Trade Policy in the Presence of a Discriminating Foreign Monopolist

Phillip McCalman
Department of Economics
University of California-Santa Cruz
mccalman@ucsc.edu

January 2009

Abstract

When a foreign monopolist sets a single market clearing price for its product, the sign of the optimal tariff is determined by the extent of pass through (also known as the terms of trade effect). However, when a foreign monopolist uses a non-linear price schedule to discriminate between domestic consumers the calculus of welfare maximization is very different. While there are still terms of trade effects from the imposition of a tariff, the existence of such effects are neither necessary nor sufficient to determine the sign of the optimal tariff. Instead the distribution of valuations within the population is the key determinant of the nature of policy intervention. This result differs significantly from the uniform price case and is driven by the incentive compatibility constraint which places the distribution of types at the center of the analysis. If there is a relatively large fraction of high valuation types in the population, then domestic information rents can be increased substantially by subsidizing imports thereby increasing the consumption of the low valuation types and moving the incentive constraint in favor of the high valuation types. However, if the share of high types in the population is relatively small then the increase in information rents will also be small but the fiscal implications of a subsidy will be large. Consequently, the optimal policy will be to impose a trade tax.

Key Words: Terms of Trade, Trade Policy, Discrimination
JEL Classifications: F13, F15

1 Introduction

Firms use a wide variety of techniques to sell their products. The diversity of methods extends far beyond the simple notion that a lower price will attract more consumers. Indeed the whole idea that firms rely on linear prices is challenged by the contracts offered for not only telecommunication services but also by the discounts for volume available in almost all retail outlets (e.g. 500ml of soft drink is more than half the price of 1000ml). The literature analyzing these techniques is vast but all studies have one feature in common, they all focus exclusively on a closed economy setting. The focus on closed economy models seems natural, especially since the working assumption regarding international trade is that it is motivated, at least in part, by differences in price. Such a motivation naturally lends itself to a relatively competitive view of international markets. However, recent empirical work challenges the idea that international trade is conducted in competitive markets characterized by relatively anonymous exchange. Indeed, Rauch (1999) finds that between 60-70% of international trade occurs outside of organized exchanges or institutions that disseminate price information. This implies that the majority of international trade occurs in settings where information about the identity of buyers and sellers is important and that one or potentially both sides of an exchange have market power.[1] At the very least this suggests there is scope for some of the vast array of selling techniques observed in the closed economy to also be applied to international transactions.

This paper considers one such technique; indirect price discrimination.[2] Specifically a foreign monopolist is assumed to serve a heterogeneous group of domestic consumers. Unfortunately for the monopolist there is no obvious way of distinguishing between the various types of customers. Instead the foreign monopolist offers the domestic consumers a menu to pick from, each option differing in terms of the quantity and total payment required. By choosing among the different options the consumers are implicitly generating a distribution of prices for the product sold by the monopolist. How is a domestic government's view of the world altered by such behavior? More specifically, is the operation of trade policy likely to be affected by the non-linear pricing structure of the foreign monopolist?[3]

[1]For an analysis of the implications of this type of environment for international trade see McLaren (2000) and Grossman and Helpman (2002).

[2]While indirect price discrimination hasn't been extensively analyzed in the literature, tariff discrimination has been. For recent contributions see Saggi (2006), Bagwell and Staiger (2004) and Bown (2004). See Ganslandt and Maskus (2007) for an example of third degree price discrimination in an international context.

[3]Another possibility in non-competitive markets is that relationship specific investment maybe undertaken. For an analysis of trade policy and the hold-up problem see Antras and Staiger (2008).

This question is interesting since an important consideration in assessing the implications of trade policy is the terms of trade effect. This effect is central to arguments about optimal trade policy since it provides a motivation for intervention by the domestic government. However, if there is not a single market clearing price it is not immediately obvious whether or how a domestic government should intervene in international trade.

This paper demonstrates that the optimal trade policy in the presence of non-linear pricing has a very different motivation than when the foreign monopolist charges a uniform price (see Brander and Spencer (1984)).[4] In particular a terms of trade effect is neither necessary nor sufficient for intervention to occur by a national welfare maximizing government.[5] This stands in direct contrast to a uniform price monopolist where the terms of trade effect not only serves as the basis for intervention but also determines whether the trade tax is positive or negative. For the uniform price case, if pass through is less than complete (i.e. a $1 trade tax raises the domestic price by less than a dollar), then it is optimal to impose a positive trade tax since the foreign firm pays part of the tax. However, if pass through is more than complete then the optimal policy is a subsidy to imports. In this instance a $1 subsidy will lower the domestic price by more than a dollar, with the terms of trade benefit once again the mechanism that motivates intervention. Finally, if pass through is complete then the optimal tariff is zero. Such logic is no longer reliable when the foreign firm uses non-linear prices.

Given that the linear and non-linear price scenarios share many common features, such as total output below the efficient level, why is there a difference in the operation of optimal trade policy? The critical difference relates to the information rents captured by a consumer under non-linear pricing. These rents are derived from the ability of high valuation types to claim to be low valuation types and the need for the monopolist to set an incentive compatible price schedule to avoid this outcome. As a result the information rents that a high type earns are a function of the bundle designed for the low types, which creates an incentive for the monopolist to reduce the low types bundle below the efficient level. By offering a subsidy on all imports the government moves production closer to the first best for the low valuation types. The increased consumption of the low valuation types from the subsidy also translates into greater information rents for the high valuation types due to a change in the incentive compatibility constraint. Will a government choose a subsidy? Clearly, more than

[4]For other work that considers optimal trade policy under asymmetric information see Bagwell and Staiger (1989) and Kolev and Prusa (1999).

[5]For a more general discussion of the role of the terms of trade effect in trade policy and international agreements see Bagwell and Staiger (2002).

complete pass through lowers the fiscal cost of the subsidy. However, the degree of heterogeneity among consumers is more important. If the fraction of high types is relatively large, then the optimal policy is more likely to be a subsidy since the gain in information rents is more than sufficient to offset the fiscal cost of the subsidy. This effect can be so strong that a subsidy is optimal even when pass through is less than complete. Conversely, when pass through is more than complete, the optimal trade tax can still be positive under non-linear pricing, with this more likely the higher the fraction of low types in the population. Hence, the incentive compatibility constraint introduces a role for consumer heterogeneity into the welfare calculus that implies the optimal policy is now determined by the interaction of the change in the terms of trade and the distribution of valuations. It is this interaction that generates the results that stand in stark contrast to those derived under the assumption of a uniform price monopolist. Moreover, the welfare implications tend to be more pronounced under non-linear pricing since the relevant welfare comparisons are no longer a dead weight loss triangle versus a slice of the tariff revenue rectangle paid by foreigners; but rather it is an information rent trapezoid versus the full tariff revenue rectangle. In other words, both gains and losses from trade policy are of a first order of magnitude under non-linear pricing.

The paper is structured as follows. To provide some evidence for the operation of non-linear pricing of international transactions, the next section considers whether all imports of a given product from a particular source country enter the US for the same price or whether there is a distribution of prices. There does indeed seem to be a distribution of prices, which persists despite controlling for a number of factors including economies of scale and distance. One possible explanation of the residual variation is non-linear pricing. Section 3 then presents a standard model of non-linear pricing by a monopolist, the main difference being that a border separates the producer from the consumers. In this context the optimal trade tax is derived and characterized in section 4. Section 5 considers a number of extensions such as the implications of more than two types, the presence of a domestic retailer as well as quality choice and ad valorem tariffs. Finally, 6 concludes and discusses avenues of potential research.

2 Variation in Unit Values and US Import Volumes

Is there any evidence of non-linear pricing or quantity discounting in international trade? The identification of quantity discounting would ideally require transaction level data. However, such data is not readily available for international sales. Nevertheless import data for the US is available at a very disaggregated level, and this data can be examined for patterns that are consistent with quantity discounting. In particular the data record imports by source country for each of the 10 digit Harmonized System (HS10) products, and this information is additionally broken down by month and US customs district.[6] An important feature of this data is that it contains information on both f.o.b. value and quantity, therefore unit values for a product can be computed by exporter, month and customs district. Data of this degree of disaggregation make it possible to explore whether or not imports of a product from a country tend to enter the US for approximately the same price or whether there is a distribution of prices.

