Why Parallel Trade May Raise Producers' Profits[1]

Horst Raff

Institut für Volkswirtschaftslehre

Christian-Albrechts-Universität zu Kiel

D-24098 Kiel, Germany

Email: raff@econ-theory.uni-kiel.de

Nicolas Schmitt

Département d'économie politique

Université de Genève

CH- 1211 Genève 4, Switzerland

Email: schmitt@ecopo.unige.ch

February 2005

Abstract

This paper shows that a manufacturer may benefit from parallel trade. In addition to an intuitive condition about the effect of demand shocks, this occurs when competitive retailers must order inventories before they know the realization of demand and for products whose sale value drops at the end of the demand period. For these types of products, letting retailers trade unsold inventories generally results in larger orders placed with the manufacturer, higher manufacturer profit and higher consumer surplus. The model provides a simple explanation as to why the volume of parallel trade is now very large and accepted by manufacturers for some products such as automobiles, clothes, toys, consumer electronics, musical recordings, cosmetics and perfumes.

JEL classification: F12

Keywords: parallel trade, distribution

1   Introduction

The purpose of this paper is to show that circumstances exist under which it pays a manufacturer to allow distributors (hereafter called retailers) to engage in parallel trade, that is trade that is not directly controlled by the trademark, the copyright or the patent owner, generally the manufacturer himself. Moreover parallel trade in these circumstances is generally welfare improving.

These results arise when four conditions are met. The first two are about the nature of the product. Retailers must place orders before they know the state of demand and the products have little value at the end of the demand period (or equivalently, it is costly to maintain them as inventories). The other two conditions are about the demand. The states of demand must be different across markets and different states of demand affect the quantity demanded rather than consumer's willingness to pay for the products.

There are several markets for which the conditions about the nature of the market apply. Many markets (for instance, automobiles, motorcycles, clothes) require manufacturers to introduce new models frequently. Other markets, such as toys, have very narrow and well-defined demand periods (e.g., pre-Christmas season). In all these cases, the value of the products decreases significantly at the end of the demand period. Moreover, many goods are not produced on a just-in-time inventory basis and exploiting scale economies often requires significant lags between production and sales. In other words, orders must be placed well before the relevant demand period. This often leads to errors in forecasting the future strength of demand and thus ex post to incentives to ship unsold inventories to markets where demand turns out to be higher than expected.

To understand why a manufacturer may have an incentive to allow this type of parallel trade, suppose it does not and is able to effectively ban parallel trade. Retailers may then be stuck with significant unsold inventories in which case they have a strong incentive to lower prices. Competition among retailers might even force the price to drop to zero as inventories are essentially sunk investments. Anticipating a possible loss, retailers will be reluctant to order large inventories. Banning parallel trade may thus have detrimental effects on the manufacturer's profitability. Allowing parallel trade, on the other hand, constitutes a simple mechanism by which the retail price does not fall dramatically when the state of the demand turns out to be low and it provides incentives to retailers to place larger orders than they

otherwise would.

Is there evidence that manufacturers like parallel trade? Data on parallel trade are notoriously hard to come by, simply because trade statistics do not distinguish between authorized and unauthorized intermediaries (Maskus, 2000). However, the volume of parallel trade has simply become too large to argue seriously that manufacturers either do not know of its existence or know about it but are unable to stop it even if it lowers their profitability. In other words, manufacturers must benefit from parallel trade at least in some markets.[1] This view is held by many experts. For instance, Lipner (1990, p.4) writes "[..] some manufacturers, while publicly opposed to gray market sales of their products, privately do little to inhibit their flow and in some instances even go so far as to encourage these transactions." Similarly, a report prepared for the EU Commission (NERA, 1999, p.11) states "[s]ome parallel trade, however, seems to be beneficial to the trademark owner. [..] If, for example, there is for some reason over-production in the source country, and the manufacturer would otherwise be left with an unsold stock, parallel trade may be a means to raise profits through additional sales. Another example is goods such as clothing which are subject to fashion waves. Previous season's clothing in one country can still yield useful revenues in other countries". Furthermore, a "dealer might [..] have over ordered, or might have excess quantities of an older or out-of-date version of the goods" (Lipner, 1990, p7).[2]

Consider the following specific examples. In the North American automobile market, retailers selling the products of a manufacturer must sign a contract forbidding them to sell in other countries cars intended for sale in one country. Yet over 200,000 new vehicles intended for the Canadian market were resold south of the border in 2001 (up from 16,000 in 1996; see

---

[1] In the US, parallel imports were estimated to be worth $7-10 billions in the mid-1980s (Cespedes, Corey and Rangan, 1988); today, estimates of $20 billion can be found (Computer Reseller News, 2001). In Europe, the volume of parallel imports varies from 5% of sales in markets like appliances, motorcars and consumer electronics to nearly 15% for musical recordings, cosmetics and perfumes (NERA, 1999). In the UK, the number of parallel imported motorcycles has increased from about 15,000 units in 1997 to 25,000 units in 1998 representing 25% of all new machines (House of Commons, 1999).

