Vertical restraints in distribution and the price impact of parallel imports*

Mattias Ganslandt†, The Research Institute of Industrial Economics
Keith E. Maskus‡, University of Colorado at Boulder

Draft 2004-04-28

Abstract

We develop a model of vertical pricing in which an original manufacturer sets wholesale prices in two markets integrated at the distributor level by parallel imports (PI). The manufacturing firm needs to set these two prices to balance three competing interests: restricting competition in the PI-recipient market, avoiding resource wastes due to actual trade, and reducing the double-markup problem in the PI-source nation. These tradeoffs imply the counterintuitive result that both wholesale and retail prices could diverge as a result of declining trading costs, even as the volume of PI increases. Thus, it may be misleading to think of PI as a force for price integration. Further, a policy of requiring uniform wholesale prices across locations would push retail prices toward convergence as transportation costs fall but this may not be optimal.

*We are grateful to the Swedish Competition Authority for financial support. Thanks to Jim Markusen and Lars Persson for valuable comments.

†IUI (The Research Institute of Industrial Economics), P.O. Box 5501, SE-11485 Stockholm, Sweden e-mail mattias.ganslandt@iui.se, fax +46-8-6654599.

‡Department of Economics, Campus Box 256, University of Colorado, Boulder, CO 80309-0256, Phone (303) 4927588, e-mail: keith.maskus@colorado.edu

INTRODUCTION

Parallel imports (PI) are goods placed legitimately onto the market in one country but are subsequently imported into another country without the authorization of the holder of intellectual property protection (patent, copyright, or trademark) in the latter market. The activity is legal if the importing country follows a policy of international exhaustion, under which the first sale of a product anywhere ends the rights-holder's ability to control distribution. It is illegal if the country pursues national exhaustion, under which any sale abroad does not restrict rights to limit imports. The European Union has adopted a rigorous regime of regional exhaustion, which states that PI are impermissible from outside the community but first sale within its territory exhausts distribution rights. The essential justification for this policy is a belief that PI generate competition at the retail level, inducing a tendency toward retail price convergence and pro-competitive gains from price integration.

In this paper we explore fully the price impacts of PI in a setting where an original manufacturer (the holder of the intellectual property rights) sells its product through independent distributors in two national markets. Our point of departure is the observation that the bulk of PI actually takes place at the wholesale level, rather than the retail level (Maskus and Chen, 2002; National Economic Research Associates, 1999). Because manufacturing firms place goods on the market initially through vertical contracts with local distributors, it is important to study the implications of such contracts for PI volumes and the consequent scope for retail price integration. Indeed, we find within our simple model that a claim that PI are a means of bringing down retail prices in expensive locations is misleading. We show that, for an important and empirically relevant range of trade costs, the existence of parallel imports can cause retail prices to diverge, precisely opposite to the conventional intuition. Thus, a reorientation of competition policy within the EU toward understanding vertical competition is in order.

The existence of a vertical control problem in the context of PI was first analyzed by Maskus and Chen (2002). They noted three essential efficiency tradeoffs in a model where a licensed distributor in one market can trade goods back to the home market of the original manufacturer. In this framework, the manufacturer had to balance the losses from a pro-competitive price effect of PI into its own market, the resource costs wasted in the activity of parallel trade, and the double-markup problem in inducing a profit-maximizing retail price in the foreign market. The authors discovered a U-shaped welfare curve in the cost of trading PI goods. If trade costs are low, the optimal policy is to permit free parallel trade, for the pro-competitive gains dominate the other efficiency losses. However, at intermediate-to-high trade costs it is efficient to ban PI.

The model advanced here builds on these insights in several important ways. First, we permit vertical distributors in two markets, opening the possibility for dual double-markup problems for the manufacturer to solve. This possibility generates a richer menu of potential price effects even within the framework of Cournot competition. Second, we focus more explicitly on retail price impacts of parallel trade within the vertical-control context. The possibility of retail price divergence under open PI is counterintuitive and has not appeared in the literature before this analysis. Finally, we consider a richer and more realistic menu of policy choices. Specifically, we analyze one policy regime in which manufacturers cannot discriminate in the setting of wholesale prices and another in which the distributor cannot charge a price to a PI firm that puts it at a competitive disadvantage. Both of these policies exist in EU competition law and both may be sources of inefficiency in a world where manufacturers will use all legal means to counter the profit-reducing effects of parallel imports.

The paper is organized as follows. In the next section we briefly review relevant prior literature. In the following section we set out a model of parallel imports in two vertically controlled markets that are linked by the possibility of PI. In the fourth section we consider the implications of two uniform-pricing policies. We offer concluding remarks in the final section.

PRIOR LITERATURE

While parallel imports may exist for a variety of reasons, two canonical models have appeared in the economics literature. Arbitrage against differences in retail prices was the focus of early literature (Tarr 1985; Hilke 1988), which found that rapid and large dollar appreciation generated rising volumes of parallel imports into the United States. In an important theoretical paper, Malueg and Schwartz (1994) argued that a regime of uniform retail pricing would be globally sub-optimal to one in which firms could price-discriminate on the basis of countries grouped by demand elasticity. Also arguing from a standard pricing model, Richardson (2002) argued that neither a global policy of uniform pricing (international exhaustion) nor of full segmentation (national exhaustion) could be supported as a Nash equilibrium, suggesting that negotiations on an international regime at the World Trading Organization would be frustrated.

The study by National Economic Research Associates (1999) reported survey evidence of significant flows of parallel trade within the European Union in the early 1990s. While the report tended to focus on retail price differences, it pointed out that the bulk of parallel trade happens at the wholesale or distributor level. Thus, in the second main approach Maskus and Chen (2002) developed a simple model of vertical price control that had the features mentioned above. They also performed straightforward econometric analysis with detailed U.S. export-price data and discovered that the international distribution of wholesale export prices did follow the predicted U-shaped relationship with U.S. tariff rates and that the level of such prices across destination regions depended on the legal treatment of parallel trade. Thus, empirical evidence points to the vertical-control problem as being central to PI.

This line of inquiry was pursued extensively within the EU context by Ganslandt and Maskus (2003). Their empirical evidence supported the view that there are multiple causes for PI. Econometric analysis of European prices suggested that both horizontal arbitrage and vertical-control problems are important practical explanations for such trade. Pricing behavior by exporters from high-price markets - such as Denmark and the United Kingdom - indicated that such firms increase export prices in countries that are in close proximity (and therefore have low trade costs) in an attempt to deter PI.

Based on this literature, we believe that there is substantial scope for understanding better the implications of parallel trade by focusing on models of vertical pricing behavior in integrated markets. It is conceivable that wholesale prices may be set in a way that offsets or even counteracts the anticipated impacts of an open PI regime. This is the point of departure for the analysis that follows.

A THEORETICAL MODEL OF PARALLEL IMPORTS

A manufacturer, M, sells its product in two countries, A and B. Firm M sells its product through an independent exclusive distributor in country A and another independent distributor in country B. The demand in A is $Q_A = 1 - p$, and that in B is $Q_B = S(1 - bp)$.

For convenience, assume $b \geq 1$ and, hence, demand is more elastic in B (for any given price p). S is the population of market B, assuming that the population in A is normalized to 1. Manufacturer M has a constant marginal cost of production normalized to zero, and the marginal cost of retailing in both countries is normalized to zero as well.[1] We assume that individual purchasers do not engage in sufficient retail arbitrage to affect the pricing decisions analyzed here.

Suppose that the manufacturing firm M can offer the distributor in market i $(i = A, B)$ any contract in the form of (w_i, T_i), where w_i is the wholesale price at which the distributor purchases from M and T_i is a transfer payment (franchise fee) from the distributor to M. However, M cannot prevent the distributor in market B from selling the product also in market A, either directly or through intermediaries, such as firms specialized in parallel trade. That is, either M cannot legally limit the distributor's territory of sales, or it is too costly for M to enforce any such constraint. However, we assume that neither the manufacturer nor any of its agents other than the licensed distributor in market B can sell in that market.