Table 1: Coefficient of Variation for Unit Values, 1994

Sample Definition	CRT Monitors	White Shirts	Fuel Oil
Whole	2.44	1.19	0.28
Monthly FE	2.14	1.17	0.26
Exporter FE	1.25	0.45	0.23
# obs	(2269)	(3874)	(670)

Since our goal is illustrative we will focus on three products that have previously been the subject of analysis.[7] These products are CRT monitors, men's white cotton shirts and fuel oil.[8] Table 1 reports the coefficient of variation for each product in 1994, where observations consist of unit values for a given product from a country going to a customs district in a month. The first row considers the degree of variation in unit values where the data have been normalized by the sample mean for each product. As previous studies have noted, even for narrowly defined products there is a large degree of variation in unit values, with the coefficient of variation ranging from a high of 2.44 for CRT monitors to a low of 0.28 for fuel oil. The ordering of these statistics also aligns with priors since CRT monitors are a relatively differentiated good while men's white shirts are somewhat standardized and fuel oil represents a commodity. The second row considers one possible source of variation, month to month fluctuations where the variation is calculated after the data are normalized by dividing through by the mean for every month for each product.[9] As can be seen, this returns values very similar to the first row, and therefore the explanation of the variation must lie elsewhere.

[6] There are approximately 15,000 products at the 10 digit level and 44 customs districts. This data is made available by the U.S. Census Bureau.

[7] See Schott (2004).

[8] The HS10 codes are 8471923200, 6205202065 and 2710000530, respectively. This data is available from http://dataweb.usitc.gov/

[9] This implies that the mean for each month is unity and the sample mean is also unity.

Following the previous literature the third row includes exporter fixed effects.[10] Once again the results are relatively intuitive with the exporter effects accounting for half the variation in unit values for CRT monitors and very little for fuel oil. However, a notable feature not identified by previous work is that there still exists a large amount of variation in unit values for both CRT monitors (1.25) and white shirts (0.45) that is unaccounted for by exporter fixed effects. This implies that these products sell for a wide range of prices, even controlling for the country of origin (i.e. the unit values vary widely and products from a particular country are not sold for the same price). In fact for both CRT monitors and men's white shirts, the variation in unit values within countries is roughly the same order of magnitude as the variation in unit values across countries.

A natural question to ask is what generates this price dispersion? There are a number of possible explanations including economies of scale, within exporter firm heterogeneity[11], variation in product quality across US customs districts (Alchian-Allen effects)[12], misreported data[13] or some form of price discrimination such as non-linear pricing. The objective here is not to attempt to test each possibility to see which explanation fits the data the best (since all are likely to play some role), but rather to see if non-linear pricing can be ruled out by controlling for the other main sources of variation. If there is residual variation, then non-linear pricing is a potential explanation. In particular, a specific form of non-linear pricing is considered: quantity discounts.[14] If quantity discounts are responsible for the variation in unit values, then shipments with larger quantities should be associated with lower unit values.

However, it is also possible that larger shipments are associated with economies of scale and therefore result in lower unit values. To account for this possibility it is important to include a measure of the overall scale of an exporter's production. The control used is aggregate quantity of a good shipped to the US in a month from an exporter. With this measure it is assumed that any price variation related to scale economies are realized within a month and are reflected uniformly in the price for that month. Aside from economies of scale another possibility is that the quality composition of within product bundles differ across US customs districts. One explanation of this difference is the Alchian-Allen effect. This says that per unit transport costs will induce changes

[10]Variation is calculated after the data have been normalized by dividing through by the mean unit value for each exporter for each product.

[11]See Melitz (2003).

[12]See Hummels and Skiba (2004).

[13]It is well known that customs data are subject to errors in reporting, especially in terms of quantities and product classification. See GOA, 1995

[14]See Busse and Rysman (2005) for evidence of quantity discounting in a domestic setting.

in relative prices, with the total per unit cost higher for markets further away from the product's country of origin. If higher quality versions of a product receive a higher price, then adding a common transport cost to all versions will move relative prices in favor of high quality goods, with this effect most pronounced for customs districts further away from the product's origin. This changes the quality composition of the product bundle that arrives at each customs district (i.e. the shipping the good apples out phenomena). To capture this mechanism a measure of the distance between the exporting country and the customs district of unlading is included.[15] To control for this and other possibilities consider the following specification:

$$ln(UV^i_{sdm}) = \beta^i_1 ln(Q^i_{sdm}) + \beta^i_2 ln(Q^i_{sm}) + \beta^i_3 ln(dist_{sd}) + \alpha^i_s + \alpha^i_m + \alpha^i_d + \epsilon^i_{sdm} \quad (1)$$

where $ln(UV^i_{sdm})$ is the log of the unit value for product i, from source country s to customs district d in month m, $ln(Q^i_{sdm})$ is the log of the quantity exported of i shipped from s to d in m, $ln(Q^i_{sm})$ is the log of total quantity of i shipped by s to the US in month m, α^i_s is a vector of exporter fixed effects, α^i_m is a vector of month fixed effects, α^i_d is a vector of customs district fixed effects and ϵ^i_{sdm} is the error term. However, this specification doesn't take into account the relationship between Q^i_{sdm} and Q^i_{sm}. To incorporate this dependence, note that Q^i_{sdm} can be expressed as $sh(Q^i_{sdm}) * Q^i_{sm}$, where $sh(Q^i_{sdm})$ represents the share of Q^i_{sm} going to d. This allows (1) to be rewritten as:

$$ln(UV^i_{sdm}) = \beta^i_1 ln(sh(Q^i_{sdm})) + \tilde{\beta}^i_2 ln(Q^i_{sm}) + \beta^i_3 ln(dist_{sd}) + \alpha^i_s + \alpha^i_m + \alpha^i_d + \epsilon^i_{sdm} \quad (2)$$

This transformation doesn't alter the estimate of β_1 but does allow the measure of economies of scale, $\tilde{\beta}^i_2 = \beta^i_1 + \beta^i_2$, to be more directly identified. Finally, to account for outliers generated by misreported shipments, robust estimation techniques are employed.[16]

The results are reported in Table 2. Note that the variables used to control for economies of scale and Alchian-Allen effects have co-efficients that align with expectations. In particular, the coefficients on the economies of scale measure have an intuitive ordering with CRT monitors having the largest economies of scale and fuel oil having no economies of scale. While using distance to control for the Alchian-Allen effect generates the result that the further a CRT monitor is shipped,

[15] For an extensive analysis of the impact of distance on trade see Disdier and Head (2008).

[16] OLS produces very similar results for both ln(sh(Q_{sdm})) and ln(Q_{sm}), but the co-efficient on ln($dist_{sd}$) for CRT monitors is sensitive to outliers. In particular, it switches from significant and negative (-0.17) to significant and positive under robust regression.

Table 2: Robust Regression Estimates

	CRT Monitors	White Shirts	Fuel Oil
$\ln(\text{sh}(Q_{sdm}))$	-0.09***	-0.02***	-0.00
$\ln(Q_{sm})$	-0.14***	-0.04***	-0.02
$\ln(\text{dist}_{sd})$	0.17**	0.01	0.00
Exporter FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Customs Dist. FE	Yes	Yes	Yes
R^2	0.87	0.82	0.73

*** p<0.01, ** p<0.05, * p<0.1

the larger the unit value, which is consistent with higher quality bundles shipped to more distance destinations. Once again the ranking of the co-efficients on distance are sensible with the co-efficient on distance positive for white shirts (though insignificant) and the co-efficient for fuel oil zero (no quality upgrading for homogeneous goods).

After controlling for economies of scale and Alchian-Allen effects, the quantity discount for CRT monitors is pronounced with each 10% increase in volume leading to a 0.9% decline in price. White shirts show a more modest, yet statistically significant, quantity discount. In contrast, fuel oil records no economically or statistically significant association between unit value and volume. While the analysis is illustrative rather than exhaustive, it is consistent with the notion of quantity discounts being offered for products where producers have greater market power.

However, there are other potential explanations for the negative correlation between unit values and quantity shipped, with one obvious candidate being price dispersion induced by firm heterogeneity as in Melitz (2003). In particular, firms with higher productivity draws will have both lower prices and sell greater quantities. However, the data used in Tables 1 and 2 are based on total shipments to a customs district in a month and therefore it potentially aggregates shipments across firms. In this case a prediction about the average unit value for goods shipped to a customs district is needed. One plausible assumption is that sub-national market entry costs vary in such a way that more varieties from an exporting country are supplied to larger markets, which predicts that unit values should be increasing in volume shipped to a customs district. That is, the lowest productivity exporters only find it profitable to serve the largest sub-national markets. While plausible, this prediction is contrary to the results reported in Table 2 and argues against the standard model of firm heterogeneity as the basis for the correlation. However, Baldwin and Harrigan (2007) develop an augmented heterogeneous firm framework that is consistent with the results in Table 2. In their model higher prices are associated with higher quality, but critically lower quality adjusted prices. This predicts a positive association between price and quantity sold at the firm level, but a negative association between average unit values and quantities at the customs district level under the entry cost assumption employed above.

