides enough richness to allow insight into why we sometimes observe a set of incumbents extending their product lines, others do nothing and some exit. Yet, in seemingly very similar circumstances we observe a set of incumbents rationalizing their product lines, others do nothing and some exit. Which outcome arises turns on the size of the endogenous implicit cost.

Intuitively, a less competitive environment results in residual demand functions that allow large rents to be extracted by firms. This is especially true at the high end of the market. However, this generates a large implicit cost; hence, only firms with a low explicit cost would consider serving the low end (and even they may not be interested). A shock that induces entry (decline in trade barriers/entry costs, an improvement in technology or market growth) changes the location of the residual demand curves. In particular, the high end residual demand curve shifts in. What about the low end? Consider an extreme situation where in the initial equilibrium no firm serves the low end of the market (so it's residual demand is not altered by competition and exists as a latent market segment). Then the inward shift of the residual demand curve at the high end reduces the implicit cost of serving the low type. At some stage this process will induce the best firms to enter the low end of the market. This is a very pro-competitive outcome; the high productivity firms always had the capability of serving the low end, but they chose not to. The end result is increased competition coinciding with product line extensions. Consistent with the seemingly confounding behavior in the watch, auto and printer markets.

If instead we consider a setting where competition is already intense, then the rents available are smaller. This produces a relatively small implicit cost. A shock that induces entry once again shifts in the residual demand functions. However, the change in implicit cost won't be large. Consequently, the "selection" mechanism dominates; Do I have a marginal cost lower enough to operate? Do I have a low enough cost to serve the low end? Since the shock shifts the residual demand curves inward, the answer to these questions for a set of incumbent firms will now be no. Some firms exit and some rationalize their product lines as competition increases. Now increased competition and product line rationalization coincide.

The differing predictions also generate differences in the magnitude of welfare changes. If product line rationalization is a feature of the equilibrium response, then the welfare benefits reflect tougher selection into the market; gains arise as high cost firms are replaced by lower cost firms. However, if a set of firms extend their product lines, this has a disproportionately large impact on the welfare of the low income consumers.

There are two sources to this gain; the usual gain from overall selection into the market, along with a new extensive margin gain that occurs exclusively for the low income segment. This new margin can result in a substantial boost in welfare. In particular, based on standard parameter choices, this new margin can predict gains from trade for the low income group to be over 10 percentage points larger than those implied by selection alone. The size of these benefits point to an important new dimension, that of the distribution of gains, in the search for the elusive pro-competitive effects of trade.

The mechanism I emphasize differs from the usual incentives assumed to underlie multi-product firms. One popular motivation is based on core competency; a firm is good at producing a specific variety and this aptitude carries over imperfectly to nearby varieties.[11] In these models, the varieties are distinct, so a consumer would be willing to add all of them to their within sector consumption basket – the emphasis is on firms introducing additional horizontally differentiated varieties. A related approach assumes that a firm receives a random draw for capabilities across multiple goods/sectors. In this case, each firm produces at most one good in each "nest" of the utility function.[12] To focus on the independent operation of the extension/selection mechanism, I rule out each of these standard motivations by assuming that firms cannot adapt to produce a related but distinct variety and that their capability is only within a single sector. Instead I highlight the ability of a firm to produce different versions of its variety. This captures Hewlett-Packard offering high and low end laser printers; both embody similar technology and the choice of the consumer is within variety. Hence, I focus on vertical product lines rather than horizontal product lines (the subject of the previous literature).

To derive and develop these results the paper has the following structure. First, a closed economy model is introduced. The equilibrium must satisfy the familiar free entry and zero cut-off profit conditions. The new dimension is that both of these conditions depend not only on the maximum cost consistent with survival in the market but also a separate lower maximum cost that defines the minimum efficiency level necessary to serve the low income segment. As is standard in the literature, the perturbation considered is variation in market size. Small markets have less competitive outcomes and are the most likely to be associated with product line extensions. An open economy version of the model is then considered. Changes in trade costs provides a realistic source of variation in the degree of competition. In line with the closed economy results, high trade barriers are consistent with less competitive outcomes and therefore are most likely to be associated with product line extensions when trade barriers are reduced. Additionally, insight is also gained into which firms add low end options in the domestic market, and which firms expand in their export markets at the low end (over and above what would be expected from changes in market access). When trade barriers are unilaterally reduced, welfare benefits occur in the short run, and also in the long run, provided the reduction in trade barriers is sufficiently large, a finding that contrasts with the existing literature. In all these cases, product line

---

[11] The key mechanisms are set out in Eckel and Neary (2010), Dhingra (2013), Nocke and Yeaple (2014) and Mayer et al. (2014).
[12] See Bernard et al. (2011).

extensions disproportionally benefit the low income segment of the market, gains which directly stem from the pro-competitive effect of either unilateral or reciprocal trade liberalization.

2. Closed economy

2.1. Preferences and consumer heterogeneity

The economy consists of two sectors, Y and Q where consumers have identical and separable preferences over these products. In particular assume:

$$u(Y, Q) = U(Y) + V(Q)$$

where $U' > 0$, $U'' < 0$, $V(0) = 0$, $V' > 0$ and $V'' < 0$. While all consumers have identical preferences, they differ in terms of income – the source of consumer heterogeneity.

To examine the role of income differences in a tractable manner I follow Tirole (1988) by adopting a quasi-linear formulation where income differences are embedded in the marginal utility of income. In particular, let I denote income and T represent total expenditure on output of the Q sector so that $I - T$ is income net of this expenditure. In addition nominate Y as the numeraire. This allows preferences to be represented as separable in net income, $U(I - T) + V(Q)$. If T is assumed to be small relative to I, then preferences can be approximated by $U(I) - TU'(I) + V(Q)$ where $U'(I)$ is the marginal utility of income. As a consequence, consumer choice in the Q sector is governed by:

$$\alpha^I V(Q) - T$$

where $\alpha^I = {}^1/_{U'(I)}$ is the inverse of the marginal utility of income. Since U is concave, consumers with high income have a low $U'(I)$ and therefore a high α^I.

To this structure I add texture and assume that the Q sector is composed of a continuum of differentiated varieties indexed by ω, where a consumer views the products comprising Q as a translated CES index:

$$Q = \left(1 + \int_{\omega \in \Omega} (q(\omega) + 1)^\rho d\omega - \int_{\omega \in \Omega} d\omega\right)^{\frac{1}{\rho}} \tag{1}$$

where Ω is the measure of products available and $q(\omega)$ can be interpreted as either quantity or quality (e.g. pages printed per minute). Since the set of products available can vary with income, it will typically be indexed by I.

Applying $V(Q) = \log(Q)$ results in an objective function for a consumer with income I:

$$\alpha^I V(Q_I) - T^I = \frac{\alpha^I}{\rho} \log\left(1 + \int_{\omega \in \Omega^I} (q(\omega) + 1)^\rho d\omega - \int_{\omega \in \Omega^I} d\omega\right) - \int_{\omega \in \Omega^I} T(\omega) d\omega$$

These preferences generate the following inverse demand (willingness to pay) system for a consumer with income I for any product ω:

$$p(\omega) = \frac{\alpha^I}{Q_I^\rho} (q(\omega) + 1)^{\rho-1}. \tag{2}$$

Under this specification it's possible that a consumer might not transact with all firms in the market. In particular, whenever $p(\omega) \geq \frac{\alpha^I}{Q_I^\rho}$, a consumer with income I will not purchase this product. Hence, the number of firms operating can depend on the income of consumers that form the market segment. Moreover, since this choke price plays an ongoing role in the analysis, we denote it more compactly as

$$\theta^I \equiv \frac{\alpha^I}{Q_I^\rho}. \tag{3}$$

This allows willingness to pay function of type I for variety ω to be written as:

$$p(\omega) = \theta^I (q(\omega) + 1)^{\rho-1}. \tag{4}$$

To emphasize that the location of this schedule is determined in equilibrium, it will be referred to as the "residual" demand curve. This highlights that any firm, in any market segment, takes Q_I^ρ as given, which means they adopt the perspective of the marginal firm in a market segment.

We are interested in analysing an environment where consumers differ in their income levels and firms are sophisticated enough to utilize this information. For the sake of simplicity, assume that a consumer has one of two incomes, $I^H > I^L$ which implies $\alpha^H > \alpha^L$.[13] Moreover, let β denote the fraction of high income types in the population. Assuming firms possess knowledge of both the level of income and its distribution, we allow then to formulate a set of non-linear prices.

These non-linear prices are implemented as a menu of options offered to consumers on a take-it-or-leave-it basis, $\{T(\omega), q(\omega)\}$, where $T(\omega)$ is the payment required for a product with attribute $q(\omega)$ by firm ω. Facing a residual demand curve given by (4) a firm evaluates the surplus from serving consumer I in the following way:

$$S^I(q) = \theta^I \int_0^q (z+1)^{\rho-1} dz = \frac{\theta^I \left((q+1)^\rho - 1\right)}{\rho} = \theta^I v(q). \tag{5}$$

where $v(q) = \frac{(q+1)^\rho - 1}{\rho}$. Note that since firms are assumed to be monopolistically competitive, they treat the θ^I as constant for

each market segment.

While a firm would like to extract all the surplus from a consumer, it is constrained by the fact it only knows the distribution of income and not the income of any individual. From the literature on second degree price discrimination, we know in this setting a firm designs the menu $\{T(q), q\}$ subject to a set of incentive compatibility (each income group prefers the option designed for them) and participation constraints (a consumer's net pay-off has to be non-negative). These constraints accommodate a wide range of possibilities, including the option to use linear prices.

3. Technology and firm behavior

Labor is the only factor of production, which can be hired at an exogenous wage rate that is normalized to unity. Entry in the differentiated product sector is costly as each firm incurs product development and production start-up costs. Subsequent production has a marginal cost, c, which is drawn from a distribution: $G(c)$ and $g(c) = G'(c)$ where $c \in [0, c_M]$, where c_M is the upper bound on the cost draws.[14] There are no per period fixed costs and the entry cost, incurred before the productivity draw, is f_e.

Since the entry cost is sunk, firms that can cover their marginal cost survive and produce. All other firms exit the industry. Surviving firms maximize their profits using the residual demand functions (4). In so doing, given the continuum of competitors, a firm takes the industry aggregate P (or equivalently Q) as given – which reflects the monopolistically competitive structure.

In maximizing profits, firms are assumed to be aware of the heterogeneity in the population but an individual consumer's income is not observable to them. Recall that β denotes the fraction of high types in the population – consumers with I^H and firms leverage this information by offering product lines; $\{T^I, q^I\}$, where T^I is the total payment required when purchasing a product with attribute q^I. These product lines are designed such that each type purchases the option intended for them, and in doing so they are left with non-negative net surplus, $S^I(q^I) - T^I \geq 0$. That is, we are considering second degree price discrimination in a monopolistically competitive setting with heterogeneous firms.