Thus, the situation we have in mind describes a protected vertical contract in one market but a vertical contract that is vulnerable to wholesale-level competition in the other market. There are several practical justifications for such an assumption. Most readily, the countries may vary in their legal treatment of PI, with international exhaustion the rule in A and national exhaustion in B. Good examples for the former are Australia and Hong Kong, which are open to PI in copyright goods, and good examples of the latter would be Japan and the United States, which are not (Maskus 2000). Next, there may be asymmetric product standards between the two countries. If the permissible standards in A are more inclusive than those in B, this kind of one-way trade would be possible. This situation could describe Japan and the United States, which are both relatively open to PI in trademark goods but the United States has less rigorous technical requirements. Finally, if there are fixed costs in transporting PI and market B is small while market A is large, a similar asymmetry is possible. Portugal and the United Kingdom within the European Union would exemplify this case, which might become more prevalent with the entry in 2004 of Cyprus, Malta, Lithuania, Estonia, and Latvia.

With this setup, suppose that the distributor in B incurs an additional constant marginal cost $t \geq 0$ in selling the good in market A. For instance, t could be the transportation cost or tariff. Subsequently the marginal cost of distributor i in market A is denoted c_i^*, where $c_A^* = w_A$ and $c_B^* = w_B + t$.

Finally, assume that if the distributor in market B sells in market A, it will compete with the distributor there in a Cournot fashion. Let the quantities sold in A by the two distributors be q_A^* and q_B^*, respectively, and the quantity sold in B by the sole distributor be q_B. A subgame-perfect Nash equilibrium is a pair (q_A^*, q_B^*) that constitute a Nash equilibrium for any (w_i, T_i) for $i = A, B$, together with an optimal choice of q_B by the distributor in market B for any (w_B, T_B) and an optimal choice of (w_i, T_i) for $i = A, B$ by the manufacturing firm M. Let w denote the vector (w_A, w_B) and T denote the vector (T_A, T_B).

Our main objective is to analyze how the manufacturing firm sets the wholesale prices and the transfer payments to maximize its profit. The manufacturing firm's profit is equal to the total revenues in equilibrium minus real costs incurred. More precisely, the objective of the manufacturing firm is to maximize:

$$\Pi(w) = p_A(w)\left(q_A^*(w) + q_B^*(w)\right) - tq_B^*(w) + p_B(w)\, q_B(w) \tag{1}$$

where the first term on the right hand side is the total revenue in market A, the second term is the real cost of trade between the markets and the third term is the total revenue in market B.

Next, solving for quantities in any subgame equilibrium is straightforward. Assuming that at least one distributor can sell in market A, for any w and T accepted by the distributors, the subgame equilibrium for $i = A, B$ is given by the well-known Cournot solution (derived in the appendix):

$$q_i^*(w) = \begin{cases} \frac{1-2c_i^*+c_{-i}^*}{3} & \text{if } 2c_{-i}^* - 1 < c_i^* < \frac{1+c_{-i}^*}{2} \\ \frac{1-c_i^*}{2} & \text{if } c_i^* \leq 2c_{-i}^* - 1 \\ 0 & \text{if } \frac{1+c_{-i}^*}{2} \leq c_i^* \end{cases} \tag{2}$$

where i refers to distributor i and $-i$ to the other distributor. The lower bound of the marginal cost of distributor i given in the first line ensures that the other distributor chooses a positive quantity in equilibrium while the upper bound ensures that distributor i sells a positive quantity. When the former condition is violated, distributor i would be the sole supplier in market A. When the latter condition is violated the other distributor would be the only seller active in equilibrium. The equilibrium quantities $q_A^*(w)$ and $q_B^*(w)$ in (2) give a retail price in market A for any w :

$$p_A(w) = \begin{cases} \frac{1+w_A+w_B+t}{3} & \text{if } 2w_A - 1 < w_B + t < \frac{1+w_A}{2} \\ \frac{1+w_B+t}{2} & \text{if } w_B + t \leq 2w_A - 1 \\ \frac{1+w_A}{2} & \text{if } \frac{1+w_A}{2} \leq w_B + t \end{cases} . \tag{3}$$

Proceeding to market B the distributor chooses a profit-maximizing quantity equal to:

$$q_B(w) = \begin{cases} S\left(\frac{1-bw_B}{2}\right) & \text{if } w_B < \frac{1}{b} \\ 0 & \text{if } w_B \geq \frac{1}{b} \end{cases} , \tag{4}$$

where the condition for the wholesale price ensures that the distributor sells a positive quantity in that market. The retail price is therefore

$$p_B(w) = \frac{1+bw_B}{2b} \quad \text{if } w_B < \frac{1}{b} \tag{5}$$

while the retail price is prohibitive for a higher wholesale price, i.e. $w_B > 1/b$.

The quantities given in equation (2) and equation (4) characterize all subgame equilibria and we can proceed to analyze the decision of the manufacturing firm. Three cases are of particular relevance: the accommodation equilibrium, the arbitrage-free equilibrium and the segmented equilibrium.

In the accommodation equilibrium the manufacturing firm sets wholesale prices to accommodate the effects of arbitrage. Accordingly, parallel imports occur in equilibrium. The quantities chosen by distributors are all positive. Inserting subgame quantities from equations (2) and (4) as well as retail prices from (3) and (5) in the profit function of the manufacturing firm, given by expression (1), we obtain

$$\begin{aligned} \Pi^{A}(w) &= \left(\frac{1+w_A+w_B+t}{3}\right)\left(\frac{2-w_A-w_B-t}{3}\right) \\ &\quad -t\left(\frac{1+w_A-2w_B-2t}{3}\right)+\frac{S}{b}\left(\frac{1-b^2w_B^2}{4}\right) \end{aligned} \tag{6}$$

where the first term is the revenue in market A, the second term is the trade cost and the third term is the revenue in market B.

In the arbitrage-free equilibrium, on the other hand, parallel trade is unprofitable and no arbitrage occurs. More precisely, the quantity sold in market A by distributor B is zero, i.e. $q_B^*(w)=0$. For this to be true the retail price in market A (p_A) minus the wholesale price in market B (w_B) must be no greater than the unit trade cost (t),

$$p_A(w)-w_B \leq t \tag{7}$$

which may be referred to as the no-arbitrage condition. Hence, in the arbitrage-free equilibrium the manufacturing firm maximizes its profit

$$\begin{aligned} \Pi^{D}(w) &= \left(\frac{1-w_A^2}{4}\right)+\frac{S}{b}\left(\frac{1-b^2w_B^2}{4}\right) \\ &\quad \text{s.t. } p_A(w)-w_B \leq t \end{aligned} \tag{8}$$

where the first term is the revenue in market A and the second term is the revenue in market B.

In the segmented equilibrium, finally, the distributor in market B can only sell in that market. This may be due either to a prohibitive trade cost or to legal restrictions, as noted above. In the segmented equilibrium the manufacturing firm can maximize its profit without a constraining arbitrage condition. This is equivalent to maximizing the profit in the arbitrage-free equilibrium (8) without the no-arbitrage condition. We refer to the markets as segmented when this condition is slack in equilibrium, even if it is not necessarily slack out-of-equilibrium. We denote the profit in the segmented equilibrium $\Pi^S(w)$.

THE PARALLEL TRADE EQUILIBRIUM

In the parallel trade (accommodation) equilibrium the manufacturing firm sets one wholesale price in market A and another in market B to maximize profits, even though arbitrage is profitable in this solution. In this equilibrium the distributor in market B finds it profitable to sell a positive quantity in market A and parallel imports occur in equilibrium. To find the optimal wholesale prices we differentiate the accommodation profit $\Pi^A(w)$ with respect to wholesale prices in the two markets.

The first order condition with respect to the wholesale price in market A, w_A, is

$$\frac{d\Pi^A(w)}{dw_A} = \left[\frac{1}{9} - \frac{2}{9}w_A - \frac{2}{9}w_B - \frac{2}{9}t\right] - \left[\frac{1}{3}t\right] = 0, \tag{9}$$

where the first term in square brackets is the "pro-competitive effect" of parallel imports in market A. That is, it characterizes the manufacturer's incentive to control the total supply by both distributors in that market. The second term is the "trade-cost effect", i.e. it captures the incentive to save resources wasted in parallel trade.

It follows immediately from the first order condition with respect to the wholesale price in market A, w_A, that the pro-competitive effect is strong when the unit trade cost is low. The incentive to moderate the pro-competitive effect of arbitrage and to charge a high wholesale price is, consequently, strong when the trade cost is low. On the other hand, the cost effect is negative when the trade cost is positive and a high unit trade cost gives a strong negative trade-cost effect. The incentive to reduce the volume of arbitrage and charge a low wholesale price in market A is therefore stronger when the trade cost is high.