To further investigate whether the negative association between unit values and quantities documented in Table 2 are generated by quantity discounts or firm heterogeneity, more disaggregated data are used.[17] This data includes information on the mode of shipment (air or sea), weight of the shipment as well as the number of shipments to the ports contained within a customs district.[18] This allows the use of within customs district variation to distinguish between a quality augmented heterogeneous firm model and a quantity discounting model since an observation is now relatively close to an individual shipment. In particular, the quality augmented model predicts a positive association between unit values and shipment size (or at least less pronounced negative association as the data becomes more disaggregated), while quantity discounting still predicts a negative relationship. To analyze the more refined data, the following specification is used:

$$ln(UV^{ij}_{spm}) = \beta^{ij}_1 ln(Q^{ij}_{spm}) + \alpha^{ij}_{sdm} + \epsilon^{ij}_{spm} \tag{3}$$

where the subscript p represents a port level observation, the superscript j represents mode of transport and α^{ij}_{sdm} is an exporter, customs district and month fixed effect. Including this fixed effect accounts for both economies of scale and Alchian-Allen effects directly. To purge outliers one final adjustment is made to the data. By combining product counts with product weights, relatively standardized products can be identified. For CRT monitors only shipments with screens that weigh between 10 and 20 pounds are included, while only shirts that weigh between 2-8 pounds per dozen and fuel oil that weighs between 130-160 pounds per barrel are considered. For comparability, these standardized samples are aggregated across customs districts, and specification (1) is used on this more aggregate data set. The results for the standardized data for both levels of aggregation are reported in Table 3.

[17]See Bureau of the Census (1990).

[18]While shipments to ports can be distinguished, there is no identifier associated with a port. Consequently we observe multiple shipments of a product from a country to a customs district within a month, but cannot consistently track which ports receive which shipments from month to month.

Table 3: OLS Estimates for Standardized Sample, Different Levels of Aggregation

		CRT Monitors	White Shirts	Fuel Oil
Port Level	$\ln(Q^{air}_{spm})$	-0.16**	-0.07**	
	$\ln(Q^{sea}_{spm})$	-0.09***	-0.08***	0.00
Customs District	$\ln(Q_{sdm})$	-0.09***	-0.05***	0.00

*** p<0.01, ** p<0.05, * p<0.1

These results are contrary to the quality augmented firm heterogeneity model since the sign of β_1 is negative and the magnitude of the co-efficients are uniformly smaller than when the more aggregated data are used. Therefore, the evidence offered does not rule out non-linear pricing as a potential explanation of the variation in unit values.

So far the analysis has concentrated on three illustrative products. However, the analysis can be extended to cover all 15,352 HS 10 products. Replicating the analysis of Table 3, 5,676 products are estimated to have negative and significant coefficients on $ln(Q_{sdm})$ when the analysis is conducted at the customs district level. Using the port level data where shipments arrive by air, the null hypothesis that the coefficient on $ln(Q_{spm})$ is greater than the coefficient on $ln(Q_{sdm})$ can be rejected in 90% of pairwise comparisons. Similarly, when shipments arrive by sea the null hypothesis that the coefficient on $ln(Q_{spm})$ is greater than the coefficient on $ln(Q_{sdm})$ can be rejected in 80% of pairwise comparisons. These results tend to favor a quantity discounting mechansim over the quality augumented firm heterogeneity model in the vast majority of cases. Furthermore, this suggests that quantity discounting is consistent with observed behavior for 20%-30% products. Thus the next section considers the implications of this behavior for the operation of trade policy.

3 Non-linear Pricing by a Foreign Monopolist

Quantity discounts can result from second degree price discrimination, a price strategy that has not been extensively analyzed in the international economics literature. To determine the implications of this selling mechanism consider the standard monopoly model of non-linear pricing.[19] The model assumes that a monopolist sells a single good to a heterogeneous set of consumers at a constant marginal cost, c.[20] At this point the only departure from the standard framework is that the

[19] See Maskin and Riley (1984) and Tirole (1988).

[20] The presentation here will treat the output produced by the monopolist as homogeneous. However, the model can also be given a quality interpretation with each consumer type buying one unit of a good where the quality of that

consumers are assumed to be located in the home country while the monopolist is located in the foreign country. Since a border separates the two sets of agents, it is natural to consider that there is some trade policy instrument that is applied at the border to imports. This policy is assumed to be a per unit trade tax, t.[21] While more complicated policies can be implemented, this simple policy has the feature that it both enables a direct comparison to the previous trade policy literature as well as capturing the actual structure of observed trade policy.[22]

The game is assumed to have four stages. In the first stage the domestic government sets trade policy. In the second stage the monopolist announces its schedule of prices and quantities, while in stage 3 consumers receive their preference draw from a common knowledge distribution, though the draw itself is private information. Finally, in stage 4 trade occurs. To solve this game we begin in stage 4 and work backwards.

3.1 Preferences

A consumer receives utility according to a quasi-linear utility function and faces a problem of maximizing utility given a pricing menu that specifies a total payment, $T(q)$, for any given quantity, q. Therefore the objective function has the following form

$$\theta_i u(q) - T(q) \qquad i = L, H$$

where $\theta_i u(q)$ is the utility function and $T(q)$ is the monetary transfer from the consumer to the monopolist. Assume that $u(q)$ is increasing in q, $u' > 0$, and strictly concave, $u'' < 0$. The parameter θ indexes a consumer's willingness to pay and is assumed to be only known by the individual consumer. For expositional simplicity, assume that there are only two types of consumers, high valuation types (θ_H) and low valuation types (θ_L), where $\theta_H > \theta_L$.[23]

good is endogenous. However, this changes the interpretation of the trade tax, and the implications of this perspective will be discussed in more detail in section 5.3.

[21]As under a uniform price monopoly the first best policy is a price ceiling at marginal cost. However, if there are fixed costs of production this policy needs to be augmented. The augmentation becomes more difficult to implement if unobservable fixed costs exist. Likewise lump-sum taxation may suffer from the same difficulty. Furthermore, if there is quality differences then a price ceiling is no longer optimal.

[22]More specifically single rate ad valorem trade taxes are the norm (which is true for the commodities discussed in section 2) and the extension of the model to such policies is straight forward and discussed in section 5.3. The motivation for considering a per unit tax is that this policy is associated with the greatest variation in policy choice in the previous literature (see Brander and Spencer (1984)).

[23]For a generalization to more types see section 5.1.

3.2 The Monopolist's Problem

Assume that the monopolist believes that the high valuation type occurs in the population with probability β. However, to keep the problem interesting we place a restriction on the distribution of types in the population. In particular, we concentrate on circumstances where both types always participate in the market under *free trade*. To ensure that this is the case we assume that $(\theta_L - \beta\theta_H)u'(0) > (1-\beta)c$ or $\beta < \frac{\theta_L u'(0)-c}{\theta_H u'(0)-c}$. Note in particular that this does not preclude the option of imposing a tariff that causes the low value type not to be served in equilibrium.

In this case the monopolist chooses $\{(q_L, T_L), (q_H, T_H)\}$ to maximize

$$\Pi = \beta(T_H - (c+t)q_H) + (1-\beta)(T_L - (c+t)q_L)$$

subject to

$$\theta_L u(q_L) - T_L \geq \theta_L u(q_H) - T_H \quad (4)$$

$$\theta_H u(q_H) - T_H \geq \theta_H u(q_L) - T_L \quad (5)$$

$$\theta_L u(q_L) - T_L \geq 0 \quad (6)$$

$$\theta_H u(q_H) - T_H \geq 0 \quad (7)$$

where (4) and (5) are the incentive compatibility constraints for the low and high types, respectively, while (6) and (7) are the corresponding participation constraints. Given our assumptions, it is well known from the mechanism design literature that only two of these constraints bind, the incentive constraint for the high type, (5), and the participation constraint for the low type, (6). Therefore the relevant constraints can be rewritten as:

$$T_H = \theta_H u(q_H) - \theta_H u(q_L) + \theta_L u(q_L) \quad (8)$$

$$T_L = \theta_L u(q_L) \quad (9)$$

This allows the profit maximization problem to be simplified to become:

$$\max_{q_L,\, q_H} \Pi = \beta\left(\theta_H u(q_H) - \theta_H u(q_L) + \theta_L u(q_L) - (c+t)q_H\right) + (1-\beta)(\theta_L u(q_L) - (c+t)q_L) \quad (10)$$

Taking first order conditions gives:

$$\frac{\partial \Pi}{\partial q_L} = (1-\beta)(\theta_L u'(q_L) - (c+t)) - \beta(\theta_H - \theta_L)u'(q_L) = 0 \quad (11)$$

$$\frac{\partial \Pi}{\partial q_H} = \theta_H u'(q_H) - (c+t) = 0 \quad (12)$$

These conditions imply the standard result that the high types are offered an efficient quantity, while the low types are offered and accept a bundle with an inefficiently low quantity.