3.1. Profit maximizing product lines

Using the surplus functions given by (5) and the information on the distribution of types in the population, a typical monopolistically competitive firm chooses a menu of $\{T^I(\omega), q^I(\omega)\}$, $I \in \{L, H\}$ to maximize

$$\pi(\omega) = \beta\left(T^H(\omega) - cq^H(\omega)\right) + (1-\beta)\left(T^L(\omega) - cq^L(\omega)\right) - f_e$$

subject to

$$
\begin{aligned}
S^H\left(q^H(\omega)\right) - T^H(\omega) &\geq S^H\left(q^L(\omega)\right) - T^L(\omega) \\
S^L\left(q^L(\omega)\right) - T^L(\omega) &\geq S^L\left(q^H(\omega)\right) - T^H(\omega),
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
S^L\left(q^L(\omega)\right) - T^L(\omega) &\geq 0 \\
S^H\left(q^H(\omega)\right) - T^H(\omega) &\geq 0.
\end{aligned}
\tag{7}
$$

where (6) are the incentive compatibility constraints while (7) are the participation constraints. In a monopoly non-linear pricing problem the ordering of the $\theta's$ is enough to ensure that the single crossing property holds – implying that only two of these

---

[13] See McCalman (2019) for an analysis with an arbitrary number of market segments.
[14] For a model with increasing marginal costs see McCalman (2019).

constraints bind, the incentive constraint for the high and the participation constraint for the low type. However, since the θ's are determined as part of an equilibrium outcome we cannot simply take for granted that $\theta^H > \theta^L$. Nevertheless, we conjecture that this ordering holds (it is in fact satisfied in equilibrium) allowing the relevant constraints to be rewritten as:

$$
\begin{aligned}
T^L(\omega) &= \theta^L v\left(q^L(\omega)\right), \\
T^H(\omega) &= \theta^H v\left(q^H(\omega)\right) - \left(\theta^H - \theta^L\right) v\left(q^L(\omega)\right).
\end{aligned}
\tag{8}
$$

These prices imply that while a firm can extract all the surplus under the residual demand curve of the low type, the high type is able to capture information rents, $(\theta^H - \theta^L)v(q^L(\omega))$, by having the low type's product as their outside option. Substitution gives the following profit function:

$$
\pi(\omega) = \beta\left(\theta^H v\left(q^H(\omega)\right) - cq^H(\omega) - \left(\theta^H - \theta^L\right) v\left(q^L(\omega)\right)\right) + (1-\beta)\left(\theta^L v\left(q^L(\omega)\right) - cq^L(\omega)\right) - f_e
$$

An equivalent, and particularly revealing, way to express this objective function is to take the information rents paid to the high types and subtract them from the marginal benefit of serving a low type. So a firm's perceived intercept (or choke price) for the "net" marginal benefit/revenue of a low type is $\theta^L - \frac{\beta}{1-\beta}(\theta^H - \theta^L) = \frac{\theta^L - \beta\theta^H}{1-\beta}$. This also defines the upper bound on the marginal cost that is consistent with a firm optimally serving the low end of the market.[15] Define this cost as $c_B \equiv \frac{\theta^L - \beta\theta^H}{1-\beta}$.

Similarly, let $c_D = \theta^H$ reference the cost of the firm who is just indifferent about remaining in the industry. Combining these two cut-offs implies the following useful feature: $\theta^L = \beta c_D + (1 - \beta)c_B$.

Using these definitions the objective function can be re-expressed in a relatively simple way. In particular, it resembles a first degree price discrimination problem over the "virtual" demand system characterized by $\{c_D, c_B\}$ as the relevant choke prices. This has the advantage that the profit maximizing choice in one segment does not depend on the choice in the other segment. Fig. 1 depicts this system along with the profit maximizing choices which I now formally confirm.

$$
\pi(\omega) = \beta\left(c_D v\left(q^H(\omega)\right) - cq^H(\omega)\right) + (1-\beta)\left(c_B v\left(q^L(\omega)\right) - cq^L(\omega)\right) - f_e
$$

The first order conditions require

$$
c_D v'\left(q^H(\omega)\right) = c \quad \Rightarrow \quad q^H(\omega) = \left(\frac{c_D}{c}\right)^{\frac{1}{1-\rho}} - 1,
\tag{9}
$$

$$
c_B v'\left(q^L(\omega)\right) = c \quad \Rightarrow \quad q^L(\omega) = \left(\frac{c_B}{c}\right)^{\frac{1}{1-\rho}} - 1.
\tag{10}
$$

Using the defining characteristic of a firm, c, and (9) and (10) the value function has the following form:

$$
\pi^*(c) = \begin{cases} \beta\left(\frac{c_D}{\sigma-1}\left(\left(\frac{c_D}{c}\right)^{\sigma-1} - 1\right) - (c_D - c)\right) + (1-\beta)\left(\frac{c_B}{\sigma-1}\left(\left(\frac{c_B}{c}\right)^{\sigma-1} - 1\right) - (c_B - c)\right) & c \in [0, c_B] \\[2ex] \beta\left(\frac{c_D}{\sigma-1}\left(\left(\frac{c_D}{c}\right)^{\sigma-1} - 1\right) - (c_D - c)\right) & c \in (c_B, c_D] \end{cases}
$$

where $\sigma = \frac{1}{1-\rho}$.

The pricing menu offered by a firm follows from (8), (9) and (10). If $c \in [0, c_B]$:

$$
T^H(c) = \frac{c_D}{\rho}\left(\left(\frac{c_D}{c}\right)^{\sigma-1} - 1\right) - (1-\beta)\frac{(c_D - c_B)}{\rho}\left(\left(\frac{c_B}{c}\right)^{\sigma-1} - 1\right),
\tag{11}
$$

$$
T^L(c) = \frac{\theta^L}{\rho}\left(\left(\frac{c_B}{c}\right)^{\sigma-1} - 1\right), \quad \text{where } \theta^L = \beta c_D + (1-\beta)c_B.
\tag{12}
$$

---

[15] Although the information rents represent an implicit cost of serving the low end of the market, netting them off the low type's residual demand function allows us to conduct the analysis in terms of explicit costs.
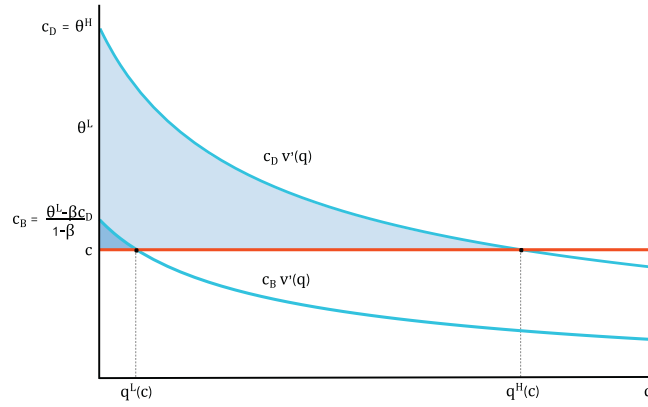
Fig. 1. Profit maximizing product lines. Redefining the surplus available from the low income segment by subtracting off the information rents conceded allows the problem of menu design under second degree price discrimination to be recast as one of menu design under first degree price discrimination. All firms face the same modified demand system, but their behavior depends on the cost draw. $c_D = \theta^H$ is the highest cost consistent with serving the high income consumers (i.e. marginal cost must be less than a high income type's choke price). To serve the low income segment, a firm must have a cost draw that's not only smaller than a low income consumer's maximum willingness to pay, θ^L, but also covers the information rents conceded to the high type (mediated by the relative frequency of the two types, $\frac{\beta}{1-\beta}$). The largest cost draw consistent with serving the low income segment is given by $c_B = \theta^L - \frac{\beta}{1-\beta}(\theta^H - \theta^L) = \frac{\theta^L - \beta\theta^H}{1-\beta}$. The shaded areas represent the profit from offering a high income type $q^H(c)$ and the darker shaded area is the profit for offering a low income type $q^L(c)$. Profits are increasing in c_D and c_B.

While for $c \in (c_B, c_D)$:

$$T^H(c) = \frac{c_D}{\rho}\left(\left(\frac{c_D}{c}\right)^{\sigma-1} - 1\right) \tag{13}$$

4. Equilibrium

An equilibrium is characterized by the two cost thresholds: c_D, c_B. These thresholds determine the location of the residual demand curves and naturally must be consistent with the entry and market segment servicing decisions of firms. Let n be the number of firms that serve the high types and n_L be the number that serve the low types. Writing this more formally, an equilibrium must satisfy:

$$\theta^H(n(c_D, c_B)) = c_D \tag{14}$$

$$\theta^L\left(n^L(c_D, c_B)\right) = \beta c_D + (1-\beta)c_B \tag{15}$$

4.1. Entry and segment servicing decisions

For any given pair, these thresholds influence the incentive to enter the industry since they determine the expected profit before a firm's cost draw is realized. Let n^e represent the number of firms that pay f_e. The thresholds also determine a firm's behavior once their marginal cost is realized; a firm only operates in a market segment if its cost draw is below the relevant threshold. Hence, an equilibrium requires that the ex ante behavior is consistent with the ex post behavior. Aligning the ex ante and ex post incentives by segment requires:

$$n^e(c_D, c_B)G(c_D) = n(c_D) \tag{16}$$

$$n(c_D)\frac{G(c_B)}{G(c_D)} = n_L(c_D, c_B) \tag{17}$$

(16) imposes that of those firms that pay f_e, the fraction with cost draws below $G(c_D)$ will serve the high end of the market, $n(c_D)$. (17) further requires of the firms operating, only those with cost draws below c_B will find it optimal to serve the low end of the

market. Which combine to give:

$$n^e(c_D, c_B)G(c_B) = n_L(c_D, c_B) \tag{18}$$

4.2. Ex ante incentives for entry

Entry occurs until all the expected profits for a firm are exhausted. Using M to measure market size, this implies

$$E\pi = M\left(\int_0^{c_D} \beta\left(\frac{c_D}{\sigma-1}\left(\left(\frac{c_D}{c}\right)^{\sigma-1}-1\right)-(c_D-c)\right)dG(c)+ \right.$$
$$\left. \int_0^{c_B}(1-\beta)\left(\frac{c_B}{\sigma-1}\left(\left(\frac{c_B}{c}\right)^{\sigma-1}-1\right)-(c_B-c)\right)dG(c)\right)-f_e = 0$$

This condition can also be expressed as:

$$\beta\left(\frac{c_D}{\sigma-1}\left(c_D^{\sigma-1}\overline{c^{1-\sigma}}-1\right)-(c_D-\overline{c})\right)G(c_D) + (1-\beta)\left(\frac{c_B}{\sigma-1}\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}}-1\right)-(c_B-\overline{c}_L)\right)G(c_B) = \frac{f_e}{M} \quad \text{(FE)}$$

where $\overline{c^{1-\sigma}} = \int_0^{c_D} c^{1-\sigma}dG(c)$, $\overline{c} = \int_0^{c_D} cdG(c)$, $\overline{c_L^{1-\sigma}} = \int_0^{c_B} c^{1-\sigma}dG(c)$ and $\overline{c}_L = \int_0^{c_B} cdG(c)$.

The FE condition reflects an ex ante perspective; what combinations of c_D and c_B are consistent with a marginal entrant just expecting to cover the entry cost, f_e. Since expected profits are increasing in both c_D and c_B, the free entry condition must reflect this trade-off. Moreover, an increase in M, market size, leads to an inward shift of the FE condition. Fig. 2 depicts the combinations of c_D and c_B consistent with free entry driving expected profits to zero and the consequences of an increase in market size.
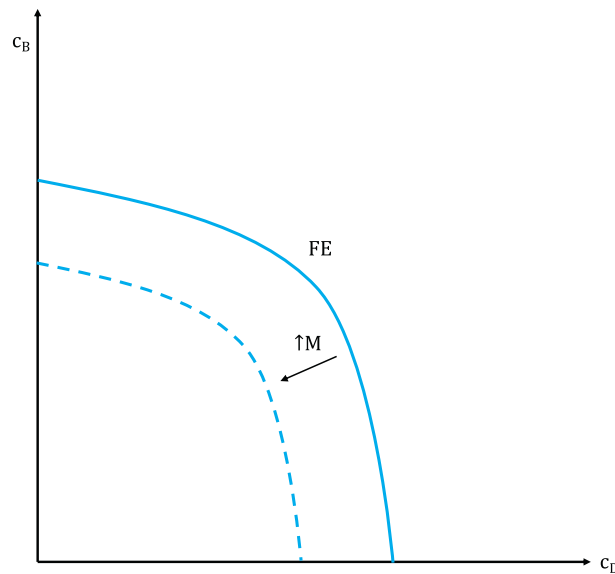
Fig. 2. Ex ante incentives and the free entry condition. Since expected profits are increasing in both c_B and c_D, the slope of FE must be negative. An increase in market size, M, shifts the FE in toward the origin.