[2] Similar points are made in Cespedes et al. (1988) about disk drives, in Computer Reseller News (2001) about IT products, and in House of Commons (1999) about motorcycles, clothing/footwear and cosmetics. The case of earth-moving equipments during the Asian crisis is interesting. The crisis redirected previously ordered equipment into the strong US market causing the gray market share of hydraulic excavator sales to rise from 4.7% in 1997 to 19.5% in 1998 (Business Week, 1998).

Automotive News, 2002). Obviously, this no-parallel-trade clause is not enforced despite the fact that retailers face heavy penalties.[3] Parallel imports arise in this market in part because of the quota system adopted by most North American manufacturers. Retailers receive a pre-determined volume allocation per model in order to force them to sell a minimum number of cars without restricting retail prices.[4] The consequence is often a mismatch between demand and supply creating strong incentives for parallel trade between Canada and the US especially when a particular model sells well in one country and not in the other.[5] The automobile producers could easily eliminate this unintended trade. The fact that they do not indicates that they are not particularly hurt by this trade. Other examples of parallel trade arise from the fact that particular models are not distributed in one country[6], or that manufacturers deliberately oversupply certain markets because they want to keep a presence there.[7]

The existing literature on parallel trade has mainly focused on the issue of price discrimination (Malueg and Schwartz, 1999; Richardson, 2002; Maskus and Chen, 2002). This literature generally finds that manufacturers want to avoid parallel trade, because the arbitrage implied by parallel trade interferes with their ability to segment markets and set different prices in different countries.[8] The welfare effects of allowing parallel trade in these models are

---

[3]Interestingly, this clause seems to have rarely been tested in court.

[4]This huge increase in parallel imports is also due to wholesale pricing policies adopted by the automobile producers in the two countries. Canadian dealers apparently benefit from lower prices than their US counterparts, because they are located in a generally thiner, weaker and more price elastic market (Automotive News, 2002).

[5]"When a customer wants a car that the dealer cannot supply, the dealer has two choices [...], they can turn the customer away or they can source the car from a Canadian exporter [...]. When a new luxury car hits North American dealer lots, it can generate year-long American waiting lists; [in Canada], they just sit on the lot" (Financial Post, 2001).

[6]In Malysia, vehicles are imported by unauthorized dealers from Japan because Lexus does not yet carry the full range of models in this country (Automotive News, 2004). See World IT Report (2003) for similar examples concerning cellular phones.

[7]For instance, Mercedes-Benz wanting to keep a presence in Barbados systematically ships too many cars there. These cars typically are re-exported to the UK (see House of Commons, 1999). According to this report, this implies that the manufacturer consents to parallel trade.

[8]Knox and Richardson (2002) show, however, that a foreign monopoly may benefit from parallel trade when a country chooses both its tariff level and whether to allow parallel trade. Ahmadi and Yang (2000), Cosac (2002) show that parallel imports may

generally ambiguous, simply because the elimination of price discrimination may or may not be welfare improving (Tirole, 1988). We see our explanation of the role of parallel trade not as a substitute to this story but as a complement. In our model, the manufacturer's incentive to allow parallel trade is particularly strong when the price elasticity of demand is similar across countries but there is uncertainty about the actual size of demand. This incentive, however, quickly disappears in the presence of significant (permanent or random) differences in the price elasticity of demand across markets. In this case, a manufacturer has every incentive to keep markets segmented in order to practice (third degree) price discrimination. The manufacturer hence typically faces a trade-off—at least for the products that satisfy our conditions—between using international price discrimination and allowing parallel trade to get retailers to order more inventory. This may explain why the North American automobile manufacturers do not always enforce the no-parallel-trade clause.[9]

Finally, our approach shows that a key to understanding parallel trade is to consider whether the incentives faced by manufacturers and retailers are aligned or not. When they are aligned, there is no need for manufacturers to impose (or enforce) no-parallel-trade contracts with their retailers. In this sense, the present paper brings the economic literature on parallel trade closer to the legal (including law and economics) literature where parallel trade (or gray markets) is mainly viewed as a contractual issue between manufacturers and intermediaries (see Lipner, 1990; Gallini and Hollis, 1999).

The rest of the paper is organized as follows. In Section 2 we illustrate the main points of the paper with the help of a simple example. Section 3 contains a general model and proofs of the main results. Section 4 concludes.

---

also benefit manufacturers when they are preceived as different products with respect to the authorized products.

[9]Although retail prices for automobiles are lower in Canada than in the US (especially in the luxury car segment), threatening retailers may be enough to limit parallel trade while secretly tolerating it. It is only during periods of very significant retail price differences between the US and Canada (whether due to a low Canadian dollar or to a demand boom in the US as during the 1999-2001 period) that manufacturers start enforcing the no-parallel-trade clause.