Two useful comparative static results are that the optimal quantity offered to a low valuation type is decreasing in the probability of being a high type:

$$\frac{dq_L}{d\beta} = \frac{\theta_H u'(q_L) - (c+t)}{(\theta_L - \beta\theta_H)u''(q_L)} < 0 \quad (13)$$

and that the quantity offered to each type is decreasing in the trade tax:

$$\frac{dq_L}{dt} = \frac{(1-\beta)}{(\theta_L - \beta\theta_H)u''(q_L)} < 0 \quad (14)$$

$$\frac{dq_H}{dt} = \frac{1}{\theta_H u''(q_H)} < 0 \quad (15)$$

4 Optimal Trade Tax

Faced with a discriminating foreign monopolist the question arises as to whether trade policy is any more or less effective than under conventional uniform pricing. If a firm implements a more sophisticated pricing strategy, does this mitigate or enhance the argument for intervention? Indeed the case for an optimal tariff in the presence of a foreign monopoly is already relatively qualified (see Brander and Spencer (1984)). Nevertheless, if intervention is warranted, what form will it take and what factors will it be conditional on? To address these questions, begin by defining the welfare of the importing country as the sum of net consumer surplus (CS) and tariff revenue (TR). In the current setting net consumer surplus is equal to the information rents captured by the domestic consumers. Since the low type is held to their reservation level of utility, only information rents accrue to the high types. In particular, $\theta_H u(q_H) - T_H = (\theta_H - \theta_L)u(q_L)$, which has the important property that it is independent of the amount consumed by the high type. Therefore domestic welfare is represented as:

$$\begin{aligned} W &= CS + TR \\ &= (1-\beta)(\theta_L u(q_L) - T_L) + \beta(\theta_H u(q_H) - T_H) + TR \\ &= \beta(\theta_H - \theta_L)u(q_L) + TR \end{aligned} \tag{16}$$

where the last line is derived using (8) and (9). This objective function differs from the standard one for two reasons; not only does the foreign monopolist use non-linear prices but it does so in the presence of incomplete information. Since both of these features are new to the analysis of optimal trade taxes it is worthwhile decomposing their influences on the optimal policy. This is most easily achieved by assuming that the foreign monopolist has complete information, and consequently it uses non-linear prices to extract all consumer surplus. To be consistent with our previous assumption that both types are served under free trade, we start by considering what happens as $\beta \to 0$.

4.1 Optimal tariffs when the foreign monopolist has complete information

If the foreign monopolist has complete information, then it can extract all of the domestic surplus by offering to sell a quantity given by $\theta_L u'(q_L) = c + t$ for a payment of $T_L = \theta_L u(q_L)$ (these results follow from (9) and (11)). In this case, domestic welfare is given by the tariff revenue: $W = tq_L$. Since tariff revenue is the only source of domestic welfare, the optimal tariff is given by:

$$\frac{\partial W}{\partial t} = q_L + t\frac{\partial q_L}{\partial t} = 0 \tag{17}$$

$$\rightarrow t^* = -\frac{q_L}{\partial q_L / \partial t} > 0 \tag{18}$$

which is the revenue maximizing tariff. A couple of observations are in order. First, the optimal tariff is always positive regardless of the shape of the demand function (this is in contrast to the usual uniform price monopoly result). Of course, this is subject to the second order conditions for a maximum for the domestic government. This condition requires $2 + q_L \frac{u'''(q_L)}{u''(q_L)} < 0$, which is the standard condition required for a uniform price monopolist to maximize profits. In a sense the ability of the monopolist to extract all the rents reverses the usual roles, with the domestic government now the monopoly supplier of market access, with the price of market access given by the tariff. By restricting market access the domestic government is able to increase domestic surplus. Finally, note that the terms of trade effect plays no role in determining the sign of the optimal tax; regardless of the extent of pass through the optimal policy is always a tax. This is a direct consequence of the ability of the foreign monopolist to extract all consumer surplus through non-linear pricing, which decouples the link between distortions in quantity consumed and domestic welfare. These results are summarized in the following proposition:

PROPOSITION 1 *If the foreign monopolist has complete information about the preferences of domestic consumers and uses optimal non-linear prices, then the optimal tariff is the revenue maximizing tariff. Moreover, the sign of the optimal tariff is independent of the degree of pass through of the tariff.*

4.2 Optimal tariff when foreign monopolist has incomplete information

We now examine the optimal policy when the foreign monopolist is incompletely informed. To help gain insight into the mechanics of the process, we start with a setting where the government can make trade taxes contingent upon the volumed shipped.

4.3 Quantity dependent trade policy

A feature of the incomplete information pricing policies is that they induce the different types to purchase bundles of different sizes. While it may not be practical to set different rates for the different quantities imported, the consideration of the optimal tax for the high and low quantity is nevertheless revealing since it provides insight into the mechanics of optimal policy choice. In this case the welfare function becomes:

$$W = \beta(\theta_H - \theta_L)u(q_L) + (1-\beta)t_L q_L + \beta t_H q_H \tag{19}$$

The associated first order conditions are:

$$\frac{\partial W}{\partial t_H} = q_H + t_H \frac{\partial q_H}{\partial t_H} = 0 \tag{20}$$

$$\frac{\partial W}{\partial t_L} = \beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial t_L} + (1-\beta)\left(q_L + t_L\frac{\partial q_L}{\partial t_L}\right) = 0 \tag{21}$$

These conditions reveal a number of interesting features. First, despite the presence of incomplete information the first order condition for the optimal tariff on the high quantity is of the same form as

(17). In this case, regardless of the distribution of types, the government will always seek to impose the revenue maximizing tariff on the high quantity shipments. This may seem somewhat surprising since the foreign monopolist doesn't capture all the surplus from the high type. On the contrary, the high type is able to capture some of the surplus due to its ability to imitate the low type. In fact, it is exactly this aspect that decouples the surplus of a high type from the quantity they consume; an attribute shared with the complete information case. Instead the surplus that a high type captures is a function of the quantity consumed by the low type. Hence, a tariff on the quantity consumed by the high type reduces the quantity they consume but not the surplus they capture, implying that the domestic government should set the revenue maximizing tariff for the high type quantity.

The second point to note is that the first order condition for the optimal tariff on the low quantity reflects the same revenue considerations as (17) but also incorporates an additional term reflecting the change in consumer surplus. In this case the government now trades-off the net fiscal implications of a trade tax $(1-\beta)(q_L + t\frac{\partial q_L}{\partial t_L})$ against the negative consumption distortion associated with a positive tariff $(\beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial t_L})$. Note in particular that the consumption distortion is experienced by the high valuation type but its magnitude is determined by the quantity response of the low type to a tariff. This is a direct consequence of the incentive compatibility constraint which links the information rents of the high type to the surplus that they can gain from imitating a low type.

The optimal tariffs with incomplete information can be expressed as:

$$t_L^* = \frac{-q_L}{dq_L/dt_L}\left(1 + \frac{\beta(\theta_H - \theta_L)u'(q_L)\frac{dq_L}{dt_L}}{(1-\beta)q_L}\right) \quad (22)$$

$$t_H^* = \frac{-q_H}{dq_H/dt_H} \quad (23)$$

Unlike the revenue maximizing tariff imposed on the high quantities, the sign of the optimal tariff on the low quantities is no longer unambiguous. In particular, whether it is optimal to tax or subsidize imports depends on the sign of $\left(1 + \frac{\beta(\theta_H - \theta_L)u'(q_L)\frac{dq_L}{dt_L}}{(1-\beta)q_L}\right)$, which depends directly on the distribution of types. The numerator of the second term reflects the government's desire to move the consumption levels for the low type closer to the efficient level since by doing so it increases the surplus captured by the high types. The denominator reflects the fiscal implications of this policy. It follows directly that the smaller the probability of a low value type (the larger is β), the more likely a subsidy is to be implemented; the cost of a subsidy is diminished while the benefit increases as β

increases. The intuition for this connection follows from the impact of a subsidy on the incentive compatibility constraint. A subsidy increases q_L which makes high types more likely to claim to be low types. To counter this the foreign monopolist must allow the high types to capture more surplus. This increase in information rents tends to be relatively big if the probability of a low type is small (which implies that q_L is relatively small and far from the first best quantity).