4.3. Ex post incentives to serve market segments

Once a firm knows its cost draw, c, the thresholds c_D and c_B determine which segments it can viably serve. In particular, for a firm to serve the high type their cost draw must be less than a high type's choke price (using (3)).

$$
\begin{aligned}
c_D &= \theta^H \\
&= \frac{\alpha^H}{Q_H^\rho} \\
&= \frac{\alpha^H}{1 + \int_{\omega \in \Omega^H} (q(\omega)+1)^\rho d\omega - \int_{\omega \in \Omega^H} d\omega} \\
&= \frac{\alpha^H}{1 + \int_0^{c_D} \left(\frac{c_D}{c}\right)^{\sigma-1} dG(c) - n} \\
&= \frac{\alpha^H}{1 + n\left(c_D^{\sigma-1}\overline{c^{1-\sigma}} - 1\right)}
\end{aligned}
$$

Rearranging this expression defines a mapping between c_D and the number of firms serving the high end which is consistent with the location of the residual demand curve:

$$
n = \frac{\alpha^H - c_D}{c_D} \frac{1}{\left(c_D^{\sigma-1}\overline{c^{1-\sigma}} - 1\right)} \tag{19}
$$

Similarly, the location of the low types residual demand must satisfy,

$$
\theta^L = \frac{\alpha^L}{1 + n_L\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}} - 1\right)},
$$

and in doing so disciplines the number of firms serving the low end to be:

$$
n^L = \frac{\alpha^L - \theta^L}{\theta^L} \frac{1}{\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}} - 1\right)}.
$$

Recalling that $\theta^L = \beta c_D + (1-\beta)c_B$, we now have a mapping between the cut-offs and the number of firms serving the low end of the market,

$$
n^L = \left(\frac{\alpha^L - (\beta c_D + (1-\beta)c_B)}{\beta c_D + (1-\beta)c_B}\right) \frac{1}{c_B^{\sigma-1}\overline{c_L^{1-\sigma}} - 1}. \tag{20}
$$

Notice that the number of firms serving the low end is not independent of the number operating in the high end. Recall from (17), given the thresholds, it must be the case that those that serve the low end are a subset of firms serving the high end, with the fraction determined by the distribution of cost draws.

Equating (17) and (20) defines the following ex post equilibrium condition:

$$
\frac{\alpha^L - \theta^L}{\theta^L} = \left(\frac{\alpha^H - c_D}{c_D}\right)\frac{G(c_B)}{G(c_D)}\left(\frac{c_B^{\sigma-1}\overline{c_L^{1-\sigma}} - 1}{c_D^{\sigma-1}\overline{c^{1-\sigma}} - 1}\right) \quad \text{(ZCP)}
$$

This condition exhibits a non-monotonic relationship between c_D and c_B that depends on β, $G(c)$, σ and α^j.[16] An important exogenous factor missing from this list is market size (or equivalently f_e). This feature plays an important role in the comparative statics with respect to market size below, and by extension trade costs.

What is the intuition underlying the non-monotonic relationship between c_D and c_B? This property reflects the shifting importance of the surplus available from serving a low type (captured by the low type's residual demand θ^L) and the implicit cost of serving a low type given by ($\theta^H - \theta^L$). Recall that c_B is a low type's choke price net of the implicit cost of serving the low type, $c_B = \theta^L - \frac{\beta}{1-\beta}(c_D - \theta^L)$. Now if c_D is large (n must be small and the surplus to be extracted from a high type is large), then the implicit cost of serving the low type is also big, $(c_D - \theta^L)$. Only firms with a low explicit cost draw, c, will serve the low types. If c_D is lowered (and n increases) then the implicit cost is also reduced since it is now less lucrative to serve the high type. However, increased n also implies increased n_L; θ^L declines as well. So an increase in c_B reflects a greater decline in

---

[16] See appendix for details.

the implicit cost $\frac{\beta}{1-\beta}(c_D - \theta^L)$ relative to the surplus available from the low type, θ^L. That is, c_D is falling faster than θ^L. The reduction in the implicit cost reflects an erosion of market power and the pro-competitive effect is reflected in an increase in c_B. This mechanism is relatively strong when c_D is high and gives rise to the negative sloped portion of ZCP in Fig. 3.

The diminished relative importance of the implicit cost as c_D is lowered means that the behavior of c_B is increasingly driven by the surplus available from serving a low type, θ^L. Beyond some point it must become the dominant factor. However, the surplus from serving the low type is also falling. This implies that c_B must also decline when θ^L is the primary determinant of whether to serve the low income segment. The mechanism now is predominately one of selection. Only firms with low explicit costs can serve low income consumers. This type of product line rationalization is consistent with the forces described in the previous literature and generates the positively sloped segment of the ZCP in Fig. 3.

I'll focus the analysis on outcomes where $c_B > 0$ and $c_D > 0$. This requires a market size greater than $\underline{M}$, which is implicitly defined by:

$$\int_0^{\frac{\alpha^L}{\beta}} \beta \left(\frac{\alpha^L}{\beta(\sigma-1)} \left(\left(\frac{\alpha^L}{\beta c} \right)^{\sigma-1} - 1 \right) - \left(\frac{\alpha^L}{\beta} - c \right) \right) g(c) dc = \frac{f_e}{\underline{M}} \tag{21}$$

Fig. 4 depicts the ZCP condition along with the FE condition. Equilibrium occurs when ex ante incentives that motivate entry, (FE), are consistent with ex post incentives for firms to serve the different market segments, (ZCP). In particular, note that this equilibrium partitions firms into three types. First, there are those with cost draws above c_D; these firms exit the market without producing. Second are the firms that draw costs below c_D but above c_B. These firms find it optimal to only serve the high end of the market and have a product line that consists of one offering. Finally, there is a set of firms with costs below c_B; the most productive firms. These firms serve both types; their product line consists of two distinct items. What happens to these product lines as the economic environment changes? We now examine this question.

4.4. Variation in market size: M

A central result of Melitz and Ottaviano (2008) is that larger markets are associated with a lower cost cut-off, c_D. A number of benefits then flow from this increased competitive pressure, including lower prices and higher welfare. To investigate whether these results are paralleled in the current setting recall that (FE) is a function of M while (ZCP) is not. Consequently, an increase in M results in an inward shift of (FE) while (ZCP) remains in place. Much like Melitz and Ottaviano (2008), larger markets are indeed associated with tougher selection; c_D is declining is M.

Does this tougher selection apply to all market segments? Fig. 5 confirms that this cannot be universally true. In particular, since (ZCP) is non-monotonic, increases in market size can result in both a decrease in c_D and an *increase* in c_B. Whenever c_B increases it must be the case that a set of firms that were previously serving only the high end of the market now extend their product lines to the low end of the market. Consequently, an increase in competitive pressure (lower c_D) is associated with a set of firms extending their product lines.

Proposition 1. For a given set of parameters $\{\alpha^l, \beta, G(c), \sigma, f_e\}$, there exists an M^* such that for $M \in [\underline{M}, M^*]$, $\frac{dc_B}{dM} > 0$. That is, a set of firms will extend their product lines to serve the low type as the market becomes more competitive (i.e. $\frac{dc_D}{dM} < 0$). For $M > M^*$, $\frac{dc_B}{dM} \leq 0$; some firms contract their product lines.
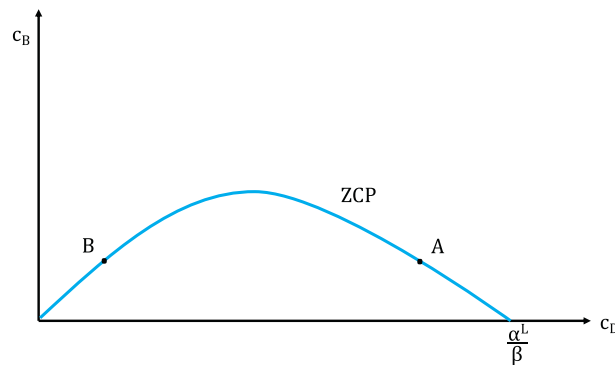
Fig. 3. Ex post incentives: zero cut-off profit condition. The *ZCP* is non-monotonic. When c_D is large, such as at point *A*, serving the high income segment alone generates correspondingly large profits (implying n must be small). Serving the low income segment reduces profits in the high income segment, generating a high implicit cost in terms of information rents. Only firms with low cost draws will have an incentive to enter the low income segment, so c_B must also be small. When c_D is small, see point *B*, the profits from serving the high income segment are also small since n must be large. The implicit cost of serving the low income segment are also small. Consequently, the decision to serve the low income segment is driven by typical selection considerations, implying that c_B must also be small.
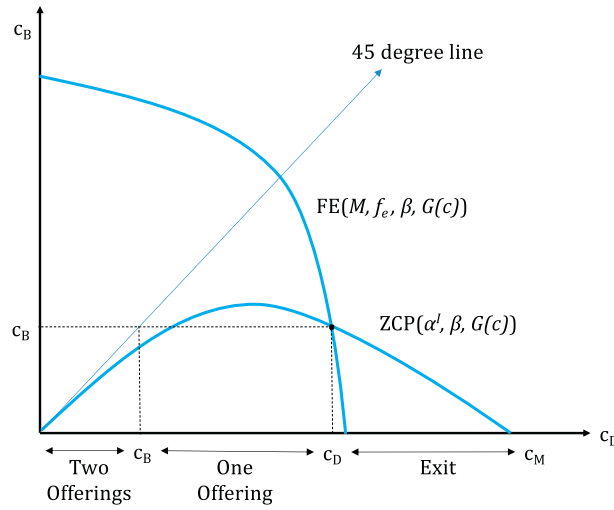
Fig. 4. Equilibrium cost thresholds and product line length. The intersection of *FE* and *ZCP* defines a point where expected profits are zero, and, when costs are realized, firms find it profit maximizing to behave according to their assigned cut-off bins. In particular, any firm with a cost draw above c_D exits, while a firm with a realized cost less than c_D but greater than c_B only serves the high income segment, a single version menu. Finally, any firm with a cost below c_B finds that they maximize profits by offering different product versions to the high and low income segments, a two version menu.

Existing firms who extend their product lines have intermediate productivity; high productivity firms already serve both segments, while the lowest productivity firms remain focused on only serving consumers with the strongest preference for the differentiated good. Consequently, there is a heterogeneous product line response to increased competition across firms.

The introduction of a low end version by incumbent firms is most likely to occur in settings where competition is not very intense. Viewed from the perspective of the FE condition, this is associated with a small market size (small M) and/or large values of the entry cost, f_e.