2 An Example

Consider a risk-neutral monopoly manufacturer selling to a continuum of risk-neutral retailers in two different countries, denoted A and B. In turn these competitive retailers sell to consumers. Retailers must deal with two key characteristics of the market: first, they must order and take possession of inventories before demand becomes known, and second, inventories left unsold at the end of the demand period have no value. Retailers face a constant unit cost of distribution that we normalize to zero. The manufacturer's production cost is also assumed to be zero, as is the trade cost; this implies that we do not need to specify in which country the manufacturer is located.

We denote the volume of realized sales in country i by q_i. The volume of realized sales may differ from the volume ordered from the manufacturer, denoted by x_i, if after the realization of demand some goods are reexported or inventory is left unsold. Demand in country A is deterministic and given by the simple inverse demand function $p_A = 1 - q_A$. The inverse demand function in B is $p_B = 1 - \frac{q_B}{s}$, where the random variable s takes the value of 1 if demand is high and $\theta < 1$ if demand is low. The low state of demand occurs with probability λ. We have picked these demand functions to illustrate two points. First, parallel imports in our model occur even if the manufacturer does not price-discriminate between markets. In fact, a monopolist selling directly to consumers would set the same price in the two markets. Second, parallel imports may go in both directions, even if one country has a larger market than the other and/or one country has deterministic demand.

The game we study has the following order of moves. The manufacturer first announces a wholesale price w_i for each market $i = A, B$ and whether retailers have the authorization to engage in parallel trade. The retailers then order inventory and, after observing the true realization of demand, decide where and how much to sell at the market-clearing price.

2.1 No Parallel Trade

We first derive the benchmark solution when the manufacturer does not authorize the competitive retailers to ship any inventory to the other country (no parallel trade). In this case, the two markets are completely segmented and equilibrium is the same as the solution to the flexible price game analyzed by Deneckere et al. (1997). It is straightforward to show that, in market A, $\hat{p}_A = \hat{w}_A = 1/2$, $\hat{q}_A = \hat{x}_A = 1/2$ and the manufacturer's profit is $\hat{\pi}_A^m = 1/4$.

Consider now market B. With competitive retailers, we can associate sales with the volume ordered from the manufacturer because retailers have every incentive to sell whatever volume they have ordered at whatever price supported by the market as the value of the product drops to zero at the end of the demand period. In particular, the retail price may effectively drop to zero when the demand turns out to be low. Hence there are two possibilities when the demand is low: the retail price may be positive or equal to zero. Of course, this will depend on the level of the wholesale price: if it is high enough, inventories will tend to be low leading to a positive retail price. To uncover the circumstance under which the retail price may fall to zero when the demand is low, consider each case separately.

The representative retailer's expected profit can be written as

$$E\pi_B^r = \lambda \max\left(0, 1 - \frac{x_B}{\theta}\right)x_B + (1-\lambda)(1-x_B)x_B - w_B x_B. \qquad (1)$$

If the retail price in B falls to zero in the low demand state ($\max\{0, 1-\frac{x_B}{\theta}\} = 0$), the representative retailer earns positive revenue only when the demand is high. Since in equilibrium the retailer's expected profit is necessarily equal to zero, $x_B = 1 - \frac{w_B}{1-\lambda}$. The manufacturer then chooses w_B that maximizes $\Pi_B^{npi} = w_B x_B$. The solution is $\widehat{w}_B = \frac{1-\lambda}{2}$, implying an order volume of $\widehat{x}_B = \frac{1}{2}(= \hat{x}_A)$, a retail price in the state of high demand of $\widehat{p}_B^h = \frac{1}{2}$, and a low-demand retail price of $\widehat{p}_B^l = 0$. In the high-demand state retailers sell their entire inventory, i.e., $\widehat{q}_B^h = \frac{1}{2}$. Realized sales in the low-demand state are equal to $\widehat{q}_B^l = \min(\theta, \frac{1}{2})$. Hence, the manufacturer sets a low wholesale price in order to induce retailers to order the level of inventory that would be optimal if demand turned out to be high.

If the retail price in B is positive irrespective of the state of the demand ($\max\{0, 1-\frac{x_B}{\theta}\} = 1 - \frac{x_B}{\theta}$), $x_B = \frac{\theta(1-w_B)}{\theta(1-\lambda)+\lambda}$ ensures that the representative retailer makes zero expected profit. The manufacturer then chooses $\tilde{w}_B = \frac{1}{2}$ to maximize his expected profit. This implies that $\tilde{x}_B = \tilde{q}_B = \frac{\theta}{2(\theta(1-\lambda)+\lambda)} < \hat{x}_A$, $\tilde{p}_B^h = \frac{2(\theta(1-\lambda)+\lambda)-\theta}{2(\theta(1-\lambda)+\lambda)}(\geq \hat{p}_A)$ and $\tilde{p}_B^l = \frac{2(\theta(1-\lambda)+\lambda)-1}{2(\theta(1-\lambda)+\lambda)}(\leq \hat{p}_A)$. Hence the manufacturer sets the same 'high' wholesale price he would if demand were certain to be high, but accepts that the inventory level is lower than in country A and therefore too low when demand turns out to be high.