While these mechanics differ from the standard case with uniform prices, it is of interest to determine whether, nevertheless, the standard logic applies; the sign of the optimal policy can be predicted by the extent of pass through. This is clearly not the case for the optimal policy on the high quantity where there is no role for the terms of trade. To determine whether this result carries over to the optimal policy associated with the low quantity define the average per unit price paid by the low valuation type as $p_L \equiv \frac{T_L}{q_L}$. Differentiating the average price with respect to the trade tax, t_L gives:

$$\frac{dp_L}{dt_L} = \frac{dT_L}{dt_L}\frac{1}{q_L} - \frac{dq_L}{dt_L}\frac{T_L}{(q_L)^2} \tag{24}$$

Totally differentiating (6) with respect to t_L, gives $\frac{dT_L}{dt_L} = \theta_L u'(q_L)\frac{dq_L}{dt_L}$. Substituting this quantity into (24) and using (6) yields:

$$\frac{dp_L}{dt_L} = \frac{dq_L}{dt_L}\left(u'(q_L) - \frac{u(q_L)}{q_L}\right)\frac{\theta_L}{q_L} \tag{25}$$

This represents the equivalent of the standard terms of trade effect for the low valuation type consumer. Note that as with the standard terms of trade effect, there is a terms of trade benefit if $\frac{dp_L}{dt} < 1$ and a terms of trade loss if $\frac{dp_L}{dt} > 1$. Combining (22) and (25) implies that the optimal tariff can be expressed as:

$$t_L^* \quad = \quad \frac{-q_L}{dq_L/dt}\left(1 - \frac{\beta(\theta_H - \theta_L)}{(1-\beta)\theta_L}\left(\frac{u'(q_L)}{\frac{u(q_L)}{q_L} - u'(q_L)}\right)\frac{dp_L}{dt_L}\right) \tag{26}$$

Since the coefficient on $\frac{dp_L}{dt_L}$ can be either greater than or less than one, it is possible that the terms of trade effect can be either magnified or muted depending on parameter values. To see this consider the benchmark of the optimal tariff if instead the foreign monopolist uses a uniform price, p^U, to sell an equilibrium quantity, Q^U:

$$t^U = \frac{-Q^U}{dQ^U/dt}\left(1 - \frac{dp^U}{dt}\right) \tag{27}$$

These two expressions ((26) and (27)) share some common features. In both cases the sign of the optimal tariff is determined by the sign of the term in parenthese. However, under a uniform price monopoly the degree of pass through ($\frac{dp^U}{dt} \gtrless 1$) completely determines whether a subsidy or a tax is imposed at the border. While the extent of pass through also plays a role in determining the optimal policy under non-linear pricing, the impact of a terms of trade effect is now more difficult isolate. A key difference between equations (26) and (27) is the critical role played by the distribution of types in determining the sign of the optimal policy under non-linear pricing. This is highlighted by using (14) and writing the optimal tariff for the low quantity in the following form:

$$t_L^* = \frac{-q_L}{dq_L/dt_L}\left(1 - \frac{\beta(\theta_H - \theta_L)}{(\theta_L - \beta\theta_H)}\epsilon(q_L)\right) \tag{28}$$

where $\epsilon(q_L) = -\left(\frac{u'(q_L)}{u''(q_L)q_L}\right)$, the elasticity of demand. In this form the role of the distribution of types is most transparent. The more elastic is the demand function, the higher is the fraction of low valuation types in the population that are consistent with an optimal tariff. In contrast, except for its impact on the degree of pass through, the distribution of types plays no direct role in determining the sign of the optimal policy under uniform pricing. This leads to the following proposition:

PROPOSITION 2 *If a foreign monopolist uses an optimal non-linear price schedule to discriminate between domestic consumers, then the optimal tariff on q_H is the revenue maximizing tariff and the optimal tariff on q_L, t_L, can be either positive or negative. However, the optimal t_L is decreasing in the fraction of high types in the population, β, the degree of heterogeneity, $(\theta_H - \theta_L)$, and the elasticity of demand, $\epsilon(q_L)$.*

One final observation is warranted. The standard optimal tariff argument is based on the comparison of the dead weight loss triangle and the tariff revenue paid by the foreign producer. A terms of trade effect ensures that the fiscal contribution of the foreign supplier is always a first order gain, while the dead weight loss is a second order loss. Under non-linear pricing, the relevant welfare comparisons involve much larger quantities. To see this consider equation (21) when $t = 0$. By imposing a small tariff the revenue raised is proportional to $(1-\beta)q_L$. However, in contrast to the standard case, <u>all</u> of this revenue is paid by the foreign producer since they capture the entire surplus from the low types. So it is no longer a thin slice of total revenue contributed by foreigners, but in fact total tariff revenue. Note also that the consumption distortion is not a triangle but a trapezoid ($\beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial t_L}$), so even small tariffs have a large impact on consumer surplus. These observations underscore the very different implications of trade policy under non-linear pricing.

4.4 Single rate trade policy

While optimal taxes that vary by the quantity shipped offer many insights, they may also prove difficult to implement especially since it is possible that small quantities are likely to be subsidized. Consequently a single rate policy is more practical, with the associated objective function of the domestic government as follows:

$$\max_t \; W \;=\; \beta(\theta_H - \theta_L)u(q_L) + t\left((1-\beta)q_L + \beta q_H\right) \tag{29}$$

To characterize the optimal tariff, differentiate domestic welfare with respect to the import tax.

$$\frac{\partial W}{\partial t} \;=\; \beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial t} + Q + t\frac{\partial Q}{\partial t} \tag{30}$$

where $Q = (1-\beta)q_L + \beta q_H$. The optimal tariff with incomplete information for the single rate case can be expressed as:

$$t^* \;=\; \frac{-Q}{dQ/dt}\left(1 - \frac{\beta(\theta_H - \theta_L)}{(\theta_L - \beta\theta_H)} s_L \epsilon(q_L)\right) \tag{31}$$

where $s_L = \frac{(1-\beta)q_L}{Q}$. As with sign of (28) the sign of (31) depends on the term in parenthesis. However, since $s_L \leq 1$ the range of parameter values that are associated with a subsidy is smaller under a single rate policy than a quantity dependent policy.

Turning to the role of the terms of trade effect, note that calculations similar to those performed above allow the optimal tariff to be expressed as:

$$t^* \;=\; \frac{-Q}{dQ/dt}\left(1 - \frac{\beta q_L}{Q}\frac{(\theta_H - \theta_L)}{\theta_L}\left(\frac{u'(q_L)}{\frac{u(q_L)}{q_L} - u'(q_L)}\right)\frac{dp_L}{dt}\right) \tag{32}$$

Since the terms of trade effect interacts with various parameters in the model, no general characterization of its role in determining optimal trade policy is possible. However, by holding preferences constant, we can make progress by considering the interaction of the terms of trade with the distribution of types.

To sharpen our understanding of the relative importance of these influences we analyze two standard demand functions, linear and constant elasticity. These demand functions are particularly interesting to study since they predict that the opposite trade policy (subsidy/tax) will be implemented in a uniform price setting as they generate different predictions about the terms of trade impact of a tariff. In the case of a linear demand curve, a tax is always optimal while under a constant elasticity demand curve a subsidy is optimal. Therefore, it is of interest to examine whether these stark predictions carry over to the non-linear pricing setting.