The second component of the comparative static outcome, $\frac{dc_B}{dM} \leq 0$ if $M \geq M^*$, is consistent with the predictions from existing multi-product trade models and is depicted in Fig. 6. This corresponds to a situation where an increase in market size decreases both c_D, selection into the market gets tougher, and c_B, selection into the lower end of the market also gets tougher. The first effect is naturally associated with exit by high cost firms. The second effect involves product line pruning by a set of firms with intermediate productivity. Once again, the response is heterogeneous across firms.

A comparison of Figs. 5 and 6 brings out an implicit feature of Proposition 1 – seemingly identical shocks can lead the same firm (as indexed by c) in one case to extend their product lines and then reduce their product lines in an apparently similar environment. This has some interesting implications. For example, consider a firm with a cost draw between c_B and c_B' – which defines the same set of viable firms in the two figures. The increase in market size leads all of these firms to add an option for the low income end of the market, even though their marginal costs haven't changed – Fig. 5. However, a further increase in market
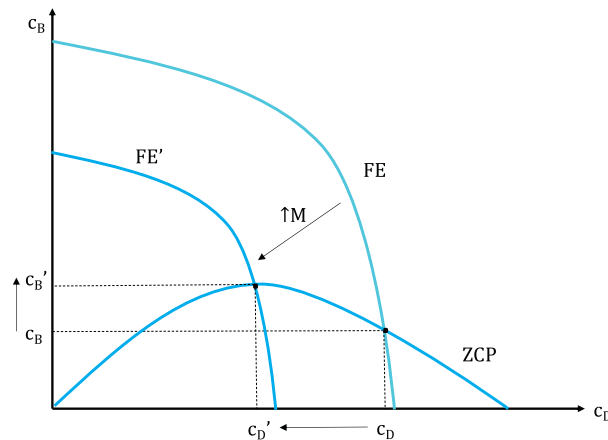
Fig. 5. Product line extensions. An increase in market size shifts the *FE* inward but does not effect the *ZCP*. A larger market causes c_D to fall to c_D', lowering the choke price and profit in the income high market segment. Since the high income segment is less lucrative, this reduces the implicit cost (information rents conceded) of serving the low income segment, inducing any firm with a cost between c_B and c_B' to extend their product lines by introducing a version of the product aimed at low income consumers. Product line extensions tend to occur when the initial difference between c_D and c_B is pronounced.
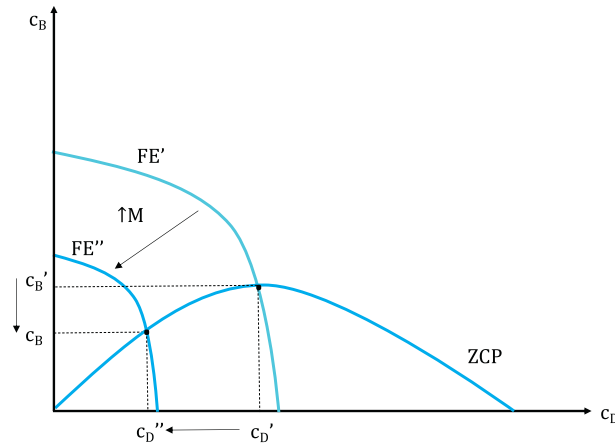
Fig. 6. Product line contractions. When the initial difference between $c_D{}'$ and $c_B{}'$ is not very large, the implicit cost of serving the low income segment is also small. Consequently, the maximum willingness to pay for a low income type, θ^L, is the dominant factor determining $c_{B'} = \theta^L - \frac{\beta}{1-\beta}(\theta^H - \theta^L)$. Since a larger market size is associated with entry in both market segments, θ^L must decline and induce a drop from $c_{B'}$ to c_B in the low income segment cost threshold. In this case, selection considerations govern behavior.

size (Fig. 6) leads exactly the same firms to eliminate their offering to the low income consumers. Analysts that follow these firms might conclude from this sequence of events that these firms mistakenly entered the low end of the market and the withdrawal is confirmation of the error. GM was subject to exactly this criticism when it launched the Saturn range (with early success despite the general scepticism) only to later dissolve the Saturn nameplate. Rather than reflecting a poor initial decision, the logic articulated in this section suggests that such behavior is a natural feature associated with an increasingly competitive environment, especially when the initial situation embodied little competitive pressure.

4.5. Efficient outcome

To gain insight into the forces at work, consider the first best outcome. A benevolent planner who chooses the number of varieties and their output levels so as to maximize the social welfare function given by the frequency and utility of each type times the number of consumers M, subject to each varieties production function and the mechanism that determines each variety's marginal cost as a random draw from $G(c)$ after f_e units of labor have been allocated to R&D.

Specifically, the planner chooses the number n^e of R&D projects and the output levels of varieties, $\{q^H(c), q^L(c)\}$, so as to maximize social welfare:

$$
\begin{aligned}
W = M\Bigg[&\beta\frac{\alpha^H}{\rho} \log\left(1 + n^e G(c_D) \int_0^{c_D} \left(q^H(c)+1\right)^\rho dG(c) - n^e G(c_D) \right) - \beta n^e G(c_D) \int_0^{c_D} c q^H(c) dG(c) \\
&+(1-\beta)\frac{\alpha^L}{\rho} \log\left(1 + n^e G(c_B) \int_0^{c_B} \left(q^L(c)+1\right)^\rho dG(c) - n^e G(c_B) \right) - (1-\beta) n^e G(c_B) \int_0^{c_B} c q^L(c) dG(c) \Bigg] - n^e f_e
\end{aligned}
$$

The first order condition that governs the choice of $q^H(c)$ is,

$$
\frac{\partial W}{\partial q^H(c)} = \frac{\alpha^H}{Q_H^\rho}\left(q^H(c)+1\right)^{\rho-1} = c \tag{22}
$$

where $Q_H^\rho = (1 + n\int_0^{c_D}(q^H(c)+1)^\rho\, g(c)dc\, n)$. The high types maximum willingness to pay is given by $\theta^H = \frac{\alpha^H}{Q_H^\rho}$, and a planner will not allocate resources to a product when c is above θ^H. Consequently the cut-off for operating is $c_D = \theta^H$. When the planner does allocate resources to a variety, it does so in the following way:

$$
c_D\left(q^H(c)+1\right)^{\rho-1} = c \;\Rightarrow\; \left(q^H(c)+1\right)^\rho = \left(\frac{c_D}{c}\right)^{\sigma-1} \tag{23}
$$

Analogous steps deliver

$$c_B\left(q^L(c)+1\right)^{\rho-1} = c \;\Rightarrow\; \left(q^L(c)+1\right)^{\rho} = \left(\frac{c_B}{c}\right)^{\sigma-1} \tag{24}$$

Using these rules, the planner's objective function becomes:

$$W = M\left[\beta\frac{\alpha^H}{\rho}\log\left(1+n^e G(c_D)\left(c_D^{\sigma-1}\overline{c^{1-\sigma}}-1\right)\right) - \beta n^e G(c_D)\left(c_D^{\sigma}\overline{c^{1-\sigma}}-\overline{c}\right)\right.$$
$$\left. + (1-\beta)\frac{\alpha^L}{\rho}\log\left(1+n^e G(c_B)\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}}-1\right)\right) - (1-\beta)n^e G(c_B)\left(c_D^{\sigma}\overline{c_L^{1-\sigma}}-\overline{c}_L\right)\right] - n^e f_e$$

The planner now chooses n_e to satisfy:

$$\frac{\partial W}{\partial n^e} = M\left(\frac{\beta\alpha^H}{\rho Q_H^{\rho}}G(c_D)\left(c_D^{\sigma-1}\overline{c^{1-\sigma}}-1\right) - \beta G(c_D)\left(c_D^{\sigma}\overline{c^{1-\sigma}}-\overline{c}\right)\right.$$
$$\left. + \frac{(1-\beta)\alpha^L}{\rho}\frac{1}{Q_L^{\rho}}G(c_B)\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}}-1\right) - (1-\beta)G(c_B)\left(c_B^{\sigma}\overline{c_L^{1-\sigma}}-\overline{c}_L\right)\right) - f_e = 0$$

Using $c_D = \frac{\alpha^H}{Q_H^{\rho}}$ and $c_B = \frac{\alpha^L}{Q_L^{\rho}}$, this condition reduces to:

$$\beta\left(\frac{c_D}{\sigma-1}\left(c_D^{\sigma-1}\overline{c^{1-\sigma}}-1\right)-(c_D-\overline{c})\right)G(c_D) + (1-\beta)\left(\frac{c_B}{\sigma-1}\left(c_B^{\sigma-1}\overline{c_L^{1-\sigma}}-1\right)-(c_B-\overline{c}_L)\right)G(c_B) = \frac{f_e}{M} \tag{25}$$

A comparison with (FE) confirms that the free entry condition is the same under both second degree price discrimination and the first best.

The market segment servicing decisions of the planner embodied in (23) and (24) also parallel those of firms engaged in second degree price discrimination, albeit where such firms account for explicit and implicit costs while the planner is only concerned with explicit costs. This has two immediate implications for our analysis. First, $c_B = \theta^L$ for an efficient outcome. And, second, as a consequence of the first, we have the following zero-profit cut-off condition for serving the low end of the market:

$$\frac{\left(\alpha^L - c_B\right)}{c_B} = \left(\frac{\alpha^H - c_D}{c_D}\right)\frac{G(c_B)}{G(c_D)}\left(\frac{c_B^{\sigma-1}\overline{c_L^{1-\sigma}}-1}{c_D^{\sigma-1}\overline{c^{1-\sigma}}-1}\right) \tag{ZCP FB}$$

This condition maps out a positive monotonic relationship between c_D and c_B and has the property that for any given c_D, the implied c_B is larger than the corresponding value that satisfies (ZCP). This property is driven by the "payment" of information
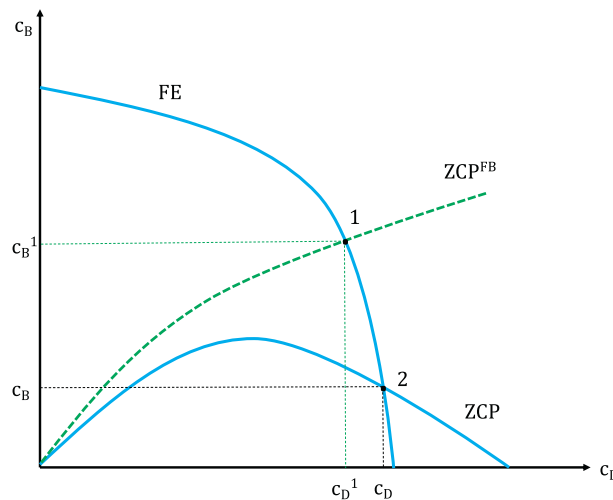
Fig. 7. Equilibrium and optimal cost thresholds. The efficient outcome occurs at 1, while the equilibrium is at 2. Since *FE* is common to both but ZCP^{FB} lies above ZCP, it follows that $c_D^1 < c_D$ and $c_B^1 > c_B$. This ordering implies the market equilibrium has too few firms in total and also too few serving the low income segment. This also identifies a distortion within firms, with high types over-served, and across firms, as the scale of high productivity firms is too small while the scale of low productivity firms is too large.

rents under second degree price discrimination, which increases the cost of serving the low end to be not only the marginal cost draw (explicit cost) but also the information rents (implicit cost). It is the absence of the implicit cost that establishes the relative position of the two curves.

This allows us to immediately conclude:

Proposition 2. The second degree price discrimination equilibrium delivers a c_D higher than the first best outcome, while the opposite holds for c_B. As a result there is less entry and fewer varieties under discrimination than efficiency would dictate. In addition, the average firm size is smaller than optimal, even though the amount devoted to the high type is larger than optimal (misallocation across types) for any firm that produces.