Correspondingly, the first order condition with respect to the wholesale price in market B, w_B, is

$$\frac{d\Pi^A(w)}{dw_B} = \left[\frac{1}{9} - \frac{2}{9}w_A - \frac{2}{9}w_B - \frac{2}{9}t\right] + \left[\frac{2}{3}t\right] - \left[\frac{Sbw_B}{2}\right] = 0, \tag{10}$$

and again the first term in square brackets is the "pro-competitive effect" of parallel imports in market A and the second term in square brackets is the "trade-cost effect". The third term is the "double-markup effect" in market B, referring to the inability of the manufacturer to set a wholesale price of zero in that market when PI are positive.

We immediately observe that the pro-competitive effect is identical in both (9) and (10). This result implies that it is only the sum of wholesale prices in the two markets that is important for the pro-competitive effect and the relative size of the two wholesale prices does not matter. In other words, the total supply in market A can be controlled by setting either the wholesale price in A, the wholesale price in B, or both. In addition, we note that the trade-cost effect is positive. To moderate the volume of arbitrage for a low t the manufacturing firm would consequently charge a high wholesale price in market B, while simultaneously charging a low wholesale price in market A.

The third term in square brackets does not appear in the first order condition with respect to the wholesale price in market A and is the double-markup effect in market B. If the manufacturing firm charges a positive wholesale price in B it suffers a double-markup problem there. The distributor in market B is a monopolist there and the retail price is consequently above the optimal (revenue-maximizing) level for the manufacturing firm whenever the wholesale price is positive. The manufacturing firm, thus, has an incentive to keep the wholesale price in B as low as possible in order to minimize the double-markup problem.

The pro-competitive effect, the trade-cost effect and the double-markup effect can be summarized as follows. The manufacturing firm has an incentive to charge a positive wholesale price in both markets to moderate the pro-competitive effect of arbitrage in market A. Moreover, the manufacturing firm has an incentive to charge a low wholesale price in A and a high wholesale price in B to reduce the volume of parallel imports and save trade costs. However, the manufacturer has the opposing incentive to keep the wholesale price in market B as low as possible to minimize the double-markup problem.

Having analyzed the incentives of the manufacturing firm we can solve the system of equations consisting of the two first order conditions (9) and (10) to find the optimal wholesale prices. The system has a unique solution:

$$w_A = \frac{1-5t}{2} - \frac{2t}{Sb}, \tag{11}$$

$$w_B = \frac{2t}{Sb}, \tag{12}$$

where both prices are continuous and linear functions in the unit trade cost t. The wholesale price in market A (B) is a decreasing (increasing) function of the trade cost. For low trade costs the wholesale prices in both markets are positive. However, for a high unit trade cost the wholesale price in market A becomes negative, meaning that the manufacturing firm sponsors the supply of distributor A in that market through a variable subsidy and recoups this subsidy by charging a higher fixed fee.

The result that wholesale prices differ is worth commentary. The manufacturer may be expected to use all legal means available to limit the negative effects of PI on its profits. These impacts include competition in market A, resource wastes in undertaking parallel trade, and an inability to set wholesale prices to avoid double-markup inefficiencies in market B. In essence, for this purpose the manufacturing firm has two instruments available: the wholesale prices it sets in the two markets. The firm will set these prices to balance these various impacts on its profit flow. It is evident that the firm would prefer the freedom to set different wholesale prices than be constrained to set a uniform price.

Seen in this light, our results support two observations. First, the manufacturing firm would choose a wholesale price in market A to moderate the pro-competitive effect of parallel imports. Simultaneously it would choose a wholesale price in market B to limit the volume of PI and avoid the waste of resources such trade incurs. The lower are the trade costs, the stronger is the incentive to reduce the pro-competitive effect by raising the wholesale price in A (note that w_A rises as t falls). Thus, for low trade costs, few resources are wasted by PI and the manufacturer would reduce competition by increasing the wholesale price in A as costs fall further. Second, however declining trade costs reduce the waste from PI and permit the firm to decrease the wholesale price in market B and reduce the double-markup distortion. Stated in the reverse fashion, the manufacturing firm would raise the wholesale price in market B in order to reduce PI when trade costs increase, despite the worsened double mark-up effect in that market. Thus, in the range of low trade costs the wholesale prices in markets A and B would diverge rather than converge as trade costs are reduced, which is opposite to the typical effect of arbitrage in price-discrimination models.

The volume of arbitrage is a function of wholesale prices. We insert the accommodation equilibrium wholesale prices (11) and (12) in equation (2), i.e. the quantity set by distributor B in market A, to obtain the volume of parallel imports:

$$q_B^* = \frac{1}{2} - \frac{2}{Sb}t - \frac{3}{2}t \tag{13}$$

which is also a continuous and linear function in the unit trade cost. The volume of arbitrage is a decreasing function in the unit trade cost. In other words, as the variable trade cost falls the volume of arbitrage increases at the same time as the wholesale prices diverge. A growing volume of arbitrage consequently goes hand in hand with price divergence between markets, again a result at variance with the usual arbitrage models.

It is also worth noting that the volume of arbitrage is an increasing function in the size S as well as the price-elasticity (captured by parameter b) in market B. The manufacturing firm has an incentive to keep the wholesale price in that market low when the size of that market is large or the customers are price-sensitive. In this case the manufacturing firm is willing to accept a larger volume of arbitrage to avoid a serious double markup problem in B.

The retail prices in markets A and B can be obtained from (3) and (5):

$$p_A = \frac{1}{2} - \frac{1}{2}t, \tag{14}$$

$$p_B = \frac{S + 2t}{2Sb} \tag{15}$$

which correspond to the pattern of wholesale prices. Thus, in the range of small trade costs a decline in t generates a divergence in retail prices, as it does in wholesale prices. The most interesting aspect of the retail prices in the parallel trade equilibrium is that at a unit trade cost of zero the prices are identical to the retail prices that would be set by a vertically integrated monopolist. In other words, the manufacturing firm can solve the vertical control problem perfectly if no real resources are wasted in arbitrage. The wholesale price in market B would be set to zero and the wholesale price in market A would be set at a prohibitive level for the distributor there, pushing it out of business. Distributor B would sell the revenue-maximizing quantity in both markets and A would be supplied from B but no real resources would be used in transportation. In addition, there would be no double markup problem in market B. Consequently, the retail prices in an integrated equilibrium would be identical to the prices set in a completely segmented equilibrium.

THE ARBITRAGE-FREE EQUILIBRIUM

In the arbitrage-free equilibrium the manufacturing firm sets wholesale prices in markets A and B such that arbitrage is unprofitable. In this equilibrium the distributor in B finds it profitable to sell a positive quantity in her own market but would not ship goods to market A and the markets are arbitrage-free. To find the optimal wholesale prices we differentiate the arbitrage-free profit function $\Pi^D(w)$ with respect to wholesale prices in the two markets, subject to the no-arbitrage condition, i.e. $p_A(w) - w_B \leq t$.

Consider a situation in which arbitrage would occur in equilibrium if the manufacturing firm charged accommodation prices. In this case, the no-arbitrage condition is not slack but binding and the wholesale price in market B is, therefore, a function of the wholesale price in market A. Inserting the retail price (3) in the no-arbitrage condition we obtain the following relationship between wholesale prices in the arbitrage-free equilibrium:

$$w_B(w_A) = \frac{1 + w_A}{2} - t \tag{16}$$

where the first term on the right hand side is the retail price in market A in an arbitrage-free situation, i.e. $q_B^*(w) = 0$. We can now replace the wholesale price in B in the arbitrage-free profit function $\Pi^D(w)$, with the function in equation (16) and differentiate with respect to the wholesale price in A. The first order condition with respect to the wholesale price in market A is

$$\frac{d\Pi^D(w)}{dw_A} = -\left[\frac{w_A}{2}\right] - \left[\frac{Sb}{4}\left(\frac{1}{2} - t + \frac{w_A}{2}\right)\right] = 0 \tag{17}$$

where the first term in square brackets reflects the effect of potential competition in market A when no arbitrage occurs in equilibrium and the second term reflects the double markup problem in market B.