4.5 Two Common Preference Specifications

We begin with an example where the optimal trade tax is always positive under uniform pricing and ask the question whether this also holds for non-linear pricing. Assume that $\theta_i u(q) = \theta_i(q - \frac{q^2}{2})$, which implies $\theta_i u'(q) = 1 - q$. With this preference structure, the profit maximizing quantities are $q_L = 1 - \frac{(1-\beta)(c+t)}{\theta_L - \beta\theta_H}$ and $q_H = 1 - \frac{(c+t)}{\theta_H}$. To determine the nature of the terms of trade effect, evaluate (25) to get $\frac{dp_L}{dt} = \frac{\theta_L(1-\beta)}{2(\theta_L - \beta\theta_H)} \gtrless 1$. In contrast to the uniform price result, the qualitative nature of the terms of trade effect is ambiguous, increasing from $\frac{1}{2}$ when $\beta = 0$ to be greater than one if $\frac{\theta_L}{2\theta_H} < \beta < \frac{\theta_L - c}{\theta_H - c}$. To avoid checking cumbersome constraints, assume $\theta_L = 2$, $\theta_H = 3$, $c = \frac{1}{3}$ and $\beta \in [0, \frac{5}{8})$. These parameters ensure that $q_L > 0$ and that pass through can be more than complete. In this context, the following correlation holds between the extent of pass through and the optimal tariffs:

$$\left.\frac{\partial p_L}{\partial t}\right|_{\beta=0} < 1 \qquad \text{and} \qquad t^* > 0$$

$$\left.\frac{\partial p_L}{\partial t}\right|_{\beta\to\frac{5}{8}} > 1 \qquad \text{and} \qquad t^* < 0$$

Although this correlation is exactly what would emerge under the standard terms of trade logic, the underlying mechanics are very different. When $\beta = 0$ there is only one type of consumer, low types, and therefore the monopolist has complete information and can extract all consumer surplus by offering an efficient quantity to the low type. It follows from the results of section 4.1 that the terms of trade plays no role and the optimal tariff is positive regardless of the extent of pass through. In contrast, as $\beta \to \frac{5}{8}$, the quantity offered to the low type approaches zero. As with the case where $\beta = 0$ the foreign monopolist can extract all of the domestic surplus from consumption, though unlike section 4.1 total production is not equal to the efficient level. Whether or not a tax or a subsidy is optimal depends on the trade-off between the change in information rents and the fiscal cost of the trade policy. Under the parameters chosen, the trade-off yields a subsidy, in part because pass through is more than complete. However, the terms of trade logic should not be over-emphasized. To see this consider what happens when there is no terms of trade effect, $\frac{\partial p_L}{\partial t}\Big|_{\beta=\frac{1}{2}} = 1$. In this case the optimal tariff is positive, $t^* > 0$, and therefore by continuity the optimal tariff can be positive even when pass through is more than complete. Thus in comparison to uniform pricing in the presence of a linear demand curve, non-linear pricing results in a more varied pass through effect and a richer optimal tariff response.

While the differences between the optimal policy in the two pricing scenarios (linear pricing and non-linear pricing) are very stark in the case of a linear demand function, they become even more pronounced if the demand function has a constant elasticity. Such a demand curve is derived from a utility function of the following form, $\theta u(q) = \theta \frac{q^{1-\frac{1}{\epsilon}}}{1-\frac{1}{\epsilon}}$, where $0 < \frac{1}{\epsilon} < 1$ is imposed to ensure the utility function is strictly concave. This implies $u' = \theta q^{-\frac{1}{\epsilon}}$ and the profit maximizing quantities are $q_L = \left(\frac{\theta_L - \beta\theta_H}{(1-\beta)(c+t)}\right)^{\epsilon}$ and $q^H = \left(\frac{\theta_H}{c+t}\right)^{\epsilon}$. Note that there is always more than complete pass through of the tariff onto the average price of the low type with $\frac{dp_L}{dt} = \frac{(1-\beta)\theta_L}{(\theta_L - \beta\theta_H)(1-\frac{1}{\epsilon})} > 1$ for $\beta \in [0,\ \frac{\theta_L}{\theta_H}]$.

Such a pronounced terms of trade effect would usually imply that the optimal trade tax is negative. However, this logic is flawed when the monopolist uses non-linear prices. To see this evaluate (31) for the constant elasticity case:

$$t^* \;=\; \frac{-Q}{dQ/dt}\left(1 - \frac{\beta(\theta_H - \theta_L)}{(\theta_L - \beta\theta_H)} s_L \epsilon\right)$$

Therefore, whenever $\beta < \frac{\theta_L}{\theta_H + (\theta_H - \theta_L)s_L\epsilon}$ the optimal tariff is positive under non-linear pricing with a constant elasticity demand curve, exactly the opposite of the result under a uniform price, which is a subsidy! The intuition is similar to that of the linear demand function case. When $\beta = 0$, it is optimal to tax regardless of the extent of pass through. However, as $\beta \to \frac{\theta_L}{\theta_H}$ and $q_L \to 0$, the optimal policy is to apply a subsidy. Consequently, even though the pass through of the trade tax is always more than complete, the optimal policy can be either a tax or a subsidy, with the nature of the optimum determined by the distribution of types in the population.

Together these examples show that the terms of trade effect is neither a necessary nor sufficient condition for trade policy intervention in the presence of a non-linear price foreign monopoly. Instead it is the distribution of valuations within the population that is the critical factor.

PROPOSITION 3 *If a foreign monopolist uses an optimal non-linear price schedule to discriminate between domestic consumers, then the optimal tariff on Q can be either positive or negative. In this setting knowledge of the extent of pass through is not sufficient to know the sign of the optimal tariff. However, the optimal tariff is decreasing in the fraction of high types in the population, β, the degree of heterogeneity, $(\theta_H - \theta_L)$, the elasticity of demand, $\epsilon(q_L)$ and the low types consumption share of output, s_L.*

5 Extensions

In this section we consider the generality of the central results to various alternative modeling assumptions.

5.1 More than two types

The results above were derived in a model with two types of consumers. While this allows the analysis to be relatively transparent, it does raise the issue of whether the results extend to a model with more consumer types. To examine the robustness of the results to more types, consider a setting with three types of consumers.[24] In particular, assume that $\theta_H > \theta_M > \theta_L$, with each type occurring in the population with probability β_i, $i \in \{H, M, L\}$. As a benchmark, assume that all types are served under free trade. Noting that the participation constraint binds for the low type, while the incentive constraints bind for both the medium (M) and high types, the objective function for the foreign monopolist can be written as:

$$\begin{aligned}\max_{q_L,\, q_M,\, q_H} \Pi = \beta_H \left(\theta_H u(q_H) - (\theta_H - \theta_M)u(q_M) - (\theta_M - \theta_L)u(q_L) - (c+t)q_M\right) \\ +\beta_M \left(\theta_M u(q_M) - (\theta_M - \theta_L)u(q_L) - (c+t)q_M\right) \\ +\beta_L(\theta_L u(q_L) - (c+t)q_L)\end{aligned}$$

[24]The extension to more than three types is direct.

Taking first order conditions:

$$\begin{aligned}
\frac{\partial \Pi}{\partial q_H} &= \theta_H u'(q_H) - (c+t) = 0 && (33)\\
\frac{\partial \Pi}{\partial q_M} &= \beta_M(\theta_M u'(q_M) - (c+t)) - \beta_H(\theta_H - \theta_M)u'(q_M) = 0 && (34)\\
\frac{\partial \Pi}{\partial q_L} &= \beta_L(\theta_L u'(q_L) - (c+t)) - (1-\beta_L)(\theta_M - \theta_L)u'(q_L) = 0 && (35)
\end{aligned}$$

These equations implicitly define the profit maximizing choices of $q_L,\ q_M,\ q_H$, and capture the response of these quantities to changes in tariffs.

Given the menu set by the foreign monopolist, the domestic government's welfare function is defined as:

$$W = \beta_H(\theta_H - \theta_M)u(q_M) + (1-\beta_L)(\theta_M - \theta_L)u(q_L) + TR$$

Once again it is useful to consider the case where the government sets a tariff targeted for each bundle size, $t_L,\ t_M,\ t_H$. In this case the optimal tariffs are:

$$\begin{aligned}
t_H^* &= \frac{-q_H}{dq_H/dt_H}\\
t_M^* &= \frac{-q_L}{dq_L/dt_L}\left(1 - \frac{\beta_H(\theta_H - \theta_M)}{((1-\beta_L)\theta_M - \beta_H\theta_H)}\epsilon(q_M)\right)\\
t_L^* &= \frac{-q_L}{dq_L/dt_L}\left(1 - \frac{(1-\beta_L)(\theta_M - \theta_L)}{(\theta_L - (1-\beta_L)\theta_M)}\epsilon(q_L)\right)
\end{aligned}$$

As is evident from comparing these tariffs to those derived earlier, the optimal tariffs are driven by the distribution of types and the elasticity of demand and not by the degree of pass through.

If the government is now assumed to only set a single tariff, the optimal tariff is:

$$t^* = \frac{-Q}{dQ/dt}\left(1 - \frac{\beta_H(\theta_H - \theta_M)}{((1-\beta_L)\theta_M - \beta_H\theta_H)}s_M\epsilon(q_M) - \frac{(1-\beta_L)(\theta_M - \theta_L)}{(\theta_L - (1-\beta_L)\theta_M)}s_L\epsilon(q_L)\right)$$

As is evident, the results contained in proposition 3 apply with only a slight modification to allow for the addition of more types.