Fig. 7 compares the price discrimination equilibrium with the first best outcome (c_D^1, c_B^1). The over-service of the high type and the under-service of the low type follows from $c_D^1 < c_D$ and $c_B < c_B^1$ and (9) and (10). Combining this result with (FE) implies that the proportional increase in c_D is less than the proportional decrease in c_B, when comparing price discrimination relative to the first best. It follows that the average firm size must be below the first best.[17] In addition, there is a misallocation of output across firms, with high productivity firms under-producing and low-productivity firms over-producing relative to the efficient outcome.

Based on these preliminaries, we can now see that the first best response when M increases is for product lines to be rationalized by a set of firms with intermediate productivity. As the market size increases, a social planner assigns more firms to the industry, which also increases the number of high productivity firms. A social planner would require these firms to serve the low end of the market (along with the high end) since they are the most efficient. Effectively, when market size expands a planner is switching out a set of low productivity firms for a larger set of high productivity firms; the low end of the market is served by more and better firms. This mirrors what is happening at the high end of the market, albeit on a smaller scale.

The rationalizing of product lines described in the second part of Proposition 1 when firms engage in second degree price discrimination most closely aligns with that of first best behavior. This isn't surprising since it arises when markets are relatively large and competitive. It also suggests that the welfare outcomes in this scenario are closest to the first best, especially at the low end of the market. However, what are the welfare effects more generally? Before addressing this question I consider the particularly elegant system that comes from a specific paramterization of $G(c)$ that will prove helpful below.

4.6. Pareto distribution

As a parametric example suppose that cost draws are from a Pareto/Power distribution: $G(c) = \left(\frac{c}{c_M}\right)^k$ and $g(c) = \frac{k}{c}G(c)$ where $k \geq 1$ and c_M is the upper bound on the cost draws. Revisiting the equilibrium conditions, we now see that the thresholds that imply zero expected profits must satisfy:

$$\beta c_D^{k+1} + (1-\beta)c_B^{k+1} = \frac{\phi}{M} \tag{26}$$

where $\phi = \frac{(k+1)(k-(\sigma-1))c_M^k f_e}{\sigma}$, is an index of technology that combines the effects of a better distribution of cost draws (lower c_M) and lower entry costs, f_e. This also requires that $k > \sigma - 1$.

Turning to the market servicing (ex post) condition, we note that the Pareto assumption implies: $\overline{c_L^{1-\sigma}} = \frac{k}{k+1-\sigma}c_B^{1-\sigma}$ and $\overline{c^{1-\sigma}} = \frac{k}{k+1-\sigma}c_D^{1-\sigma}$. Using these results, the ex post condition (ZCP) simplifies to:

$$\frac{\alpha^L - (\beta c_D + (1-\beta)c_B)}{\beta c_D + (1-\beta)c_B} = \left(\frac{\alpha^H - c_D}{c_D}\right)\left(\frac{c_B}{c_D}\right)^k \tag{27}$$

While the ex post condition for the efficient outcome simplifies to:

$$\frac{\alpha^L - c_B}{c_B} = \left(\frac{\alpha^H - c_D}{c_D}\right)\left(\frac{c_B}{c_D}\right)^k \tag{28}$$

4.7. Welfare

A feature worth highlighting is that for the low income group, each firm is able to fully extract the surplus under the residual demand function. This might give the impression that those with a low income derive no net benefit from consuming any of the differentiated goods. However, this misses the fact that a firm is only able to extract surplus at the margin; each firm views themselves as the marginal firm (i.e. takes industry output in each segment as given). Intuitively, the utility function allows the varieties to interact with one another to generate welfare. Since an individual firm takes the output of all other firms as given, they

---

[17] A property of the mean is that it is homogeneous of degree one.

don't account for this interaction. It is this component that generates positive net surplus for a low income consumers from consumption of the differentiated goods. This can be seem most clearly by deriving the consumer surplus of a low income type.

$$
\begin{aligned}
CS_L &= \alpha^L V(Q_L) - T_L \\
&= \frac{\alpha^L}{\rho} \log\left(1 + \int_{\omega \in \Omega^L} (q(\omega)+1)^{\frac{1}{\rho}} d\omega - \int_{\omega \in \Omega^L} d\omega\right) - \int_{\omega \in \Omega^L} T_L(\omega) d\omega \\
&= \frac{\alpha^L}{\rho} \log\left(1 + n^L\left(\overline{(q_L+1)^\rho}-1\right)\right) - \frac{\alpha^L}{\rho} \frac{n^L\left(\overline{(q_L+1)^\rho}-1\right)}{\left(1 + n^L\left(\overline{(q_L+1)^\rho}-1\right)\right)}
\end{aligned}
$$

where the third line uses (5) and $\overline{(q_L+1)^\rho}$ is an expected value. Also recall that (3) can be used to identify the associated number of firms serving the low end of the market:

$$
n^L = \frac{\alpha^L - \theta^L}{\theta^L} \frac{1}{(q_L+1)^\rho - 1}, \tag{29}
$$

so that,

$$
CS_L = \frac{\alpha^L}{\rho}\left(\log\left(1 + \frac{\alpha^L - \theta^L}{\theta^L}\right) - \frac{\alpha^L - \theta^L}{\theta^L}\left(\frac{\theta^L}{\alpha^L}\right)\right) \tag{30}
$$

Since $\alpha^L \geq \theta^L$, it follows that $CS^L \geq 0$, where equality only holds when $\alpha^L = \theta^L$. Heuristically, $\alpha^L V(Q_L)$ is an increasing concave function of n_L, which means that the average surplus of a given n_L is greater than the marginal surplus evaluated at that n_L. So even though a firm can fully extract the marginal surplus it creates, this doesn't resign a low income type to a consumer surplus of zero. Instead, it is the "love of variety" property of the preferences that allows a low type to capture positive consumer surplus.

Moreover, this discussion applies equally to the first best outcome which coincides with first degree price discrimination. In this case, consumer surplus for both income groups has the same form as (30). The only situation where an income group receives an additional form of surplus is under second degree price discrimination where the high income type also captures information rents. Nevertheless, the consumer surplus for the high income group in this case can be derived along similar lines:

$$
\begin{aligned}
CS_H &= \alpha^L V(Q_L) - T_H \\
&= \frac{\alpha^H}{\rho}\left(\log\left(1 + \frac{\alpha^H - c_D}{c_D}\right) - \frac{\alpha^H - c_D}{c_D}\left(\frac{c_D}{\alpha^H}\right)(1-\delta)\right),
\end{aligned} \tag{31}
$$

where $\delta = (1-\beta)\left(1 - \frac{c_B}{c_D}\right)\frac{G(c_B)}{G(c_D)}\left(\frac{c_B^{\sigma-1}\overline{c_L^{\sigma-1}}-1}{c_D^{\sigma-1}\overline{c^{\sigma-1}}-1}\right)$.

Written in this way, the consumer surplus expressions, (30) and (31), embody the outcomes for both the second degree price discrimination and first best cases, and provide a mapping between the cut-offs and welfare outcomes. This mapping is contingent on the allocative mechanism. For the low income group, $\theta^L = \beta c_D + (1-\beta)c_B$ under second degree price discrimination and $\theta^L = c_B^1$ under the first best. For the high income types, $\theta^H = c_D$ under second degree price discrimination and $\theta^H = c_D^1$ under the first best. However, $\delta \in (0,1)$ when second degree price discrimination is practised, while $\delta = 0$ under the first best. The three sources of difference (cut-offs, the definition of θ^L and δ) all contribute to the welfare outcomes diverging from the first best. Whether these differences are diminished as market size expands turns on whether or not fighter brands are introduced.

As a benchmark consider how efficiency dictates welfare should change for each group as market size expands. This also isolates the outcomes when only the selection mechanism is present as $\frac{dc_B^1}{dM} < 0$ and $\frac{dc_D^1}{dM} < 0$. From (30) and (31) it is also apparent that the welfare of each income group just depends on the cut-off for serving that group. Straightforward differentiation of these expressions confirms,

$$
\frac{dCS_L}{dM} = -\left(\frac{\alpha^L - c_B^1}{\alpha^L c_B^1}\right)\frac{dc_B^1}{dM} > 0 \quad \& \quad \frac{dCS_H}{dM} = -\left(\frac{\alpha^H - c_D^1}{\alpha^H c_D^1}\right)\frac{dc_D^1}{dM} > 0. \tag{32}
$$

A feature which differentiates the response to a change in market size under second degree price discrimination is that the welfare of each group is a function of not only the cut-off required for serving that group but, also the threshold cost of serving the other group. This adds a layer of richness to the welfare analysis that makes general statements about welfare responses

difficult to characterize. While it can be shown that both income groups do gain from an increase in market size, an interesting question is how this change compares to the efficient path?

The welfare change for the low income group is especially interesting since they are most directly affected by the non-monotonicity associated with increases in market size. When considering $M < M^*$, we know from Proposition 1 that, despite overall selection into the market, $\frac{dc_D}{dM} < 0$, no firm exits the low end of the market, on the contrary, firms only enter this market segment. In this setting of a small initial market size, firms exercise their market power by not serving the low end, and when they do, it is with an offering inferior to the first best. That is, both n^L and $\overline{q}_L$ tend to be small, which in turn means that $n^L((\overline{q_L + 1})^\rho - 1) = \frac{\alpha^L - \theta^L}{\theta^L}$ must also be "small". These characteristics allow us to take advantage of an approximation that clarifies the relative importance of the extension and selection mechanisms.

$$
\begin{aligned}
\alpha^L V(Q_L) &= \frac{\alpha^L}{\rho} \log\left(1 + \frac{\alpha^L - \theta^L}{\theta^L}\right) \\
&\approx \frac{\alpha^L}{\rho}\left(\frac{\alpha^L - \theta^L}{\theta^L}\right) \equiv V^L
\end{aligned}
\tag{33}
$$

were the approximation is better the closer θ^L is to α^L. Note that V^L is used to make the distinction from $V(Q_L)$ explicit. When these conditions apply, a low income type's consumer surplus is approximated by:

$$
\begin{aligned}
CS_L^{2nd} &\approx V^L - n^L \overline{T}^L = \frac{\alpha^L}{\rho}\left(\frac{\alpha^L - \theta^L}{\theta^L}\right) - \frac{\alpha^L - \theta^L}{\rho} \\
&= \frac{\left(\alpha^L - \theta^L\right)^2}{\rho \theta^L} \\
&= \frac{\theta^L}{\rho}\left(\frac{\left(\alpha^H - c_D\right)}{c_D}\frac{G(c_B)}{G(c_D)}\left(\frac{c_B^{\sigma-1}\overline{c_L^{\sigma-1}} - 1}{c_D^{\sigma-1}\overline{c^{\sigma-1}} - 1}\right)\right)^2
\end{aligned}
$$

where the last line uses the (ZCP).