When the manufacturer does not authorize parallel trade, his optimal strategy in B is straightforward. Comparing profits, he chooses a high wholesale price, $\widetilde{w}_B = \frac{1}{2}$, when he expects the demands in the two states to be relatively similar ($\frac{1-\lambda}{2-\lambda} \leq \theta \leq 1$) and he chooses a low wholesale price, $\widehat{w}_B = \frac{1-\lambda}{2}$,

otherwise ($\theta < \frac{1-\lambda}{2-\lambda}$). As a result, the manufacturer's total expected profit is

$$\Pi^{npt} = \begin{cases} \frac{1}{4}\left[1 + \frac{\theta}{\theta(1-\lambda)+\lambda}\right] & \text{if } \frac{1-\lambda}{2-\lambda} \le \theta \le 1 \\ \frac{1}{4}\left[2 - \lambda\right] & \text{if } \theta < \frac{1-\lambda}{2-\lambda} \end{cases} \tag{2}$$

2.2 Parallel Trade

We now consider the case where the manufacturer allows the retailers to engage in parallel trade. The direct effect of allowing parallel trade is that it provides retailers in B an opportunity to sell unsold inventories in A when demand is low. Of course, the possibility of having to compete with parallel imports from B means that retailers in A also face random (residual) demand and may end up re-exporting part of their inventory to B when demand there turns out to be higher than expected. Hence even with demand shocks in only one country, parallel trade may take place in either direction.

Below, we denote the volume of parallel trade originating in B by m^j, where $j = h, l$ denotes the state of demand in B. Since x_i represents the volume of orders, the volume of sales q_i in the presence of parallel trade is equal to $x_i + m^j$ ($i = A, B$; $j = h, l$) in the country receiving parallel imports and $x_i - m^j$ in the other country. It should be clear that when the demand in B is low ($j = l$) then parallel trade necessarily goes from B to A ($m^l > 0$). However, when the demand in B is high ($j = h$), parallel trade could potentially originate in B ($m^h > 0$) or in A ($m^h < 0$).

Given these assumptions, the expected retail profit in B when $m^h > 0$ is equal to

$$E\pi_B^r = \lambda\{(1 - \frac{x_B - m^l}{\theta})\left(x_B - m^l\right) + (1 - \left(x_A + m^l\right))m^l\}$$
$$+ (1 - \lambda)\{(1 - \left(x_B - m^h\right))\left(x_B - m^h\right) + (1 - (x_A + m^h))m^h\} - w_B x_B, \tag{3}$$

and when $m^h < 0$, the expected retail profit in B is

$$E\pi_B^r = \lambda\{(1 - \frac{x_B - m^l}{\theta})\left(x_B - m^l\right) + (1 - \left(x_A + m^l\right))m^l\}$$
$$+ (1 - \lambda)(1 - \left(x_B - m^h\right))x_B - w_B x_B. \tag{4}$$

We make the further assumption that the difference in demands that triggers parallel imports between the two countries is sufficiently small that the

7

volume of parallel imports equalizes retail prices. Hence

$$1 - \frac{x_B - m^l}{\theta} = 1 - \left(x_A + m^l\right) \quad \text{or} \quad 1 - \left(x_B - m^h\right) = 1 - \left(x_A + m^h\right).$$

It follows that the volume of parallel trade is

$$m^l = \frac{x_B - \theta x_A}{1 + \theta} \quad \text{or} \quad m^h = \frac{x_B - x_A}{2}. \tag{5}$$

Substituting (5) into (3) or (4) and setting the representative retailer's profit to zero (competition at the retail level), the relationship between wholesale price and order volume such that the representative retailer expects zero profit is

$$w_B = 1 - \frac{1 + \lambda + \theta(1 - \lambda)}{2(1 + \theta)}(x_A + x_B), \tag{6}$$

irrespective of the direction of parallel trade.