5.2 Implications of Wholesale-Retail Link

A maintained assumption is that the monopolist sells directly to consumers or a set of perfectly competitive retailers. However, in many cases retailers play an important role in product distribution since they have some form of market power. An example of this behavior is in the US automobile industry where car dealerships cannot be owned by the manufacturers. Moreover these dealers operate under exclusive territory contracts giving them a degree of market power. These vertical relationships raise a number of questions about the implementation of quantity-price menus and the implications for trade policy.

5.2.1 Retail Market Power

Consider the two type model outlined above except now the manufacturer uses a single retailer to sell their products. Therefore the timing of the game is as follows. In the first stage the government sets trade policy while in stage two the monopolist announces its schedule of quantities and wholesale prices $(w_H,\ w_L)$. In stage 3 the distributors choose which bundles to offer and at what prices. In stage 4 consumers receive their preference draw from a common knowledge distribution, though the draw itself is private information, and trade occurs. Since we have already characterized the consumers decision we pick up the analysis in stage 3 and work backwards.

5.2.2 Retailers Optimal choice of products and prices

Taking the product quantities $\{q_L,\ q_H\}$ and wholesale prices $\{w_L,\ w_H\}$ as given, the retailer maximizes its profits by choosing which products to offer and the associated retail prices. Therefore the problem can be written as:

$$\begin{aligned}\max_{p_L,\ p_H,\ I_{LL},\ I_{LH},\ I_{HH},\ I_{HL}} \Pi_R = & \ \beta(p_H - w_H)I_{HH} + \beta(p_L - w_L)I_{LH} \\ & +(1-\beta)(p_H - w_H)I_{HL} + (1-\beta)(p_L - w_L)I_{LL} \end{aligned} \tag{36}$$

where I_{jk} is an indicator variable that equals one if bundle of quantity j is designed for sale to type k. Note that the problem faced by the retailer is one where the quantity range is given, therefore they are faced with a problem of choosing over a discrete number of options and finding the optimal prices associated with each option. Given the discrete nature of the problem and the constraints imposed by (4)-(7) then the potential configurations are as follows:

$$I_{HH} = I_{LL} = 1,\ p_H = \theta_H u(q_H) - (\theta_H - \theta_L)u(q_L),\ p_L = \theta_L u(q_L) \quad (37)$$

$$I_{HH} = 1,\ p_H = \theta_H u(q_H),\ p_L = \infty \quad (38)$$

$$I_{LH} = I_{LL} = 1,\ p_H = \infty,\ p_L = \theta_L u(q_L) \quad (39)$$

$$I_{HH} = I_{HL} = 1,\ p_H = \theta_L u(q_H),\ p_L = \infty \quad (40)$$

5.2.3 Foreign Monopolists choice of qualities and wholesale prices

The introduction of the retailer changes the nature of the foreign manufacturer's problem since not only does the producer need to take into account the choices of the consumers, it must do so through the filter of a retailer. Consequently, the constraints faced by the producer are no longer the participation and incentive constraints of the consumer directly, but the incentive and participation constraints of the retailer. To investigate when the foreign producer offers a full range of sizes requires solving the following program:

$$\max_{q_L,\ q_H,\ w_H,\ w_L} \Pi = \beta(w_H - (c+t)q_H) + (1-\beta)(w_L - (c+t)q_L) \quad (41)$$

subject to

$$\Pi_R(I_{HH} = I_{LL} = 1) \geq \Pi_R(I_{LH} = I_{LL} = 1) \quad (42)$$

$$\Pi_R(I_{HH} = I_{LL} = 1) \geq \Pi_R(I_{HH} = 1) \quad (43)$$

$$\Pi_R(I_{HH} = I_{LL} = 1) \geq \Pi_R(I_{HH} = I_{HL} = 1) \quad (44)$$

$$\Pi_R(I_{HH} = I_{LL} = 1) \geq \Pi_R(I_{LH} = 1) \quad (45)$$

$$\Pi_R(I_{HH} = I_{LL} = 1) \geq 0 \quad (46)$$

Using conditions (37)-(40) these constraints can be shown to imply:

$$w_H = \theta_H u(q_H) - \frac{1}{(1-\beta)}(\theta_H - \theta_L)u(q_L)$$

$$w_L = \frac{1}{(1-\beta)}\theta_L u(q_L) - \frac{\beta}{(1-\beta)}\theta_H u(q_L)$$

A couple of observations are in order. First, these wholesale prices are strictly less than those offered by the producer in the absence of a retail link. Obviously this is likely to have important implications for quantity choice and the desirability of producing the full range of product sizes. Second, since these wholesale prices are lower than the associated retail prices, the participation constraint does not bind for the retailer. In an equilibrium where both types are served a different quantity, the retailer earns $\frac{\beta}{(1-\beta)}(\theta_H - \theta_L)u(q_L)$. Note that this is increasing in β, decreasing in $(1-\beta)$ and decreasing in t.

Substituting in the optimal wholesale prices gives the following simplified program:

$$\max_{q_L,\, q_H} \Pi = \beta(\theta_H u(q_H) - (c+t)q_H) + (1-\beta)(\theta_L u(q_L) - (c+t)q_L) - \frac{\beta(2-\beta))}{(1-\beta)}(\theta_H - \theta_L)u(q_L)$$

Taking first order conditions generates:

$$\frac{\partial \Pi}{\partial q_L} = (\theta_L u'(q_L) - (c+t)) - \frac{\beta(2-\beta))}{(1-\beta)^2}(\theta_H - \theta_L)u'(q_L) = 0 \quad (47)$$

$$\frac{\partial \Pi}{\partial q_H} = \theta_H u'(q_H) - (c+t) = 0 \quad (48)$$

Somewhat surprisingly there is still no distortion for the high types, even though there is a double marginalization problem. Instead the full impact of the double marginalization is felt by the low types, that have the size of their bundle reduced further below the first best since $\frac{\beta(2-\beta))}{(1-\beta)^2} > \frac{\beta}{(1-\beta)}$.

5.2.4 Optimal Tariffs with a Foreign Monopolist and Domestic Retailer

In addition to consumer surplus and tariff revenue, the domestic economy also receives surplus from the profits of the domestic retailer. Given this additional source of surplus the domestic welfare function is written as:

$$W = \beta(\theta_H - \theta_L)u(q_L)\frac{(2-\beta)}{(1-\beta)} + TR \quad (49)$$

If tariffs are used that depend on the size of the bundle imported then the optimal tariffs have the following form:

$$t_H^* = \frac{-q_H}{dq_H/dt_H} \quad (50)$$

$$t_L^* = \frac{-q_L}{dq_L/dt_L}\left(1 - \frac{\beta(2-\beta)(\theta_H - \theta_L)}{(\theta_L - \beta(2-\beta)\theta_H)}\epsilon(q_L)\right) \tag{51}$$

Note that regardless of whether or not a retailer serves as an intermediary, the optimal tariff on the high type bundle is the revenue maximizing tariff. The tariff on the low type bundle also has a relatively familiar form, though since $(2-\beta) \geq 1$, a subsidy is optimal over a greater range of parameter values when there is a domestic retailer. If only a single rate is set for all imports then the optimal tariff is:

$$t^* = \frac{-Q}{dQ/dt}\left(1 - \frac{\beta(2-\beta)(\theta_H - \theta_L)}{(\theta_L - \beta(2-\beta)\theta_H)}s_L\epsilon(q_L)\right) \tag{52}$$

Once again the central role of β is evident.