The manufacturing firm would set a wholesale price in market B to block PI completely, which saves resources in trade. The higher the trade cost, the more protected is market A by the natural barrier and, further, the wholesale price in market B can be lowered to reduce the double markup effect there. The wholesale price would be set to block PI while limiting the price distortion in market B. When t reaches its prohibitive level (t=1/2), the wholesale price becomes zero and the price distortion disappears.

Next we can solve the first order condition (17) to find the optimal wholesale prices. The solution is unique and arbitrage-free equilibrium wholesale prices are:

$$w_A = \frac{Sb\,(2t-1)}{4+Sb}, \tag{18}$$

$$w_B = \frac{2\,(1-2t)}{4+Sb}, \tag{19}$$

where both prices are continuous and linear functions in the unit trade cost (t). Contrary to the wholesale prices in the parallel trade equilibrium, the wholesale prices in the arbitrage-free equilibrium diverge as the unit trade cost increases.

In this case the wholesale price in A is negative for any non-prohibitive unit trade cost and it is an increasing function in t. The intuition for this result again is that the manufacturer has an incentive to sponsor distributor A with a negative wholesale price to keep the markets arbitrage-free. The potential competition from parallel trade, however, results in a price which is below the revenue-maximizing monopoly level in a segmented situation. The manufacturing firm consequently has an incentive not to sponsor the distributor in market A more than is necessary. Moreover, the unit trade cost moderates the potential competition from parallel trade and the manufacturing firm can thus charge a higher price in A when the trade cost is high. As a consequence the pro-competitive effect is weaker for high unit trade costs.

In market B, on the other hand, the wholesale price is positive and a decreasing function of t. A positive wholesale price in B results in a double markup problem there. The manufacturing firm has an incentive to set the wholesale price in this market at the lowest possible level. When the unit trade cost is high markets are segmented by a natural barrier and the manufacturing firm can charge a low wholesale price in market B. When the unit trade cost is low, the manufacturer must charge a high wholesale price there (in combination with a low wholesale price in A) to keep markets arbitrage-free.

The wholesale prices can be inserted in (3) and (5) to obtain the retail prices in both markets in the arbitrage-free equilibrium:

$$p_A = \frac{2+Sbt}{4+Sb}, \tag{20}$$

$$p_B = \frac{(2+S)\,b+4\,(1-bt)}{2\,(4+Sb)\,b}, \tag{21}$$

and again both prices are linear in the unit trade cost. The retail price in market A rises and the retail price in B falls in the unit trade cost. Thus, retail prices diverge in the arbitrage-free equilibrium when the trade cost increases. This should be expected, for the potential competition from arbitrage is strong when the trade cost is low and weak when the trade cost is high.

PARALLEL IMPORTS AND TRADE COSTS

So far we have studied the parallel trade equilibrium and the arbitrage-free equilibrium separately, without specifying the parameter values for which one or the other is indeed the equilibrium. This section focuses on which of the two possible candidate set of prices (accommodation prices or arbitrage-free prices) the manufacturing firm would choose in equilibrium for a given unit trade cost t.

The manufacturer would set accommodation prices rather than arbitrage-free prices if the accommodation profit is higher than the arbitrage-free profit and vice-versa. We can determine the manufacturing firm's behavior in equilibrium by comparing its profits in the two cases for a given set of parameters $\{S, b, t\}$. The firm's profit with accommodation prices is:

$$\Pi^A\,(S,b,t) = \left(\frac{1}{4}-\frac{1}{4}t^2\right) - t\left(\frac{1}{2}-\frac{2}{Sb}t-\frac{3}{2}t\right) + \frac{S}{b}\left(\frac{1}{4}-\frac{t^2}{S^2}\right) \tag{22}$$

and, correspondingly, the profit with arbitrage-free prices is

$$\Pi^D\,(S,b,t) = \frac{1}{4}\left(1-\left(\frac{Sb\,(2t-1)}{4+Sb}\right)^2\right) + \frac{S}{4b}\left(1-b^2\left(\frac{2\,(1-2t)}{4+Sb}\right)^2\right). \tag{23}$$

Both functions are continuous in S, b and t. For any S and b it is possible to find a t such that the accomodation and arbitrage-free prices yield the same profit for the manufacturing firm. We refer to this unit trade cost as the threshold level. Set $\Pi^A(S,b,t) = \Pi^D(S,b,t)$ and solve for t to obtain

$$\widetilde{t} = \frac{Sb}{4+3Sb}, \tag{24}$$

which takes values on the open interval $(1/7, 1/3)$ for $Sb > 1$. The threshold level is consequently strictly positive and strictly less than the prohibitive unit trade cost $(t = 1/2)$ for relevant values of S and b. It can easily be checked that this is also the trade cost at which the distributor in market B finds it just unprofitable to ship goods to market A at accommodation prices. In other words, no arbitrage takes place in equilibrium even at accommodation prices and the no-arbitrage condition is binding, i.e. $p_A(w) - w_B = \widetilde{t}$.

The manufacturing firm finds it profitable to accommodate parallel trade at a lower unit trade cost, i.e. $t < \widetilde{t}$, while arbitrage-free prices result in a higher profit for a higher unit trade cost, i.e. $t > \widetilde{t}$. The intuition for this result is that arbitrage-free prices would result in a large pro-competitive effect in market A and a large double markup problem in market B for very low unit trade costs. At the same time the manufacturing firm could solve the vertical control problem quite well with accommodation prices. Very little real resources are used in parallel trade and accommodation is consequently more profitable for low unit trade costs.

For high trade costs, on the other hand, arbitrage-free prices are relatively close to segmented prices and the pro-competitive effect in market A and the double markup problem in market B are both small. Accommodation prices, however, would result in significant shipping of goods between the markets and wasted resources in arbitrage. Accordingly, arbitrage-free prices are more profitable than accommodation prices at high trade costs. We have the following result:

Proposition 1 *The model has a unique equilibrium. For unit trade costs below the threshold level* $\widetilde{t}$*, the manufacturing firm sets accommodation prices and parallel trade occurs. For unit trade costs above thre threshold level* $\widetilde{t}$*, the manufacturing firm sets arbitrage-free prices and no parallel trade occurs. The equilibrium values of* w_A *and*

w_B *are given by*

$$w = \begin{cases} w_A = \frac{1-5t}{2} - \frac{2t}{Sb} & w_B = \frac{2t}{Sb} \quad if \quad t < \widetilde{t} \\ w_A = \frac{Sb(2t-1)}{4+Sb} & w_B = \frac{2(1-2t)}{4+Sb} \quad if \quad t \geq \widetilde{t} \end{cases}. \tag{25}$$

Further, it is straightfoward to find the corresponding retail prices. We simply insert the equilibrium wholesale prices from the proposition above in the optimal retail price functions. The equilibrium retail prices in Country A and B are given by

$$p = \begin{cases} p_A = \frac{1}{2} - \frac{1}{2}t, & p_B = \frac{S+2t}{2Sb} \quad if \quad t < \widetilde{t} \\ p_A = \frac{2+Sbt}{4+Sb}, & p_B = \frac{(2+S)b+4(1-bt)}{2(4+Sb)b} \quad if \quad t \geq \widetilde{t} \end{cases}. \tag{26}$$

Thus, in the range of low trade costs the wholesale prices set by the manufacturing firm imply that the retail price in A *(B)* would decrease (increase) as trade costs rise. As a result, retail prices actually would diverge as trade costs decline in this range. In the limit, as trade costs approach zero (a situation in which markets are completely integrated for firms while still segmented for consumers) the retail prices would approach the same retail prices that would pertain in a completely segmented equilibrium. This surprising result stems from the fact that the manufacturing firm can solve the vertical distribution problem perfectly by letting one retailer, in this case retailer B, supply both markets and set the profit-maximizing monopoly prices in both markets. However, as trade costs rise within the range of low t, the retail price in market A increasingly would fall below its monopoly level due to the pro-competitive effect of PI. The intuition for this result is that the manufacturing firm would set wholesale prices simultaneously to reduce the volume of PI and restrict this pro-competitive effect. To put this result in simpler terms, the existence of trade costs induces the firm to have two distributors to reduce shipping costs but it then faces the threat of competition.