Using the same reasoning as above, it is easy to verify that the relationship in country A between w_A and $x_A + x_B$ is identical to (6). This means that the wholesale price in each country depends on the manufacturer's overall shipment only and that $w_A = w_B$. Of course, this is due to the assumption that parallel trade equalizes retail prices across countries. As a result, we can only determine the total volume of orders that maximizes the manufacturer's overall profit, $\Pi^{pt} = w_B(x_A + x_B)$, where w_B is given by (6). It is easy to show that this overall profit is maximized for $\bar{w}_B = \frac{1}{2}$ irrespective of the expected demand in B. This results in an overall shipment of $\bar{X} = \frac{1 + \theta}{(1 + \lambda + \theta(1 - \lambda))}$ and a manufacturer's overall profit of

$$\Pi^{pt} = \frac{1 + \theta}{2(1 + \lambda + \theta(1 - \lambda))}. \tag{7}$$

Hence, comparing (2) and (7), we obtain

Result 1: *The manufacturer's expected overall profit increases with parallel trade if demand in the low-demand state is not too low $\left(\theta > \frac{1 - \lambda}{3 - \lambda}\right)$ and decreases otherwise.*

Figure 1 illustrates the manufacturer's total equilibrium profit as a function of θ with and without trade trade. Except at $\theta = 1$, parallel trade raises the manufacturer's profit when the retail price in B remains positive even when demand is low (corresponding to $\theta > \frac{1 - \lambda}{2 - \lambda}$ in Figure 1). It is also the case when the retail price in B drops to zero in the low-demand state, at least

8

if $\theta > \frac{1-\lambda}{3-\lambda}$. The reason why parallel trade raises the manufacturer's profit can be seen most easily in the case where $\theta > \frac{1-\lambda}{2-\lambda}$. Since wholesale prices are the same with and without parallel trade, the increase in profit clearly comes from the fact that retailers overall order more inventory when parallel trade is authorized. The same effect is at work if $\frac{1-\lambda}{3-\lambda} < \theta < \frac{1-\lambda}{2-\lambda}$: parallel trade gives retailers an incentive to order more inventory even without the manufacturer having to reduce the wholesale price (as we saw he would do without parallel trade). Finally, if $\theta < \frac{1-\lambda}{3-\lambda}$ —meaning that demand in B in the low-demand state is extremely low—the effect of parallel trade on the overall level of inventory is exactly reversed. While still giving an incentive to retailers in B to order more, this positive effect is more than offset by the fact that parallel trade gives a strong disincentive to retailers in A who expect to be swamped with parallel imports when demand in B is low. Hence, the manufacturer wants to prohibit parallel trade, precisely because he would ship too little inventory overall.

2.3   Welfare

We now investigate the welfare effects of allowing parallel trade, starting with the effect on consumer surplus. Suppose first that parallel trade is not allowed and let CS_i^{npt} denote the expected consumer surplus in country i in this case. In country A, consumer surplus is $CS_A^{npt} = \frac{1}{8}$. The expected domestic consumer surplus in B is

$$CS_B = \lambda(1 - p_B^l)\frac{q_B^l}{2} + (1 - \lambda)(1 - p_B^h)\frac{q_B^h}{2}. \tag{8}$$

The specific level of expected consumer surplus will depend on whether the retail price drops to zero when the state of the demand is low. If it does, using the respective prices and quantities in (8) yields $CS_B^{npt} = \frac{4\lambda\theta + 1 - \lambda}{8}$. If the retail price does not drop to zero, we obtain $CS_B^{npt} = \frac{\theta}{8(\theta(1-\lambda)+\lambda)}$.

Suppose now that parallel trade is allowed, and let CS_i^{pt} denote expected consumer surplus in country i. Above, we only found that $x_A + x_B = \bar{X} = \frac{1+\theta}{(1+\lambda+\theta(1-\lambda))}$. Hence, we do not know the volume of orders from each country and the volume of parallel trade. To compute consumer surplus for each country, all we need, however, is individual country sales which can be found in the following way. Suppose demand in B is low. Since parallel trade equalizes retail prices, it must be true that $1 - \frac{q_B}{\theta} = 1 - q_A$ and thus $q_A = \frac{q_B}{\theta}$.

9

In addition, parallel trade makes sure that, across the two countries, total sales must be equal to total orders (i.e., $q_A + q_B = x_A + x_B$). Combining these two conditions, we get

$$q_B^l = (x_A + x_B)\,\frac{\theta}{1+\theta} \quad \text{and} \quad q_B^h = \frac{x_A + x_B}{2}$$

when demand in B is low, respectively high. Once we have computed retail prices in B, we can use (8) to obtain

$$CS_A^{pt} = \frac{4\lambda + (1-\lambda)(1+\theta)^2}{8(1+\lambda+\theta(1-\lambda))^2} \quad \text{and} \quad CS_B^{pt} = \frac{4\lambda\theta + (1-\lambda)(1+\theta)^2}{8(1+\lambda+\theta(1-\lambda))^2}.$$

A comparison of expected consumer surplus in the two countries with and without parallel imports yields the following result.

Result 2: *Parallel trade raises expected consumer surplus in country A. It raises expected consumer surplus in country B provided that the retail price is positive without parallel trade ($\theta \geq \frac{1-\lambda}{2-\lambda}$), and reduces it otherwise.*

When the demand is very low in the low-demand state, the retail price in B drops to zero, and this can only benefit consumers. Allowing parallel trade makes this possibility less likely and this effect is detrimental to consumers. However, parallel trade also induces retailers to place larger orders, which is good for consumers. It is this effect that explains why consumer surplus in B is higher with parallel trade when θ is high enough to guarantee positive retail prices in all states of demand. Since the retail price never drops to zero in A, this effect also explains why consumer surplus in A always increases with parallel trade. Clearly, allowing parallel trade is profitable for the manufacturer exactly for the same reason that consumers in both countries like it, namely because retailers order more inventory—at least for sufficiently high values of θ.