5.3 Product Quality and Ad Valorem Tariffs

The analysis so far has considered the case of a homogeneous product that is sold to different types at different average prices. This requires that resale opportunities are restricted. While there are products where this property is a reasonable assumption primarily due to transaction costs involved (e.g. most people don't buy the larger quantity of soft drink in order to gain a quantity discount and then attempt to find another consumer to share both the drink and the expense), it may seem more natural to consider products that differ in quality. Such products have the property that they typically can't be divided for arbitrage purposes (e.g. the price of a 40 inch TV might be less than double the price of 20 inch TV, but you cannot divide a 40 inch screen in half to get two 20 inch screens). Adopting this interpretation requires very little change to the analysis of the decisions by consumers and the foreign monopolist. In this case, a consumer would like to buy one unit of a product and the variable q indexes the quality of the product. Similarly the foreign monopolist now chooses the quality that it provides to each type rather than the quantity. However, the representation of trade policy is altered by a quality interpretation. In particular, if specific tariffs are applied they are based on the quantity imported, which is one unit regardless of quality. Consequently, ad valorem tariffs that are based on the value of the imports are a more natural instrument to consider. To see how this changes the analysis, let τ denote the ad valorem tariff rate. In this case the objective function of the foreign monopolist and the resulting first order conditions can be expressed as:

$$\max_{q_L,\, q_H} \Pi = \beta\left((1-\tau)[\theta_H u(q_H) - \theta_H u(q_L) + \theta_L u(q_L)] - cq_H\right) + (1-\beta)((1-\tau)[\theta_L u(q_L)] - cq_L) \quad (53)$$

Taking first order conditions gives:

$$\begin{aligned} \frac{\partial \Pi}{\partial q_L} &= \beta(1-\tau)(\theta_H - \theta_L)u'(q_L) - ((1-\tau)\theta_L u'(q_L) - c))(1-\beta) = 0 \\ &\Rightarrow (\theta_L - \beta\theta_H)u'(q_L) - \frac{(1-\beta)c}{(1-\tau)} = 0 \quad (54) \\ \frac{\partial \Pi}{\partial q_H} &= \theta_H u'(q_H) - \frac{c}{(1-\tau)} = 0 \quad (55) \end{aligned}$$

Once again the high type is offered an efficient quality, conditional on the tariff, while the low type is offered an inefficiently low quality good.[25]

With an ad valorem tariff the government's objective function becomes:

$$\max_{\tau} \; W = \beta(\theta_H - \theta_L)u(q_L) + \tau\left((1-\beta)T_L + \beta T_H\right) \quad (56)$$

To characterize the optimal tariff, differentiate domestic welfare with respect to the import tax.

$$\frac{\partial W}{\partial \tau} = \beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial \tau} + T + \tau T' = 0 \quad (57)$$

where $T = (1-\beta)T_L + \beta T_H$ and $T' = \beta\frac{\partial T_H}{\partial q_L}\frac{\partial q_L}{\partial \tau} + (1-\beta)\frac{\partial T_L}{\partial q_L}\frac{\partial q_L}{\partial \tau} < 0$.

Solving for the optimal ad valorem tariff yields:

$$\begin{aligned} \tau^* &= \frac{-T}{T'}\left(1 + \frac{\beta(\theta_H - \theta_L)u'(q_L)\frac{\partial q_L}{\partial \tau}}{T}\right) \\ &= \frac{-T}{T'}\left(1 - \frac{\beta(\theta_H - \theta_L)}{(\theta_L - \beta\theta_H)}\frac{(1-\beta)q_L c}{T(1-\tau)^2}\epsilon(q_L)\right) \quad (58) \end{aligned}$$

Once again the sign of the optimal ad valorem tariff depends on the distribution of types. When $\beta = 0$ the optimal tariff is a tax and coincides with the revenue maximizing ad valorem rate. However as $\beta \to \frac{\theta_L}{\theta_H}$, the optimal rate declines and a subsidy becomes more likely. Furthermore, under a uniform price foreign monopoly, the optimal tariff is non-negative for any demand function

[25] Krishna (1990) examines the implications of exogenous trade frictions for the quality composition of imports when the foreign monopolist chooses the quality of its product line. However, she does not characterize the optimal policy.

that is less convex than a constant elasticity demand curve. In fact, for the constant elasticity of demand case, the optimal ad valorem tariff is zero. Consequently, the constant elasticity case serves as a useful benchmark and leads to the following proposition:

PROPOSITION 4 *If a foreign monopolist uses an optimal non-linear price schedule to quality discriminate between domestic consumers, and the domestic demand curves have a constant elasticity, then the optimal ad valorem tariff is positive for β close to zero and negative for β sufficiently close to $\frac{\theta_L}{\theta_H}$. This contrasts with the optimal ad valorem policy when facing a uniform price foreign monopolist which is free trade.*

6 Conclusion

While the practice of non-linear pricing is well documented in domestic models of industrial organization, the implications of such techniques for international transactions are relatively unexplored. This paper represents an attempt to fill this gap. It finds three things. First, there is at least some empirical support for the notion that non-linear pricing is used in international trade. Second, it finds that the calculus of welfare maximization is very different when the foreign monopolist implements second degree price discrimination. While there are still terms of trade effects from the imposition of a tariff, the existence of such effects are neither necessary nor sufficient for the government to intervene. Instead the distribution of valuations across the population is the key determinant of the nature of policy intervention. If there is a relatively large fraction of high types in the population, then domestic information rents can be increased substantially by subsidizing imports thereby increasing the consumption of the low valuation types and moving the incentive constraint in favor of the high valuation types. However, if the fraction of high types is relatively small then the increase in information rents will also be small but the fiscal implications of a subsidy will be large. Consequently, the optimal policy will be to impose a trade tax. Finally, the welfare magnitudes involved with the operation of trade policy are now much larger, with trade policy having first order implications for both the tariff revenue paid by the foreigner and also for domestic consumer surplus.

These finding have important implications for the conduct of trade policy more broadly and specifically for the design of trade agreements. The standard analysis of trade agreements places the terms of trade effect at the center of the analysis since this is the mechanism by which international externalities arise. However, the above analysis suggests there are settings in which the terms of trade are not driving trade policy. This raises the question of whether this new motivation for trade policy has additional implications for the design of trade agreements. The fact that the international externalities are much larger under non-linear pricing suggests that this maybe the case.

Another avenue to explore is the incentives to undertake FDI for a discriminating foreign monopoly, especially in the context of quality discrimination. The conditions for Proposition 4 provide a neat perspective from which to consider these incentives. If the monopolist employs linear prices, then the optimal policy is free trade and the foreign monopolist has no incentive to consider a strategy of FDI. However, with non-linear prices the optimal policy is not free trade, and the foreign monopolist may consider jumping the trade policy. If it does consider the option of FDI, the quality discrimination scenario raises a new twist on the standard tariff jumping model since the monopolist has a choice of whether or not to relocate both the high and low quality production lines or just one of the production lines. If only one production line is relocated, which will it be?

Reference

Antras, P. and Staiger, R. W. (2008). Offshoring and the value of trade agreements. manuscript.

Bagwell, K. and Staiger, R. W. (1989). The role of export subsidies when product quality is unknown. *Journal of International Economics*, *27*(1-2), 69–89.

Bagwell, K. and Staiger, R. W. (2002). *The Economics of the World Trading System*. MIT Press.

Bagwell, K. and Staiger, R. W. (2004). Multilateral trade negotiations, bilateral opportunism and the rules of gatt/wto. *Journal of International Economics*, *63*(1), 1–29.

Baldwin, R. and Harrigan, J. (2007). Zeros, quality and space: trade theory and trade evidence. NBER Working Papers 13214, National Bureau of Economic Research, Inc.

Bown, C. P. (2004). Trade policy under the gatt/wto: empirical evidence of the equal treatment rule. *Canadian Journal of Economics*, *37*(3), 678–720.

Brander, J. A. and Spencer, B. J. (1984). Trade warfare: tariffs and cartels. *Journal of International Economics*, *16*(3-4), 227–242.

Bureau of the Census (1990). *U.S. Exports and Imports of Merchandise on CD-ROM [machine-readable data file]*.

Busse, M. and Rysman, M. (2005). Competition and price discrimination in yellow pages advertising. *RAND Journal of Economics*, *36*(2), 378–390.

Disdier, A.-C. and Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics*, *90*(1), 37–48.

Ganslandt, M. and Maskus, K. E. (2007). Vertical distribution, parallel trade, and price divergence in integrated markets. *European Economic Review*, *51*(4), 943–970.

GOA (1995). *U.S. Imports: Unit Values Vary Widely for Identically Classified Commodities*. Report GAO/GGD-95-90.

Grossman, G. M. and Helpman, E. (2002). Integration versus outsourcing in industry equilibrium. *The Quarterly Journal of Economics*, *117*(1), 85–120.

Hummels, D. and Skiba, A. (2004). Shipping the good apples out? an empirical confirmation of the alchian-allen conjecture. *Journal of Political Economy*, *112*(6), 1384–1402.

Kolev, D. R. and Prusa, T. J. (1999). Tariff policy for a monopolist in a signaling game. *Journal of International Economics*, *49*(1), 51–76.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *RAND Journal of Economics*, *15*(2), 171–196.

McLaren, J. (2000). Globalization and vertical structure. *American Economic Review*, *90*(5), 1239–1254.

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

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, *48*(1), 7–35.

Saggi, K. (2006). Preferential trade agreements and multilateral tariff cooperation. *International Economic Review*, *47*(1), 29–57.

Schott, P. K. (2004). Across-product versus within-product specialization in international trade. *The Quarterly Journal of Economics*, *119*(2), 646–677.

Tirole, J. (1988). *The Theory of Industrial Organization.* MIT Press.