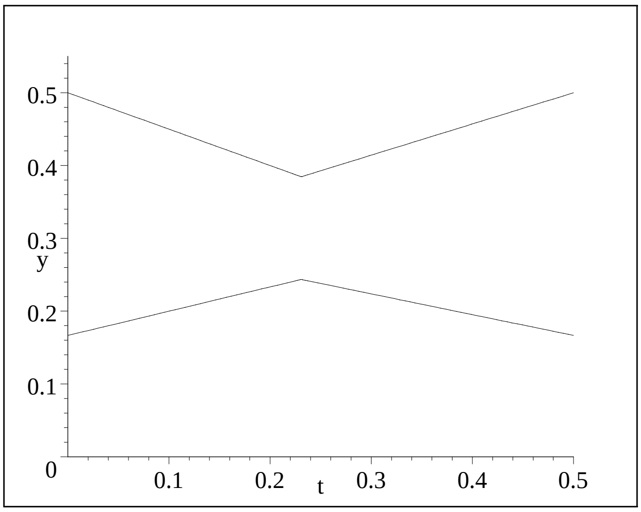

Fig 1: Retail prices in market A and market B.

In the range of low trade costs it is profitable for the manufacturing firm to reduce the volume of PI with a combination of a lower wholesale price in market A and a higher wholesale price in market B as costs go up. The effect on the retail price in market A of the lower wholesale price in that market would dominate the effect of the higher wholesale price in market B. As a result,the retail price would fall below its monopoly level. Also remember that the volume of parallel imports is:

$$q_B^* = \frac{1}{2} - \left(\frac{4+3Sb}{2Sb}\right) t \tag{27}$$

which is positive för unit trade cost $t < \widetilde{t}$ and negative otherwise. We thus have:

Corollary 1 *For low trade costs, $t < \widetilde{t}$, there is parallel importing from country B to country A in equilibrium and the difference in retail prices between markets A and B is a decreasing function in unit trade cost t.*

In contrast, in the range of high trade costs the retail price in B would fall with t while it would increase in A (see Figure 1). The retail price in market B would also decline with trade costs since the manufacturing firm sets a wholesale price in this market to eliminate PI. This positive wholesale price reflects a double markup distortion that diminishes with higher t. Thus, for higher trade costs the manufacturing firm would eliminate PI with a lower wholesale price and the corresponding double markup problem would be weaker. For high trade costs we have:

Corollary 2 *For high trade costs, $t > \widetilde{t}$,there is no parallel importing from country B to country A in equilibrium and the difference in retail prices between markets A and B is a strictly increasing function in t.*

It is evident that parallel imports reduce the profits of the manufacturer, because they create competition in the country receiving them, incur additional transactions costs and prevent the manufacturer from achieving efficient vertical pricing (see Figure 2). When the manufacturer is unable effectively to impose territorial restraints, it can still reduce or eliminate PI by raising the wholesale price to the distributor, but this leads to a less profitable retail price in the country where the imports originate. In equilibrium, the manufacturer balances the need to exercise optimal vertical price control and to limit parallel imports.

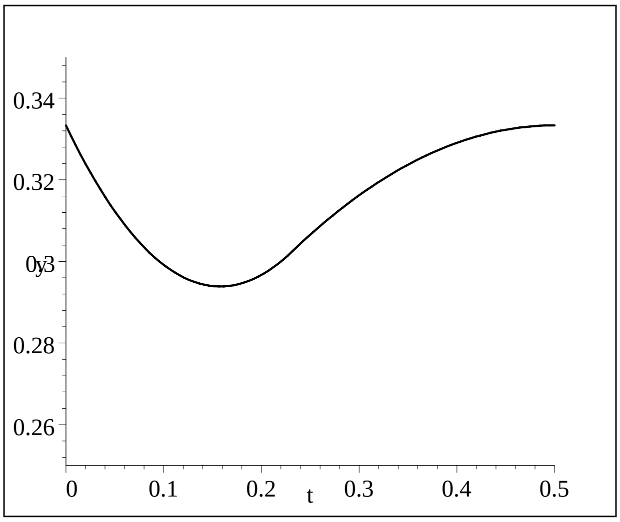

Fig 2: The manufacturing firm's profit.

More formally, we differentiate the manufacturer's profit in the accommodation equilibrium with respect to the trade cost,

$$\frac{d\Pi^A(S,b,t)}{dt} = \left(\frac{4+5Sb}{2Sb}\right)t - \frac{1}{2} \tag{28}$$

to get the following result

Corollary 3 *The manufacturing firm's profit decreases in the trade cost t when $t < \frac{Sb}{4+5Sb}$, increases in t when $t > \frac{Sb}{4+5Sb}$, and has two global maxima at $t = 0$ and $t \geq \frac{1}{2}$.*

It is worth noting that the manufaturing firm's profit has its lowest level in the range of accommodation prices (see Figure 2). In this range real resources are wasted in parallel trade at the same time as the manufacturing firm is unable to achieve efficient vertical pricing.

NON-NEGATIVE WHOLESALE PRICES

Previously we considered the possibility that the manufacturing firm could charge both positive and negative wholesale prices to control distribution in markets A and B. It may be more reasonable to assume that the manufacturer can only set non-negative wholesale prices. This would change the equilibrium for intermediate and high unit trade costs, which is the range where the manufacturing firm would prefer to charge a negative wholesale price in A if possible.

The unit trade cost for which the accomodation price in market A is zero can be found by setting (11) to zero. The relevant threshold level for the unit trade cost is

$$\underline{t} = \frac{Sb}{4+5Sb} \tag{29}$$

and this is also the unit trade cost for which the manufacturing firm's profit is lowest. For unit trade costs above the threshold level, $\underline{t}$, the firm maximizes its profit subject to the constraint that the wholesale price in market A is zero. This constraint adds an additional distortion and further complicates the vertical control problem.

Consider the intermediate range of trade costs ($t > \underline{t}$), which imply significant losses from resource waste in PI. The first order condition for the manufacturing firm is

$$\frac{d\Pi^A(w)}{dw_B} = \left[\frac{1}{9} - \frac{2}{9}w_B - \frac{2}{9}t\right] + \left[\frac{2}{3}t\right] - \left[\frac{1}{2}Sbw_B\right] = 0 \tag{30}$$

where as before the first term reflects the pro-competitive effect in A, the second term is the trade-cost effect and the third term reflects the double markup problem in B.

The important difference compared to the first order condition when the manufacturing firm could charge a negative wholesale price, in eq (10), resides in the first term. The manufacturing firm has an incentive to charge a lower wholesale price in market B when $w_A = 0$, to avoid generating a double markup problem in market A. Hence, the wholesale price in A is higher, and that in B lower, than they are when w_A can be negative. The accommodation wholesale prices for intermediate trade costs are

$$w_A = 0 \quad w_B = 2\left(\tfrac{1+4t}{4+9Sb}\right) \ . \tag{31}$$

As t rises within this range, the manufacturer would reduce the volume of PI by raising the wholesale price in market B. However, it is not profitable for the manufacturing firm to block PI completely by this strategy because it would create an excessively large double markup effect and consequently reduce the firm's profit. The manufacturer cannot limit the scope for PI by reducing the price in market A further as the wholesale price must be non-negative. The higher the trade cost, the stronger is the incentive to reduce the volume of PI. This incentive is a combination of two effects: a higher trade cost would waste more resources in PI but also would protect market A from competition. The latter effect in turn would reduce the net cost of limiting PI with an inefficient wholesale price in market B. Accordingly, the wholesale price in B would increase with trade costs while the wholesale price in A would be kept at the minimum level of zero.

Finally, consider the range of high trade costs. The manufacturing firm would set a wholesale price in market B to eliminate PI completely. The no-arbitrage condition (7) in combination with a non-negative price in A give the optimal wholesale prices

$$w_A = 0 \quad w_B = \tfrac{1}{2} - t \ . \tag{32}$$

The higher the trade cost, the more protected is market A by the natural barrier and, further, the wholesale price in market B can be lowered to reduce the double markup effect there. This price would be set both to block PI and limit the price distortion in B. When t reaches its maximum level, the wholesale price becomes zero and the price distortion disappears.