Finally we turn to world social welfare, which consists of the sum of consumer surplus and the manufacturer's profit (the retailers expect zero profit) and thus to $W = CS_A + CS_B + \Pi$. Figure 2 illustrates W with and without parallel trade. We find:

Result 3: *Parallel trade raises expected world welfare if θ is sufficiently high ($\theta \geq \frac{1-\lambda}{2-\lambda}$) and reduces it otherwise.*

The key insight is that the interests of consumers and the manufacturer will often be aligned when it comes to the kind of parallel trade examined here: when demand shocks are not too large ($\theta \geq \frac{1-\lambda}{2-\lambda}$), parallel trade raises

both the manufacturer's profit and consumer surplus in both countries. This simple result is in sharp contrast with the conclusions of the previous literature that views parallel trade as a price discrimination issue. Parallel trade in that literature typically produces a disagreement between consumers and producers on the one hand, and between consumers in different countries on the other (see Malueg and Schwartz, 1994).

3 Generalization

Consider now a more general specification of the demand and the distribution of the demand shock. Let demand in country A be given by the function $D^A(p)$ and demand in B by $D^B(p, \theta)$, with $D_p^i < 0$, where $i = A, B$ and $\theta \in [\theta^-, \theta^+]$ is the (non-negative) realization of a random variable with density $f(\theta)$ and cumulative distribution $F(\theta)$. We still assume that in the presence of parallel trade prices are strictly positive for all realizations of demand. Note, however, that we no longer restrict the market in B to be smaller than the one in A.

Writing the inverse demand functions as $p_A = p_A(q_A)$ and $p_B = p_B(q_B, \theta)$ and letting $m > 0$ denote the volume of parallel trade going from B to A, we can express the expected retail revenue in A, ER_A^r, as

$$
\begin{array}{ll}
\int_{\theta^-}^{\theta^+} p_A(x_A + m)x_A f(\theta)d\theta & \text{if } m \geq 0 \\
\int_{\theta^-}^{\theta^+} \{p_A(x_A + m)(x_A + m) + p_B(x_B - m, \theta)(-m)\} f(\theta)d\theta & \text{if } m < 0,
\end{array} \tag{9}
$$

and that in B, ER_B^r, as

$$
\begin{array}{ll}
\int_{\theta^-}^{\theta^+} \{p_B(x_B - m, \theta)(x_B - m) + p_A(x_A + m)m\} f(\theta)d\theta & \text{if } m \geq 0 \\
\int_{\theta^-}^{\theta^+} p_B(x_B - m, \theta)x_B f(\theta)d\theta & \text{if } m < 0.
\end{array} \tag{10}
$$

Since the expected retail profits, $E\pi_i^r = ER_i^r - w_i x_i$, $i = A, B$, must equal zero, we can write the manufacturer's problem of maximizing the expected value of $\Pi = w_A x_A + w_B x_B$ as

$$
\max_{x_A, x_B} \int_{\theta^-}^{\theta^+} \{p_A(x_A + m)(x_A + m) + p_B(x_B - m, \theta)(x_B - m)\} f(\theta)d\theta. \tag{11}
$$

The volume of parallel trade, $m = m(x_A, x_B, \theta)$, is the flow of goods that equalizes retail prices across the two countries ex post for each level of x_A,

11

x_B and θ. It is implicitly defined by

$$p_A(x_A + m) = p_B(x_B - m, \theta). \tag{12}$$

Suppose that the manufacturer's profit is maximized for $m = 0$, and let the corresponding profit-maximizing order volumes be denoted by x_A^* and x_B^*; also assume, for the time being, that retail prices are non-negative at these volumes for all realizations of θ. Then the derivative of (11) with respect to m must be zero at $m = 0$,

$$\int_{\theta^-}^{\theta^+} \left\{ p_A(x_A^*) + x_A^* p_A'(x_A^*) - p_B(x_B^*, \theta) - x_B^* p_B'(x_B^*, \theta) \right\} f(\theta)d\theta = 0. \tag{13}$$

That is, the manufacturer will set wholesale prices ex ante so that the volumes ordered by the retailers, x_A^* and x_B^*, equalize expected marginal revenues in A and B. In addition, expected marginal revenues must equal marginal cost, which is zero,

$$p_A(x_A^*) + x_A^* p_A'(x_A^*) = \int_{\theta^-}^{\theta^+} \left\{ p_B(x_B^*, \theta) + x_B^* p_B'(x_B^*, \theta) \right\} f(\theta)d\theta = 0. \tag{14}$$