Adopting the Pareto distribution allows this expression to be simplified as:

$$
V^L - n^L \overline{T}^L = \frac{\theta^L}{\rho}\left(\frac{\left(\alpha^H - c_D\right)}{c_D}\left(\frac{c_B}{c_D}\right)^k\right)^2.
\tag{34}
$$

A similar approximation applies to the first best outcome for a low type:

$$
CS_L^{1st} \approx V^L - n_L^1 \overline{T}_L^1 = \frac{c_B^1}{\rho}\left(\frac{\left(\alpha^H - c_D^1\right)}{c_D^1}\left(\frac{c_B^1}{c_D^1}\right)^k\right)^2.
\tag{35}
$$

Relative welfare for a low income individual can then be approximated by:

$$
\frac{CS_L^{2nd}}{CS_L^{1st}} = \frac{\theta^L}{c_B}\left(\frac{c_B}{c_B^1}\right)^{2k+1}\left(\frac{\alpha^H - c_D}{\alpha^H - c_D^1}\left(\frac{c_D^1}{c_D}\right)^{k+1}\right)^2
\tag{36}
$$

While this says that the welfare of the low income group under indirect discrimination is below the first best, it also implies:

$$
\widehat{CS}_L^{2nd} - \widehat{CS}_L^{1st} = s_H(\hat{c}_D - \hat{c}_B) + (2k+1)\left(\hat{c}_B - \hat{c}_B^1\right) + \psi \hat{c}_D - \psi^1 \hat{c}_D^1
\tag{37}
$$

where $\hat{x}$ denotes a proportional change in x, with $s_H = \frac{\beta c_D}{\theta^L}$, $\psi = \frac{(k+1)\alpha^H - kc_D}{\alpha^H - c_D}$ and $\psi^1 = \frac{(k+1)\alpha^H - kc_D^1}{\alpha^H - c_D^1}$. This can be further simplified since $\psi\hat{c}_D - \psi^1\hat{c}_D^1$ tends to be small and is zero when $c_D \approx c_D^1$ – which occurs in the short-run (see below) or if the Pareto parameter, k, is "large". In this case:

$$\widehat{CS}_L^{2nd} - \widehat{CS}_L^{1st} = s_H(\hat{c}_D - \hat{c}_B) + (2k+1)\left(\hat{c}_B - \hat{c}_B^1\right) \tag{38}$$

This equation reveals two things. First, it highlights the role of firm heterogeneity (as measured by k) while the product differentiation parameter does not appear (as measured by σ). Second, the approximation is most accurate when the initial M is small, hence both θ^L and c_B^1 are close to α^L. Consequently, an increase in M results in a subset of firms extending their product lines, which delivers: $\hat{c}_D < 0$, $\hat{c}_B > 0$ and $\hat{c}_B^1 < 0$. These product line extensions dis-proportionally benefit the low end of the income distribution. For example, in the extension range there must exist a point where $|\hat{c}_D| = |\hat{c}_B| = \hat{c} >> |\hat{c}_B^1| > 0$, then $\widehat{CS}_L^{2nd} - \widehat{CS}_L^{1st} = (2k-1)\hat{c} + (2k+1)|\hat{c}_B^1|$. This translates into disproportionately large gains for the low income group as k increases.

The pro-competitive benefits to the low types when product lines are extended reflect the reduction in a distortion that negatively affects the low income groups. However, it is associated with the capture of information rents by the high income groups – which is not part of an efficient outcome. This is reflected by δ in (31) which is non-monotonic in market size due to the non-monotonicity of c_B in M. In particular, a sufficiently small M leads to $c_B \to 0$ and $\delta \to 0$, while a sufficiently large M results in $c_B \to c_D$ and $\delta \to 0$. Once again the Pareto distribution provides an especially neat characterization:

$$\delta = (1-\beta)(1-\gamma)\gamma^k$$

where $\gamma = \frac{c_B}{c_D} \in (0,1)$. This is a strictly concave function which attains a maximum when $\gamma = \sqrt{\frac{k}{k+1}}$.

The aggregate consequences for the efficiency of second degree price discrimination are also potentially non-monotonic. Table 1 provides an example where the proportional change in welfare from an increase in market size under second degree price discrimination can at times lag that of the first best, but at other times exceed it. While it would be convenient to restrict parameters and rule out the small number of cases where the (relative) inefficiency increases with market size, this would require ruling out values of k that fall in the empirically relevant range.

More generally changes in market size, M, are relatively difficult to engineer within a country. The most direct parallel to the above analysis is complete integration between two countries – free trade. Since this is also a relatively rare outcome, the next section considers the implications of positive trade costs and how equilibrium outcomes are shaped by reciprocal and unilateral trade liberalization.

5. Open economy and trade costs

When trade costs are introduced, the analogy between an increase in market size and trade liberalization becomes less precise. For example, with heterogeneous firms, trade costs imply that not all products are available in all markets. Indeed, when firms utilize product lines, the design and number of items offered by a firm can also vary across countries. To explore these issues consider two countries, h and f, with M^h and M^f consumers located in each country.

5.1. Internationally segmented markets

Assume that each national market is segmented. This captures the national jurisdiction of intellectual property rights which allow firms to limit arbitrage opportunities. More generally, differences in regulations across countries can also restrict international arbitrage – the automotive industry is a good example where arbitrage is ruled out due to differences in regulation across countries (see Freund and Oliver (2015)).

Table 1
$\widehat{CS}^{2nd} - \widehat{CS}^{1st}$ when market size increases.

			k					
			1	2	3	4	5	6
M	Small	$\hat{c}_B > 0$	+	+	+	+	−	−
	Medium	$\hat{c}_B = 0$	+	+	+	+	+	+
	Large	$\hat{c}_B < 0$	+	+	+	+	+	+

Shocks, $\hat{M} > 0$, are calibrated to first best $\widehat{CS}_H = 1\%$.
Parameters: $\alpha^L = 1$, $\alpha^H = 2$, $f_e = 1$, $c_M = \alpha^H$, $\beta = .6$.

Let trade costs be of the usual iceberg form; in order for one unit of q to arrive in the overseas location i, $\tau_i > 1$ units need to be shipped. Denoting the cut-offs in each location as c_D^i and c_B^i, then any firm that wants to serve i must have a cost below c_D^i, and if they want to serve the low income segment in i they need a cost draw below c_B^i. For an exporter these cut-offs are naturally inclusive of the transport costs.

Since a firm treats the two countries as segmented, the profits from optimally serving the local (π_d^i) and overseas (π_x^i) markets can be expressed as:

$$\pi_d^i = \left(\beta_i \left(\frac{c_D^i}{\sigma-1} \left(\left(\frac{c_D^i}{c} \right)^{\sigma-1} - 1 \right) - \left(c_D^i - c \right) \right) + \left\{ (1-\beta_i) \left(\frac{c_B^i}{\sigma-1} \left(\left(\frac{c_B^i}{c} \right)^{\sigma-1} - 1 \right) - \left(c_B^i - c \right) \right) \; | c_B^i \geq c \right\} \right) M^i$$

$$\pi_x^i = \left(\beta_j \left(\frac{c_D^j}{\sigma-1} \left(\left(\frac{c_D^j}{\tau_j c} \right)^{\sigma-1} - 1 \right) - \left(c_D^j - \tau_j c \right) \right) + \left\{ (1-\beta_j) \left(\frac{c_B^j}{\sigma-1} \left(\left(\frac{c_B^j}{\tau_j c} \right)^{\sigma-1} - 1 \right) - \left(c_B^j - \tau_j c \right) \right) \; | c_B^j \geq \tau_j c \right\} \right) M^j.$$

Entry is unrestricted in both countries and firms choose a production location prior to entry and pay the sunk entry cost. In order to focus our analysis on the effects of market size and trade cost differences, we assume that countries share the same technology referenced by the entry cost f_e and cost distribution $G(c)$. To further simplify the expressions I will make use of the Pareto/power specification for $G(c)$.

Free entry of domestic firms in country i implies zero expected profits in equilibrium, hence:

$$M^i \left(\beta_i \left(c_D^i \right)^{k+1} + (1-\beta_i) \left(c_B^i \right)^{k+1} \right) + \nu_j M^i \left(\beta_j \left(c_D^j \right)^{k+1} + \left(1-\beta_j \right) \left(c_B^j \right)^{k+1} \right) = \phi$$

where $\nu_j = \left(1 - \dfrac{k\left((k+1-\sigma)(1-\tau_j) + \frac{(k+1)}{(\sigma-1)}(1-\tau_j^{1-\sigma})\right)}{\sigma} \right)$.

To help characterize the implications of free entry, let $C_i = M^i(\beta_i(c_D^i)^{k+1} + (1 - \beta_i)(c_B^i)^{k+1})$. Hence, the system of free entry conditions can be written as:

$$C_i + \nu_j C_j = \phi \tag{39}$$

$$\nu_i C_i + C_j = \phi \tag{40}$$

When firms are active in both locations the solution is $C_i = \left(\dfrac{1-\nu_j}{1-\nu_j \nu_i} \right) \phi$ and $C_j = \left(\dfrac{1-\nu_i}{1-\nu_j \nu_i} \right) \phi$. Consequently, we can derive a compact free entry condition for each market:

$$\beta_i \left(c_D^i \right)^{k+1} + (1-\beta_i) \left(c_B^i \right)^{k+1} = \left(\frac{1-\nu_j}{1-\nu_j \nu_i} \right) \frac{\phi}{M_i} \tag{41}$$

Fig. 8. Exporter product line extensions and reciprocal trade liberalization. Reciprocal trade liberalization induces the thresholds dictating export market participation to move from A to C. This can be decomposed into improved market access, A to B, and a greater incentive to serve the low income segment due to a lower implicit cost (lower information rents), B to C. The move from B to C reflects a pro-competitive effect.

5.2. *Eqn 41 Reciprocal trade liberalization*

The comparative static implications of this general formulation can be quite rich. Not only is location choice influenced by differences in market size, but also asymmetries in trade costs and the distribution of consumer types. To help isolate the role of trade costs on product line design assume: $\beta_i = \beta_j$, $M_i = M_j$, $\tau_i = \tau_j > 1$. These symmetry assumptions rule out home market effects and (Eqn 41) becomes:

$$\beta c_D^{k+1} + (1-\beta)c_B^{k+1} = \frac{\phi}{M(1+\nu)} \tag{42}$$

Consequently, when market size and trade costs are symmetric, the free entry condition is a straightforward generalization of (26). It is evident that reciprocal changes in trade costs vary the position of the free entry condition just like variation in market size. Since trade costs don't alter the threshold cut-off cost conditions for positive production in either market segment, the zero cut-off profit condition remains the same and is given by (27).

Due to the similarity of (42) and (26), we are immediately able to conclude that a reciprocal lowering of transport costs can be consistent with product line extensions. In particular, this is more likely to occur if trade costs are initially high and/or M is small.

Proposition 3. For a given set of parameters $\{\alpha^l, \beta, k, \phi\}$, there exists an M^* such that for $M(1+\nu) \in [\underline{M}, M^*]$, $\frac{dc_B}{d\nu} > 0$. That is, a set of firms will extend their product lines to serve the low type as trade barriers are reduced. For $M(1+\nu) > M^*$, then $\frac{dc_B}{d\nu} < 0$; some firms trim product lines as trade costs fall.

Compared to Proposition 1 this result introduces an interaction between domestic market size and trade barriers. If the domestic market size is relatively large, then a reciprocal reduction in trade barriers has a conventional impact on the thresholds required to serve a given market segment (i.e. $\frac{dc_B}{d\nu} < 0$ and $\frac{dc_D}{d\nu} < 0$). However, if the domestic market size falls below the threshold defined by Proposition 1 and trade barriers are sufficiently high, then reciprocal liberalization leads to a set of firms to add a low income option to their menu, even though they now face tougher competition (i.e. $\frac{dc_B}{d\nu} > 0$ and $\frac{dc_D}{d\nu} < 0$).

A natural question is which firms add a low end option; and more generally, what are the dynamics of product line redesign when trade barriers are reciprocally reduced? To gain insight into these questions, split firms into local producers (serving local consumers only) and exporters. It is clear that the set of firms adding a low end version includes local producers with the behavior of these firms characterized by c_D and c_B. Moreover, since these firms don't benefit from improved market access abroad, this product line extension is purely a response to greater competition.