The upper threshold level for which the accommodation price is arbitrage free (subject to the constraint $w_A = 0$) can be computed by setting the wholesale price in B in the accommodation case (31) equal to the wholesale price in B in the arbitrage-free case (32). Accordingly, the upper threshold level is

$$\bar{t} = \frac{3}{2}\left(\frac{Sb}{4+3Sb}\right) = \frac{3}{2}\widetilde{t}, \tag{33}$$

where $\widetilde{t}$ is the critical threshold level for arbitrage-free prices when the manufacturing firm can charge a negative price in market A. It follows immediately that wholesale prices are arbitrage-free for substantially lower trade costs when the manufacturing firm is free to set both positive and negative wholesale prices.

For the case of non-negative wholesale prices we have the following result:

Proposition 2 *The model has a unique subgame perfect Nash equilibrium with non-negative wholesale prices, i.e.* $w_i \geq 0$ *for* $i = A, B$. *The equilibrium values of* w_A *and* w_B *are given by*

$$w = \begin{cases} w_A = \left(\frac{1-5t}{2} - \frac{2t}{Sb}\right) & w_B = \frac{2t}{Sb} & if \quad t < \underline{t} \\ w_A = 0 & w_B = 2\left(\frac{1+4t}{4+9Sb}\right) & if \quad \underline{t} \leq t < \bar{t} \\ w_A = 0 & w_B = \frac{1}{2} - t & if \quad t \geq \bar{t} \end{cases}. \tag{34}$$

Further, it is straightfoward to find the corresponding retail prices. We simply insert the equilibrium wholesale prices from the proposition above in the optimal retail price functions. The equilibrium retail prices in Country A and B are given by

$$p = \begin{cases} p_A = \frac{1}{2} - \frac{1}{2}t, & p_B = \frac{S+2t}{2Sb} & if \quad t < \underline{t} \\ p_A = \frac{2+4t+3Sb(1+t)}{4+9Sb}, & p_B = \frac{4+9Sb+2b+8tb}{2(4+9Sb)b} & if \quad \underline{t} \leq t < \bar{t} \\ p_A = \frac{1}{2}, & p_B = \frac{1}{4} + \frac{1}{2b} - \frac{t}{2} & if \quad t \geq \bar{t} \end{cases}. \tag{35}$$

Thus, in the range of low trade costs the wholesale prices set by the manufacturer imply that the retail price in A (B) would decrease (increase) as trade costs rise (see Figure 3). This range corresponds to the case analyzed in the section above (without restrictions on the wholesale price in market A).

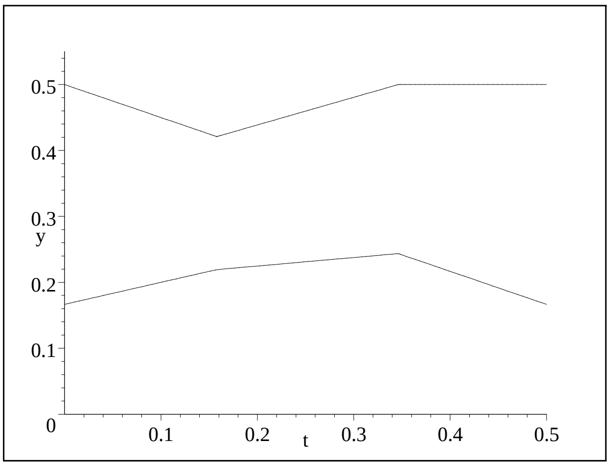

Fig 3: Retail prices in market A and B ($w_A \geq 0$)

Next, the range of intermediate trade costs is particularly interesting. In this range, both retail prices would rise with trade costs (see Figure 3). Moreover, for declining trade costs the retail price difference between markets would decrease. These effects follow from the manufacturing firm's optimal strategy to set a wholesale price in market B to reduce the volume of PI while keeping the wholesale price in market A at its minimum level. For lower trade costs the manufacturer firm would accommodate a larger volume of PI. More precisely, the firm would set a lower wholesale price in market B for lower trade costs. As a result, the pro-competitive effect of PI in market A would be stronger and the double markup effect in market B would be weaker. The retail prices in both markets consequently would fall as trade costs go down.

Finally, in the range of high trade costs the retail price in B would fall with trade costs while it would remain constant at the monopoly level in A (see Figure 3). The retail price in B would decline with trade costs because the manufacturing firm sets a wholesale price in this market to block PI. This wholesale price reflects a double-markup distortion that diminishes with higher t. Thus, for higher trade costs the manufacturing firm would eliminate PI with a lower wholesale price and experience a correspondingly weaker double.

It is straightforward to see that this arbitrage-free equilibrium is Pareto-dominated by a completely segmented equilibrium. This follows directly from the fact that the retail price in market B is higher than the segmented retail price (because of the remaining double-markup problem), the retail price in market A equals the segmented retail price and the profit of the manufacturing firm is lower than it would be under full segmentation.

CONSUMER WELFARE

Consumer welfare in markets A and B depends on the consumer prices in both markets. Aggregated consumer surplus in the two markets is

$$CS = \left[\left(Q_A - \frac{Q_A^2}{2}\right) - p_A Q_A\right] + \left[\frac{1}{b}\left(Q_B - \frac{Q_B^2}{2S}\right) - p_B Q_B\right] \tag{36}$$

where the first term is consumer surplus in A and the second term is consumer suplus in B. We insert the demand functions in markets A and B to get consumer surplus as a function of the retail price vector $p = (p_A, p_B)$:

$$CS(p) = \frac{1}{2}(1 - p_A)^2 + \frac{S}{2b}(1 - bp_B)^2 \tag{37}$$

which is continuous, concave and decreasing in both p_A and p_B.

Consider the case with the unconstrained wholesale prices, which may be either positive or negative. We insert accommodation retail prices to get consumer surplus in the accommodation equilibrium

$$CS^A = \frac{1}{8}(1 + t)^2 + \frac{1}{8Sb}(S - 2t)^2 \tag{38}$$

This may be differentiated with respect to the unit trade cost to study how joint consumer suplus depends on trade costs

$$\frac{dCS^A}{dt} = \frac{1}{4} + \frac{1}{4}t - \frac{1}{2b} + \frac{t}{Sb} \tag{39}$$

which is strictly positive for $b > 2$.

Correspondingly we insert arbitrage-free retail prices to get joint consumer surplus in the arbitrage-free equilibrium

$$CS^D = \frac{1}{2}\left(1 - \frac{2+Sbt}{4+Sb}\right) + \frac{S}{2b}\left(1 - \frac{(2+S)\,b + 4\,(1-bt)}{2\,(4+Sb)}\right)^2 \tag{40}$$

and differentiate with respect to the unit trade cost to determine how the consumer suplus depend on trade costs

$$\frac{dCS^D}{dt} = \frac{S\,(1-b\,(1-t))}{(4+Sb)} \tag{41}$$

This expression is strictly negative for $b > 2$. In other words, consumers benefit from partial market integration (at intermediate trade costs) as long as the difference between the two markets in terms of price elasticity is sufficiently large. The intuition for this result is that for intermediate trade costs the manufacturer is forced to set wholesale prices that result in a relatively small retail price-gap between the two markets, which increases total consumer welfare.

In the case of non-negative wholesale prices, the effect of parallel imports on consumer welfare is slightly more complicated. For a sufficiently high b such that

$$b > \frac{2\left(5S - 3 + \sqrt{9 + 14S + 25S^2}\right)}{11S}, \tag{42}$$

where the threshold level on the right hand side takes values on $\left[\frac{4}{3}, \frac{20}{11}\right]$, we have the following result

$$\frac{\partial CS}{\partial t} = \begin{cases} \frac{S(b-2)+(Sb+4)t}{4Sb} > 0 & if \quad t < \underline{t} \\ \frac{-Sb(2-t)-2(S+1)+4t}{4+9Sb} < 0 & if \quad \underline{t} \le t < \overline{t} \\ \frac{1}{8}S\,(2-b\,(1-2t)) > 0 & if \quad t \ge \overline{t} \end{cases}. \tag{43}$$

More precisely, combined consumer surplus in markets A and B has its unique global minimum at $t = \overline{t}$. If $b > b^*$ (the threshold value of b in equation (42), it has its unique global maximum at $t = \underline{t}$. If the difference in elasticities between the markets is small, i.e. $b < b^*$, consumer surplus has its global maxima at $t = 0$ and $t = \frac{1}{2}$.