Ex post, however, when the demand shock has been realized, marginal revenues may no longer be equalized for x_A^* and x_B^*. That is, we may have

$$MR_A^* \equiv (p_A(x_A^*) + x_A^* p_A'(x_A^*)) = 0 \gtrless p_B(x_B^*, \theta) + x_B^* p_B'(x_B^*, \theta) \equiv MR_B^*(\theta). \tag{15}$$

In this case, an ex-post reallocation of output between markets would raise the manufacturer's expected profit. This can best be seen if we suppose that the manufacturer could do the reallocation himself, that is if distribution and production were vertically integrated. For realizations of θ for which $MR_B^*(\theta) > 0$, the manufacturer would ship goods from A to B until marginal revenues are equalized, thereby raising his ex-post profit. When the realized value of θ is such that $MR_B^*(\theta) < 0$, the manufacturer would not reallocate goods but sell only enough units to make marginal revenue in B equal to zero, leaving some inventory unsold in the process. The ability to reallocate goods ex post would thus increase the manufacturer's profit for high values of θ and leave profit unchanged for low values. It follows that from an ex-ante perspective the manufacturer's expected profit would be higher if he had the option of reallocating goods ex post.

Of course, in the scenario we consider the goods have been sold to independent retailers. Hence the question is whether simply allowing them to engage in parallel trade can bring about an ex-post reallocation of goods that would raise the manufacturer's expected profit. The retailers will carry out parallel trade whenever there are differences in retail prices across countries. Retailers that can sell their goods in another country where prices are higher gain, those that see retail prices fall due to parallel imports lose. Will the aggregate retail profit in the two countries rise for a given realization of θ when parallel trade eliminates retail price differences? The answer is yes, if for the given θ the equalization of retail prices also moves marginal revenues closer to each other. Otherwise, aggregate retail profit for this value of θ falls.

To learn whether ex ante expected aggregate retail profit rises or falls when parallel trade is allowed, one has to weight ex-post retail profits with the density and integrate over θ. If the expected aggregate retail profit in A and B rises, then this implies that for given wholesale prices retailers have to order more inventory in order to keep the zero-profit constraint implied by perfect competition satisfied. This in turn means that the manufacturer's expected profit, expected consumer surplus (in the two countries taken together) and therefore also expected world welfare increase. By contrast, if the expected aggregate retail profit falls when parallel trade is allowed, then the manufacturer's expected profit, aggregate consumer surplus and world welfare decrease.

Therefore the crucial question is whether there exist sufficient conditions under which parallel trade raises expected aggregate retail profits and under which it lowers them. We prove the following result:

Proposition 1 *(a) If $D^A(p) = D(p)$ and $D^B(p, \theta) = \theta D(p)$, and $F(\theta)$ puts sufficiently little probability mass on realizations of θ for which retail price falls to zero without parallel trade, then allowing parallel trade raises the manufacturer's expected profit, expected world consumer surplus and expected world welfare. (b) If the inverse demand functions take the form $p_A = p(q)$ and $p_B = \theta p(q)$, then permitting parallel trade reduces the manufacturer's expected profit, expected world consumer surplus and expected world welfare.*

Proof: (a) In this case, it is straightforward to show that the price elasticity of demand, $\varepsilon(p)$, does not depend on θ. Hence marginal revenues in A and

B are equalized ex post if

$$p_A\left(1 - \frac{1}{\varepsilon(p_A)}\right) = p_B\left(1 - \frac{1}{\varepsilon(p_B)}\right), \tag{16}$$

i.e. whenever $p_A = p_B$. Hence parallel trade, by equalizing retail prices for every value of θ, must raise expected aggregate retail profits. This is true under the implicit assumption that in the absence of parallel trade retail prices are non-negative for all realizations of demand.

Suppose now that without parallel trade the choice of x_A^* and x_B^* causes the retail price in B to drop to zero for low realizations of demand. In particular, let $\tilde{\theta} \equiv \tilde{\theta}(x_B^*)$ be defined by $p_B(\frac{x_B^*}{\tilde{\theta}}) = 0$, so that price is zero for $\theta \in [\theta^-, \tilde{\theta}]$.[10] We then have to modify (14) as follows:

$$p_A(x_A^*) + x_A^* p_A'(x_A^*) = \int_{\tilde{\theta}}^{\theta^+} \left\{p_B(\frac{x_B^*}{\theta}) + \frac{x_B^*}{\theta} p_B'(\frac{x_B^*}{\theta})\right\} f(\theta)d\theta = 0. \tag{17}$$

For any $\theta \in [\theta^-, \tilde{\theta}]$, goods will flow from B to A if we allow parallel trade. This will have no effect at first on retailer profits in B because the price there is zero. The retailer profits in A, however, must fall as successive units are shipped there, since x_A^* was chosen optimally and additional units cause marginal revenue to become negative. Thus parallel trade may lower ex-post aggregate retail profit for low values of θ. For $\theta \in (\tilde{\theta}, \theta^+]$ parallel trade will raise overall ex-post retail profit for the reasons given in the preceding paragraph. Hence the overall effect of parallel trade will depend on the distribution of θ. In particular, if there is sufficiently little probability mass on $[\theta^-, \tilde{\theta}]$, the expected aggregate retail profit will increase with parallel trade.