To isolate the behavior of c_x and c_{Bx}, we can use $c_x = c_D/\tau$, $c_{Bx} = c_B/\tau$ and (42) to derive an export market free entry condition:

$$\beta c_x^{k+1} + (1-\beta)c_{Bx}^{k+1} = \left(\frac{\tau^{-(k+1)}}{1+\nu}\right)\frac{\phi}{M}$$

Similarly the exporter ZCP can be written as:

$$\frac{\alpha^L/\tau - (\beta c_x + (1-\beta)c_{Bx})}{\beta c_x + (1-\beta)c_{Bx}} = \frac{\alpha^H/\tau - c_x}{c_x}\left(\frac{c_{Bx}}{c_x}\right)^k \tag{43}$$

These conditions are just re-scaled versions of (27) and (42). Nevertheless, they clarify the difference between changes in market access due to lower trade barriers, $\hat{\tau}$, and changes in equilibrium thresholds, $\hat{c}_B$ and $\hat{c}_D$. Fig. 8 depicts the changes in exporter cut-offs. Point A represents the initial position, and the distance AB captures the increase in market access if the decline in trade costs were the only source of variation. Market access improves proportionally in both segments. The shift from B to C shows the net impact after allowing for an equilibrium response to the reduction in trade costs. Market access at the high end is mitigated by overall tougher selection into the market, $\hat{c}_D < 0$. In contrast, market access to the low end improves further since $\hat{c}_B > 0$. This

Table 2
$\widehat{CS}_L^{2nd}$ - $\widehat{CS}_L^{1st}$ reciprocal trade liberalization.

			k					
			1	2	3	4	5	6
M	*Small*	$\hat{c}_B > 0$	1.8	3.4	5.5	7	11	12
	Medium	$\hat{c}_B = 0$	1.2	1.3	1.9	2.2	2.9	3.4
	Large	$\hat{c}_B < 0$	0.6	0.65	0.67	0.7	0.72	1.1

Shocks, $\hat{\tau} < 0$, are calibrated to first best $\widehat{CS}_H = 1\%$.
Parameters: $\alpha^L = 1$, $\alpha^H = 2$, $f_e = 1$, $c_M = \alpha^H$, $\beta = .6$.

says that exporting at the low end improves not only because of a reduction in trade costs but also because of a reduction in monopoly power. Once again the mechanism is familiar. Greater entry/competition in the high market segment reduces the rents extracted from high types, lowering the implicit cost of serving the low end of the market. In this instance, a reduction in trade costs is the catalyst for not only improved market access but also a pronounced pro-competitive effect that leads both domestic and foreign firms to serve the low income market segment to a greater extent.

6. Gains from trade liberalization

Since the changes in cut-offs mirror those of an increase in market size, it follows that the welfare outcomes are also similar. In particular, while the high income groups always gain from selection (decrease in c_D), the low income group gains disproportionately when reciprocal liberalization generates an increase in c_B – existing firms have an incentive to enter and compete in the low income segment. This is most likely to occur when trade barriers are high and the domestic market is relatively small, circumstances which impose a high implicit cost of serving the low income group causing firms to concentrate on extracting rents from the high income consumers.

To explore this issue we start with the same underlying parameter values and consider a shock to trade costs, $\hat{\tau} < 0$, that delivers a 1% increase in the high income groups welfare under the **first best**. Note that the associated welfare gain for the low income group is greater than 1%. These benefits are derived exclusively from the selection mechanism that restricts the firms in each market to have a marginal cost below a progressively lower cut-off for each segment.

How does the low income group fare when the allocation mechanism is second degree price discrimination, which admits the possibility of both selection and extension forces, each of which can dominate in different settings? Table 2 presents the relative welfare implications of reciprocal trade liberalization for the low income group as the size of the Pareto shape parameter, k, and the initial market size vary. To set the scene, recall that any increase in market size generates selection effects $\hat{c}_B^1 < 0$ and $\hat{c}_D < 0$, while the new dimension relates to $\hat{c}_B \gtrless 0$.

When the initial market size is large, both models predict that trade liberalization tightens selection into the low end of the market $\hat{c}^B < 0$. The relative welfare change in this scenario are displayed in the third row which shows that the differences across the two allocation mechanisms are modest. Similarly, for an initial market size associated with $\hat{c}^B = 0$, the proportional change in consumer surplus is higher under second degree price discrimination and increases with k.

However, the most dramatic changes occur when $\hat{c}^B > 0$.[18] This arises when initial market size is relatively small (implying high c_D and low c_B). As is clear from the first row, the welfare benefits to the low types in this case range from 1.8 to 12 percentage points above that predicted by a selection based model. To reiterate, for a shock that raises a low types welfare due to selection forces, the same shock is capable of increasing the low type's welfare by over 10 percentage points **more** if the liberalization results in some firms extending their product lines to serve low income groups. This suggests that even modest increases in competition can translate into large welfare gains for those at the lower end of the income distribution. Combining this with the insights of Table 2, we see that it is possible for the proportional increase in aggregate welfare to be below that associated with an efficient (selection based) outcome but the gains to the low income group to greatly exceed that predicted by selection alone. These distributional consequences are significant and not part of conventional thinking about the gains from trade.

6.1. Unilateral trade liberalization

If instead a country considers unilateral liberalization then it can face contrasting short and long run consequences. The long run welfare impact of unilateral trade liberalization can be particularly diverse, depending not only on the size of the domestic market but also on the scale of unilateral trade liberalization. This last feature, in particular, is under appreciated in the current literature. To isolate the different outcomes we follow Melitz and Ottaviano (2008) and begin our analysis with the short-run responses to unilateral liberalization by country i.
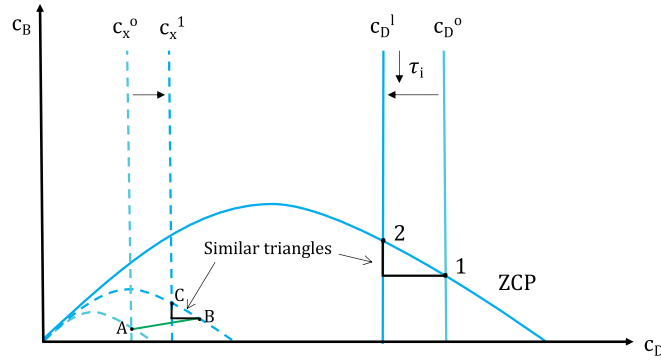
6.1.1. Short run

To make things especially clear, consider an initial equilibrium where countries are symmetric. In this case, all operating firms have $c \leq c_D^0$ (where the superscript 0 denotes the initial equilibrium value). In the short run there is no entry or exit (though firms can choose not to operate). This implies the maximum cost in both locations is given by $\bar{c}_M = c_D^0$. Defining $\bar{N}_D^i$ as the number of firms operating in the initial symmetric equilibrium we can then determine the number of firms that were initially serving country i as $\bar{N}_D^i + \tau_i^{-k}\bar{N}_D^j$.

If there is a negative shock where some firms shutdown (such as a decrease in τ_i to τ_i'), then

$$\left(\frac{k+\sigma-1}{\sigma-1}\right)\frac{\left(\alpha^H-c_D^i\right)}{c_D^i} = \left(\frac{c_D^i}{\bar{c}_M}\right)^k\left[\bar{N}_D^i + (\tau_i')^{-k}\bar{N}_D^j\right] \tag{44}$$

---

[18] To ensure that the results are not driven by an initial equilibrium where $c_B = 0$, and $\hat{c}_B$ can be unbounded, the underlying equilibrium in the first row is disciplined so that $\frac{c_B}{c_D} \geq .2$ for each k.

Fig. 9. Unilateral liberalization and product line extensions in the short run. In the short run *FE* is replaced by c_D^0 - the highest cost draw among incumbent firms consistent with positive output and the equilibrium is at point 1. An analogous condition for export market participation, c_x^0, also exists and the relevant trading thresholds are defined at A. If trade barriers are unilaterally reduced, the upper bound on viable costs shifts down to c_D^1 and the domestic thresholds are now defined by 2. The lower trade barriers induce a set of domestic firms to extend their product lines to the low income segment in a pro-competitive manner. Foreign exporters enter due to improved market access (transition from A to B) which would see a proportional improvement in access for both segments. These effects are further amplified for the low income segment by the movement from B to C. Consequently, the low income segment is better served due to both improved market access and a reinforcing pro-competitive effect.

This is the analogue of (41) in the short run.

Since unilateral trade liberalization by i increases τ_i^{-k}, this condition implies that c_D^i must decrease for any given c_B^i. Once again the ZCP/ex-post condition is unaffected by trade barriers and can be used to determined the equilibrium outcome for c_B^i. In particular, if the initial equilibrium coincides with the upward sloping segment of the ZCP, then product line extensions will be added by both local and exporting firms.

Fig. 9 provides an example where unilateral liberalization leads a set of both domestic and foreign exporters to add low end options to their product lines. A set of domestic firms adding options at the low end follows immediately from $\hat{c}_B^i > 0$. As with the reciprocal liberalization case, foreign exporters have an incentive to expand product lines purely because of improved market access, $\hat{\tau}_i < 0$. However, once again we see that there is an additional group of exporters that add low end options due to a decrease in market power ($\hat{c}_D^i < 0$) which provides greater incentive to serve the low end ($\hat{c}_B^i > 0$), implying $\hat{c}_{Bx}^i > |\hat{\tau}_i|$.

The short-run also provides an interesting setting in which to evaluate the level of welfare and how it changes, both in the aggregate and in composition. Since outcomes which feature product line extensions provide the most striking results, that's where I'll concentrate the analysis. Consider first the high income group and compare the welfare outcome under both the first best and second degree price discrimination starting from $c_D = c_D^0$ – the initial cut-off for serving the high income group which defines a fixed number of initially viable firms in the market. Since CS_H^{1st} is solely a function of c_D the first best welfare for the high income group is defined by (31). However, since $\delta < 1$ it immediately follows that the welfare of the high income group under second degree price discrimination exceeds the first best. Moreover, when $\frac{d\delta}{d\tau_i} < 0$, which must be true when product line extensions occur, then the increase in welfare for a high income consumer also exceeds the first best (the absolute gain is greater though the relative gain maybe smaller).

Naturally, there must be a downside since aggregate welfare can't be greater than the first best and the welfare of the low income group is well below the efficient level. Nevertheless, when product line extensions occur due to unilateral trade liberalization, it must be the case that the proportional change in welfare for the low income group is greater than if the selection mechanism alone operates – with (38) offering a concise approximation of the difference.

6.1.2. Long run

While the short run offers welfare gains for the liberalizing country, the long run can be a different matter. When Melitz and Ottaviano (2008) consider the long run implications of unilateral liberalization, they find as long as no country is specialized in the numeraire, the delocation effects of unilateral liberalization will reduce welfare of the liberalizing country. This property carries over to the present model using (41) and the fact that the ZCP isn't effected by changes in trade costs. However, the negative consequences of unilateral liberalization are also limited by the potential for delocation. Implicit in the Melitz and Ottaviano (2008) result is a requirement that either both countries are relatively large or if the liberalizing country is small, that unilateral liberalization occurs on a small scale. For this latter case, what happens when unilateral liberalization is large and the liberalizing country ends up specialized in the production of the numeraire good? As we'll see the consequences of unilateral liberalization no longer reflect the Melitz and Ottaviano (2008) prediction.[19]

---

[19] The Australian automotive industry provides a motivating example. In the 1980's Australian auto production received a 57.5% tariff (along with quantitative restrictions that carried tariffs of over 100%). The very existence of the industry was dependent on high trade barriers. It seems intuitive that the unilateral reduction in trade barriers should raise welfare, not only in the short run, but also the long run.