This result shows that consumers in both markets may jointly benefit from parallel imports for intermediate trade costs as long as the difference in the price elasticity of demand is sufficiently large between the two consumer groups. The inutition for this result is that the pro-competitive effect of PI in A would dominate the double-markup effect in B since the optimal price in the latter market would be sufficiently low to induce a significant price reduction in the former market.

POLICY ANALYSIS: UNIFORM PRICING

Next consider a situation in which supranational competition authorities forbid the manufacturing firm from charging differentiated wholesale prices. More precisely, assume that the manufacturing firm cannot charge wholesale prices that put the parallel importing firm at a competitive disadvantage. For instance, a firm with a dominant EU position could be forbidden under competition law to price-discriminate between distributors in different member countries. Competition rules might be written this way in the belief that permitting a manufacturer to set differentiated prices would provide scope for blocking PI and reducing competitive gains in the higher-price market. In this case the manufacturing firm would set a uniform wholesale price w to maximize the joint industry profit in the two countries.

In the accommodation equilibrium with a uniform price, the first order condition with respect to the wholesale price in market A, w_A, is

$$\frac{d\Pi^A(w)}{dw_A} = \left[\frac{2}{9} - \frac{8}{9}w - \frac{4}{9}t\right] + \left[\frac{1}{3}t\right] - \left[\frac{1}{2}Sbw\right] = 0, \tag{44}$$

where again the first term in square brackets is the pro-competitive effect of PI in A, the second term is the trade-cost effect, and the third term reflects the double-markup problem in B.

For a low trade cost the manufacturing firm has an incentive to set a positive uniform wholesale price to moderate the pro-competitive effect in market A. In addition, for slightly higher trade costs the manufacturing firm can be expected to raise the wholesale price to reduce the volume of costly arbitrage. However, the incentive to moderate the pro-competitive effect and reduce PI volume is offset by the incentive to minimize the double-markup problem caused by a positive wholesale price in B. The optimal accommodating wholesale price is a trade-off between these different interests.

For sufficiently high trade costs, the manufacturer would block parallel imports by setting a low common wholesale price. The manufacturing firm would establish the wholesale price to block PI and consequently solves:

$$\frac{1 - w - 2t}{3} = 0 \tag{45}$$

with the unique solution

$$w = 1 - 2t \tag{46}$$

The trade cost is prohibitive for $t > 1/2$ as in the previous case with price differentiation.

We are now ready to characterize the equilibrium with a uniform wholesale price. We define the following threshold level for t:

$$\widehat{t} = \frac{1}{2}\left(\frac{4 + 3Sb}{5 + 3Sb}\right) \tag{47}$$

and we have the following result:

Proposition 3 *The model has a unique subgame perfect Nash equilibrium with a uniform wholesale price, $w_A = w_B$. The equilibrium value of w is given by*

$$w = \begin{cases} w_A = w_B = \frac{2(2-t)}{16+9Sb} & if \quad t < \widehat{t} \\ w_A = w_B = 1 - 2t & if \quad t \geq \widehat{t} \end{cases}, \tag{48}$$

To prove this proposition we first consider trade costs $t < \widehat{t}$. In this range of trade costs the manufacturing firm would set a uniform wholesale price to maximize the joint profit in both markets. More precisely, the solution to the sufficient first order condition (44) is

$$w = 2\left(\frac{2 - t}{16 + 9Sb}\right) \tag{49}$$

It follows that at this wholesale price the PI quantity would be non-negative for $t < \widehat{t}$, which is consequently a necessary condition for the equilibrium. For high trade costs, $t > \widehat{t}$, the volume of PI would be negative at accommodation prices and the manufacturing firm would set a uniform wholesale price to eliminate the trade:

$$w = 1 - 2t. \tag{50}$$

Define the threshold level for t as the unit trade cost for which the accomodation prices results in zero arbitrage:

$$q_B^* = \left(\frac{1-w-2t}{3}\right) = 0 \tag{51}$$

We insert the accomodation wholesale price

$$w = \frac{2(2-t)}{16+9Sb} \tag{52}$$

and solve the equation to find the critical trade cost

$$t = \frac{1}{2}\left(\frac{3Sb+4}{3Sb+5}\right) \tag{53}$$

which concludes our proof.

It is worth noting that for any non-prohibitive trade cost the uniform wholesale price is strictly positive. The intuition for this result is that the manufacturing firm has multiple offsetting interests. For relatively high trade costs, it has an incentive to reduce the volume of PI by charging a wholesale price that makes it unprofitable for the distributor in market B to ship goods to market A. On the other hand, the manufacturing firm is interested in keeping the wholesale price as low as possible to reduce the double-markup effect in both markets. Unlike the case with price differentiation the manufacturing firm would be forced to charge a wholesale price that results in a double mark-up effect in both markets, rather than only in B. The higher the trade cost, the stronger is the incentive to rely on the natural barrier to restrict and reduce the volume of PI. Accordingly, the uniform wholesale price decreases as trade costs rise.

For relatively low trade costs, the manufacturer has two other offsetting interests. On the one hand, the firm desires to reduce the pro-competitive effect of PI in market A, which can be done by charging a higher uniform wholesale price. On the other hand, the manufacturing firm prefers to avoid the double-markup problem in market B, which can be achieved by charging a lower wholesale price. The optimal wholesale price is determined by this tradeoff. For low trade costs the pro-competitive effect is dominant and the manufacturing firm would charge a higher wholesale price. For somewhat higher trade costs the pro-competitive effect is secondary and the manufacturer would charge a lower wholesale price. Again, the wholesale price decreases as trade cost goes up.

It is straightforward to find the corresponding retail prices by inserting the equilibrium wholesale prices in the price functions:

$$p = \begin{cases} p_A = \frac{1}{3} + \frac{2}{3}\left(\frac{2(2-t)}{16+9Sb}\right) + \frac{t}{3} & \\ p_B = \frac{1}{2b} + \frac{1}{2}\left(\frac{2(2-t)}{16+9Sb}\right) & if \quad t < \widehat{t} \\ p_A = 1 - t & \\ p_B = \frac{1}{2b} + \frac{1}{2} - t & if \quad t \geq \widehat{t} \end{cases} . \tag{54}$$

Very high (that is, close to prohibitive) trade costs result in a double-markup problem in both markets. The manufacturing firm would charge a uniform wholesale price to block PI, which would result in higher retail prices than they would be in the segmented equilibrium. (This may be seen in Figure 4, where the thin lines replicate retail prices in the differentiated-price case (see Figure 3) and the thick lines depict retail prices in the uniform-price case.) This outcome is obviously Pareto-dominated by market segmentation. Consumers in both markets as well as the manufacturing firm would prefer the segmented equilibrium to an equilibrium in which trade costs are high and PI is blocked. More interestingly, the outcome is Pareto-dominated by an equilibrium in which the manufacturer can differentiate its price between distributors. The retail prices in markets A and market B are both strictly higher in the uniform-price case compared to the equilibrium with differentiated prices. The manufacturing firm is obviously worse off in the uniform-price case compared to the differentiated-price case. This follows directly from the fact that he has more degrees of freedom when maximizing profit with two instruments than with one.

For low and intermediate trade costs, however, the retail price in market B (A) decreases (increases) as trade cost goes up under uniform wholesale pricing (see Figure 4). As trade cost rises the pro-competitive impact of PI in A is diminished. However, the double-markup problem in B becomes weaker as the uniform wholesale price is reduced, bringing down with it the retail price. In this range, trade liberalization (or reduced transport costs) would result in a reduced retail price gap between the two markets. Thus, our model indicates that a competition policy rule stating that a manufacturing firm is not permitted to differentiate wholesale prices, in order to limit the scope for PI, has a market-integrating effect in the sense that the retail price difference between markets diminishes as trade costs decrease.

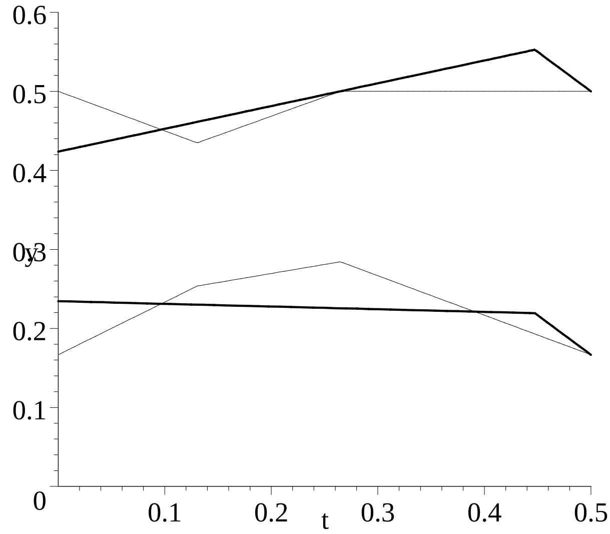

Figure 4: Retail Prices with a Uniform Wholesale Price.

It is straightforward to show in this case that consumer surplus in market A achieves a global maximum at complete trade liberalization (t=0) and a global minimum at $t = \widetilde{t}$, where retail price would reach its maximum. Correspondingly, consumer surplus in market B has a global minimum at complete trade liberalization and a global maximum at complete segmentation $t > 1/2$. From this analysis we conclude that a policy of forcing uniform pricing would be preferred to differentiated pricing by consumers at intermediate trade costs (around $\bar{t}$). Differentiated prices, on the other hand would be strictly preferred by all consumers for high trade costs $t > \widetilde{t}$. For very low trade costs consumers in market A would prefer uniform pricing while consumers in market B would enjoy differentiated pricing. An interesting implication is that a policy of uniform wholesale pricing with diminishing trade costs effectively would permit consumers in A to free-ride on the larger double-markup problem in B, which raises consumer price there.

The key result, however, is that if retail price convergence is the policy objective, a rule mandating uniform wholesale pricing would be appropriate in combination with efforts to reduce trade costs. However, this policy would not be optimal if trade costs are high and, even with declining transport costs it would set up a conflict of interests between consumers in the different markets.

Consider briefly an alternative policy that resonates in EU competition law. Suppose that distributors cannot charge prices to PI firms that put them at a disadvantage. Specifically, constrain the problem so that the price to PI firms, who are independent purchasers from the distributor in B, cannot be greater than the wholesale price in A, though the latter charge may be set by the manufacturer and may be differentiated from the wholesale price in B. In this case the earlier analysis pertains and as trade costs decline in the low range, the firm would raise the wholesale price in A in order to limit the force of competition from PI firms. In consequence, this policy would generate price divergence as transportation costs go down, as shown by the thin lines in Figure 4, and would harm consumers in the recipient market.

CONCLUDING REMARKS

We developed a model in which a manufacturing firm owns an intellectual property right in two markets but its ability to limit parallel imports from one market to the other is exhausted. In this environment, the firm has the ability to set differential wholesale prices to its independent distributors in the two locations. It will use these instruments to maximize profits within the vertical-control framework. There are three essential tradeoffs for the manufacturer. It wishes to restrict the extent of competition from PI in the A market, limit the amount of PI because it wastes real resources in transport costs, and avoid the double-markup problem in market B arising from the inability to set an efficient (zero) wholesale price.

Our analysis turned up some interesting and counterintuitive results. Because of the cross-cutting effects of PI on wholesale prices in the two markets, it is possible to observe a divergence in wholesale prices as trade costs are reduced within a low range. As a result, retail prices may move apart as well. Arguably, the EU is in a situation of low and declining internal trade restrictions. In this context, paradoxically, the policy of free parallel trade among member states may be a force for retail price divergence. For intermediate and high trade costs, the firm might wish to set a negative wholesale price in A but could be constrained to a minimum price of zero. As trade costs rise within those ranges, the volume of PI declines and the double-markup problem in B diminishes. Retail prices rise in both locations as transport costs increase within the intermediate range, then diverge at high costs as these prices achieve levels expected with wholesale-market segmentation.

A central implication is that a policy requiring uniform pricing among a manufacturing firm's distributors may be effective in achieving integrated retail prices in an environment of trade liberalization or reduction in transportation costs. However, this policy would not be optimal if trade costs are high and, even with declining transport costs it would set up a conflict of interests between consumers in the different markets. Finally, a policy of uniform wholesale pricing with diminishing trade costs effectively would permit consumers in A to free-ride on the larger double-markup problem in B, which raises consumer price there. As is typical in the analysis of arbitrage, consumers in low-priced economies would be harmed by uniform pricing as market become more integrated.

REFERENCES

Ganslandt, M. and K. E. Maskus (2003), "Vertical Restraints, Distribution, and the Price Impact of Parallel Imports: Implications for the European Union and Sweden," report to Swedish Competition Authority.

Hilke, J. C. (1988), "Free Trading or Free Riding: an Examination of the Theories and Available Evidence on Gray Market Imports," *World Competition* 32: 75-92.

Malueg, D. A. and M. Schwartz (1994), "Parallel Imports, Demand Dispersion, and International Price Discrimination," *Journal of International Economics* 37: 187-196.

Maskus, K. E. (2000), "Parallel Imports," *The World Economy* 23: 1269-1284.

Maskus, K. E. and Y. Chen (2002), "Parallel Imports in a Model of Vertical Distribution: Theory, Evidence, and Policy," *Pacific Economic Review*7: 319-334.

National Economic Research Associates (1999), *The Economic Consequences of the Choice of Regime in the Area of Trademarks* (London: National Economic Research Associates).

Richardson, M. (2002), "An Elementary Proposition Concerning Parallel Imports," *Journal of International Economics* 56: 233-245.

Tarr, D. G. (1985), "An Economic Analysis of Gray Market Imports," U.S. Federal Trade Commission, Manuscript.

APPENDIX

Distributor A's profit is equal to:

$$\pi_A^* = q_A^* \left[1 - (q_A^* + q_B^*) - w_A\right], \tag{55}$$

while the corresponding profit in market A for distributor B is

$$\pi_B^* = q_A^* \left[1 - (q_A^* + q_B^*) - w_B - t\right]. \tag{56}$$

The first-order conditions, which are also sufficient here, are:

$$1 - 2q_A^* - q_B^* - w_A = 0, \tag{57}$$

$$1 - q_A^* - 2q_B^* - w_B - t = 0, \tag{58}$$

provided $w_B + t \leq 1/2$.

Therefore, given any (w_i, T_i) that is accepted by L_i for $i = A, B$, there exists a unique Nash equilibrium in A, $\left(q_A^*(w_A, w_B), q_B^*(w_A, w_B)\right)$, given by

$$\begin{array}{c} q_A^*(w) = \frac{1-2w_A+w_B+t}{3}; \\ q_B^*(w) = \frac{1+w_A-2w_B-2t}{3} \end{array}, \text{ if } w_B \leq \frac{1}{2} - t \tag{59}$$

$$\begin{array}{c} q_A^*(w) = \frac{1-w_A}{2} \\ q_B^*(w) = 0 \end{array}, \text{ if } \frac{1}{2} - t < w_B \tag{60}$$

The equilibrium retail price in Country A, as a function of w_B, is

$$p_A(w) = \begin{cases} \frac{1+w_A+w_B+t}{3} & if \quad w_B \leq \frac{1}{2} - t \\ \frac{1+w_A}{2} & if \quad \frac{1}{2} - t < w_B. \end{cases} \tag{61}$$

When $\frac{1}{2} - t \geq w_B$,

$$\pi_A^*(w) = \frac{(1 - 2w_A + w_B + t)^2}{9}, \tag{62}$$

$$\pi_B^*(w) = \frac{(1 + w_A - 2(w_B + t))^2}{9}. \tag{63}$$

When $\frac{1}{2} - t < w_B$, we have

$$\pi_A^*(w) = \frac{(1 - w_A)^2}{4}, \ \pi_B^*(w) = 0 \tag{64}$$

We next consider output and price in Country B, again taking as given any (w_B, T_B) that is accepted by L_B. Distributor L_B solves

$$\max_p \{S(1 - bp_B)(p_B - w_B)\}. \tag{65}$$

The equilibrium (optimal) retail price and quantity in B thus are:

$$p_B(w) = \frac{1 + bw_B}{2b}; \ q_B(w) = S\left(\frac{1 - bw_B}{2}\right). \tag{66}$$

Firm L_B's operating profit in B, excluding T, is

$$\pi_B(w) = \frac{S}{b}\left(\frac{1 - bw_B}{2}\right)^2. \tag{67}$$