To explore this issue consider the relatively divisive situation where no one country is large enough to house a competitive industry in isolation (i.e. each country's market size and trade costs result in a symmetric outcome with cut-offs on the positively sloped portion of the ZCP). This is a setting tailor-made for nationalistic arguments about local industrial capability and the benefits which flow from maintaining its presence. Indeed, as noted above, unilateral liberalization can lead to welfare losses due to delocation. Moreover, delocation is extra costly in this setting since the extension mechanism operates in reverse. This would appear to be an situation ill-suited to unilateral liberalization, with liberalization best achieved through complex rounds of trade negotiations or only viable between liked-minded liberalization inclined countries.

However, this prediction is too pessimistic. While advocates of the benefits of liberalization can point to the short-run gains, they can also claim that the degree of unilateral liberalization is too limited. To see this recall we are in a setting where no country is large enough to have a competitive market. Let's also raise the stakes and say that the level of trade costs in an initial symmetric outcome improve on autarky but still present significant impediments to trade. This means that unilateral liberalization has the potential to not only lead to some reduction in the number of local firms, it can lead to the complete delocation of the domestic industry. This captures the worst case scenario feared by proponents of maintaining domestic industrial capacity. The complete delocation of the domestic industry in country i due to unilateral liberalization occurs at the highest τ_i that satisfies

$n_i^e = \frac{c_M^k}{1-(\tau_j\tau_i)^{-k}}\left[\frac{N^i}{(c_D^i)^k}-\tau_i^{-k}\frac{N^j}{(c_D^i)^k}\right] = 0$. Call this trade cost $\tilde{\tau}_i$. This trade cost results in both a higher c_D^i and a lower c_B^i than the

original pre-liberalization equilibrium, lowering welfare for all income groups in the liberalizing country.

However, this delocation "low point" also confirms that the liberalizing country has been too timid in its efforts. The complete delocation in country i implies that the free entry condition in country i no longer forms part of the equilibrium conditions. Instead, the equilibrium cut-offs are now derived from (40), (ZCP) – one for each country – and

$$\frac{\alpha^H - c_D^i}{(c_D^i)^{k+1}} = \tau_i^{-k}\left(\frac{\alpha^H - c_D^j}{(c_D^j)^{k+1}}\right) \tag{45}$$

This condition defines a negative relationship between τ_i^{-k} and c_D^i (i.e. unilateral liberalization decreases c_D^i). Furthermore, if $\tau^i \to 1$, then $CS^i \to CS^j$ (i.e. both locations have the same welfare). Since c_D^i is monotonically decreasing as country i unilaterally liberalizes, it immediately follows that (1) welfare is higher under unilateral free trade in country i than the initial symmetric trade cost equilibrium, (2) there exists a $\tau_i^* > 1$ where liberalization to this point leaves country i indifferent between the initial symmetric trade cost equilibrium and the asymmetric unilateral liberalization outcome. This suggests that gains from unilateral liberalization are most likely to arise if the degree of liberalization is sufficiently large.[20] Moreover, once agglomeration in country j is complete, τ_j plays no role in the equilibrium outcome. This provides scope for liberalization on the part of country j as well.

7. Conclusion

The standard prediction of international trade models is that increased integration leads to specialization/concentration of production. This mechanism has been utilized at the country, industry and firm level to gain many valuable insights. Nevertheless, the notion that international trade will lead firms to rationalize their product portfolios and concentrate on their "best" products doesn't always square with reality. On the contrary, there are important and prominent exceptions where firms extend their product range when confronted with more intense competition. These examples raise a series of questions. Why do some firms expand their offerings and not others? Why do we observe these extensions only some of the time, while we observe product line rationalization at other times? Do these outcomes have different welfare implications? Are these differences big?

This paper offers answers to all of these questions using a standard trade model augmented by consumer heterogeneity and populated by firms trying to leverage these differences to their advantage. Depending on parameter values that determine the degree of competition, tougher competition can be associated with either the standard prediction of product line rationalization or the contrasting outcome of product line extensions. That is, both types of behavior can arise in equilibrium. Since trade costs directly influence competitive pressure, their variation can have important implications for product line design. While any reciprocal liberalization generates efficiency gains, these welfare benefits are magnified greatly by the introduction of "fighter brands". In particular, the gains from trade in this case can be in excess of 10 percentage points higher than predicted by the standard framework for groups that are under-served.

These results provide some nuance to the "pro-competitive effect" that has proven so elusive. In particular, the standard selection/specialization mechanism speaks to a setting where competition is already likely to be intense. So while pro-competitive gains are possible, they maybe modest for all market segments. In contrast, when firms exercise market power by excluding or under-serving certain groups, the potential pro-competitive effect delivers much larger benefits to low income groups that might reasonably be described as impressive.

---

[20] Nevertheless, optimal unilateral tariffs under specialization are unlikely to be zero. See McCalman (2010).

Appendix A. Appendix

A.1. Proof that ZCP is non-monotonic

Rearranging (ZCP) as

$$\left(\alpha^L - \theta^L\right) c_D G(c_D) H(c_D) - \left(\alpha^H - c_D\right) \theta^L G(c_B) H(c_B) = 0 \tag{46}$$

where $H(c_D) = c_D^{\sigma-1}\overline{c^{1-\sigma}} - 1$ and $H(c_B) = c_B^{\sigma-1}\overline{c_L^{1-\sigma}} - 1$. Totally differentiate this condition:

$$
\begin{aligned}
dc_D \Big(&-\beta c_D G(c_D) H(c_D) + \left(\alpha^L - \theta^L\right) G(c_D) H(c_D) + \left(\alpha^L - \theta^L\right) c_D G'(c_D) H(c_D) \\
&+ \left(\alpha^L - \theta^L\right) c_D G(c_D) H'(c_D) + \theta^L G(c_B) H(c_B) - \beta\left(\alpha^H - c_D\right) G(c_B) H(c_B) \Big) \\
-dc_B \Big(&(1-\beta) c_D G(c_D) H(c_D) + (1-\beta)\left(\alpha^H - c_D\right) G(c_B) H(c_B) + \left(\alpha^H - c_D\right) \theta^L G'(c_B) H(c_B) \\
&+ \left(\alpha^H - c_D\right) \theta^L G(c_B) H'(c_B) \Big) = 0
\end{aligned}
\tag{47}
$$

From the (46), $c_B = 0$ implies that c_D is either 0 or $\frac{\alpha^L}{\beta}$. If the sign of the slope of (ZCP) differs at these two points, then the function is non-monotonic. From (47)

$$\left.\frac{dc_D}{dc_B}\right|_{c_B=0, c_D=\frac{\alpha^L}{\beta}} = -\frac{(1-\beta)}{\beta} \tag{48}$$

In contrast, using l'Hopital's rule and (47) confirms:

$$\left.\frac{dc_D}{dc_B}\right|_{c_B=0, c_D=0} > 0. \tag{49}$$

A.2. Proof of Proposition 1

From the (ZCP), $c_B = 0$ implies that c_D is either 0 or $\frac{\alpha^L}{\beta}$. As c_D approaches $\frac{\alpha^L}{\beta}$, the slope of (ZCP) approaches $-\frac{1-\beta}{\beta}$. Since an increase in M shifts the FE toward the origin, it follows that in this neighborhood, $\frac{dc_B}{dM} > 0$. However, as c_D approaches zero, the slope of (ZCP) is positive. Hence, increasing M from at starting point in the neighborhood of $\underline{M}$ eventually encounters a point where $\frac{dc_B}{dL} = 0$. Call the M where this first occurs M^*. Therefore, for $M^* > M > \underline{M}$ it must be $\frac{dc_B}{dM} > 0$.

References

Arkolakis, Costas, Costinot, Arnaud, Donaldson, Dave, Rodrguez-Clare, Andrés, 2018. The elusive pro-competitive effects of trade. Rev. Econ. Stud. 86 (1), 46–80.
Bernard, A., Redding, S., Schott, P., 2011. Multi-product firms and trade liberalization. Q. J. Econ. 1271–1318.
Deneckere, Raymond J., Preston McAfee, R., 1996. Damaged goods. J. Econ. Manag. Strategy 5 (2), 149–174.
Dhingra, Swati, 2013. Trading away wide brands for cheap brands. Am. Econ. Rev. 103 (6), 2554–2584.
Eckel, Carsten, Peter Neary, J., 2010. Multi-product firms and flexible manufacturing in the global economy. Rev. Econ. Stud. 77 (1), 188–217.
Fajgelbaum, P., Grossman, G.M., Helpman, E., 2011. Income distribution, product quality, and international trade. J. Polit. Econ. 119 (4), 721–765.
Flam, Harry, Helpman, Elhanan, 1987. Vertical product differentiation and north-south trade. Am. Econ. Rev. 810–822.
Freund, Caroline, Oliver, Sarah, 2015. Gains From Harmonizing us and eu Auto Regulations Under the Transatlantic Trade and Investment Partnership. Peterson Institute for International Economics (Policy Brief no. PB15-10).
Johnson, Justin P., Myatt, David P., 2003. Multiproduct quality competition: fighting brands and product line pruning. Am. Econ. Rev. 93 (3), 748–774.
Jung, Jae Wook, Simonovska, Ina, Weinberger, Ariel, 2019. Exporter heterogeneity and price discrimination: a quantitative view. J. Int. Econ. 116, 103–124.
Maskin, Eric, Riley, John, 1984. Monopoly with incomplete information. RAND J. Econ. 15 (2), 171–196.
Mayer, Thierry, Melitz, Marc J., Ottaviano, Gianmarco I.P., 2014. Market size, competition, and the product mix of exporters. Am. Econ. Rev. 104 (2), 495–536.
McCalman, Phillip, 2010. Trade policy in a "super size me" world. J. Int. Econ. 81 (2), 206–218.
McCalman, Phillip, 2018. International trade, income distribution and welfare. J. Int. Econ. 110, 1–15.
McCalman, Phillip, 2019. International Trade, Product Lines and Welfare: The Roles of Firm and Consumer Heterogeneity. University of Melbourne (Working paper).
Melitz, Marc J., Ottaviano, Gianmarco I.P., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75 (1), 295–316.
Moon, Youngme, 2004. The Birth of the Swatch. Harvard Business School, Boston, MA.
Nocke, Volker, Yeaple, Stephen, 2014. Globalization and multiproduct firms. Int. Econ. Rev. 55 (4), 993–1018.
Porter, Michael E., 1980. Competitive Strategy: Techniques for Analyzing Industries and Competitors. Simon and Schuster.
Ritson, Mark, 2009. Should you launch a fighter brand? Harv. Bus. Rev. 87 (10), 86–94.
Shapiro, Carl, Varian, Hal R., 1998. Versioning: the smart way to. Harv. Bus. Rev. 107 (6), 107.
Teisberg, Elizabeth, Clark, Theodore, 1994. The Desktop Printer Industry in 1990. Harvard Business School Case 9-390-173.
Tirole, Jean, 1988. The Theory of Industrial Organization. MIT Press.
Tushman, Michael, Radov, Daniel, 2000. Rebirth of the Swiss Watch Industry, 1980–1992 (B): Hayek and Thomke at SMH. Harvard Business School.