(b) In this case one can easily show that an equalization of marginal revenues requires equal shipments to each market independent of θ,

$$p(x_A)\left(1 - \frac{1}{\varepsilon(x_A)}\right) = p(x_B)\left(1 - \frac{1}{\varepsilon(x_B)}\right), \tag{18}$$

and hence almost always results in different retail prices. Allowing retailers to engage in parallel trade necessarily pushes marginal revenues apart in the two countries for every value of θ and hence reduces expected aggregate retail profit.∎

---

[10]Note that we let $p_B = p_B(\frac{q_B}{\theta})$ be the inverse of $D_B(p, \theta) = \theta D(p)$.

The proposition shows that the form of demand uncertainty plays a crucial role in determining whether parallel trade is desirable or not. One way to interpret the demand shock in (a) is to assume that consumers in both countries have identical individual demand functions, but that the manufacturer does not know how many consumers there will be for his product. Hence, at any given price, the price elasticity of demand is the same across the two countries, but the manufacturer does not know how much he will sell at that price. Both manufacturers and retailers have the same incentives with respect to parallel trade: let retailers ship goods ex post to the market where demand is highest so that it pays them to order more inventory. Case (b) can be viewed as a situation in which consumers in the two countries differ in their willingness to pay, but the manufacturer is uncertain about how much they differ. This situation requires that the retail price of the good adjust ex post to the realized willingness to pay. Without parallel trade, perfect competition between retailers ensures that this adjustment takes place. Parallel trade, however, interferes with this price adjustment because a rise in the retail price tends to draw parallel imports into the country. In anticipation of this, retailers will order less inventory, thereby reducing the manufacturer's profit, as well as consumer surplus and welfare.

4   Conclusions

This paper demonstrates that there are circumstances under which it pays a manufacturer to allow retailers to engage in parallel trade. Specifically, parallel trade gives retailers an incentive to place larger orders than they otherwise would. Moreover, parallel trade in these circumstances is generally welfare improving. These results arise when three conditions are met: first, retailers must place orders before they know the state of the demand; second, the states of the demand are different across markets; and third the products have little value at the end of the demand period (or equivalently, it is costly to maintain them as inventories).

The type of demand shock also matters. In particular, we should expect to see parallel imports encouraged by manufacturers in those industries in which consumers' willingness to pay is relatively similar across markets but for which there is uncertainty about how many consumers will actually choose to buy. In short, these would be products that are relatively standard but for one reason or another the manufacturers have a difficult time anticipating the

volume of sales. The toy market seems like a good example of such market, but so are certain segments of the fashion, motorcycle or even automobile market, as well as the electronics market.

It is important to note that there are other possible mechanisms that would allow a manufacturer to reduce destructive competition. Three come to mind: vertical integration with retailers, resale price maintenance (see Deneckere et al., 1997), or a return policy for unsold inventories. The point of the paper is that parallel trade constitutes a particularly simple mechanism to achieve this goal. We would expect this to be true, especially if the manufacturer has less information than retailers about local market conditions, as seems especially likely if the manufacturer is located overseas. Resale price maintenance, for instance, would involve considerably more checking and monitoring of retailers than allowing parallel imports. Similarly a manufacturer's return policy is costly, not only because it might allow well informed local retailers to shirk on sales effort, but also because a foreign manufacturer may have no particular physical facilities in its export markets to handle returned merchandise. Finally, vertical integration may be difficult for a foreign manufacturer again due to informational asymmetries, but also because the market volume may be too small to justify the investment.

References

[1] Automotive News (2004), 'No Matter What Color that Lexus is Asia Quite Often Is Gray', 78, Issue 6078, February 2.

[2] Automotive News (2002), 'Gray Market Yields Gold', 76, Issue 5972, February 25.

[3] Business Week (1998), 'The Earth is Shifting under Heavy Equipment', Issue 3572, April 6.

[4] Computer Reseller News (2001), 'Vendors Aim to Squash Gray Market', Issue 963, September 17.

[5] Cespedes, F.; E.R. Corey abd V.K. Rangan (1988), 'Gray Markets: Causes and Cures', *Harvard Business Review,* July-August, 75-82.

[6] Cosac, Teodora (2002), 'Vertical Restraints and Parallel Imports with Differentiated Products', Unpublished PhD Thesis, Simon Fraser University.

[7] Deneckere, Raymond, Howard P. Marvel and James Peck (1997), 'Demand Uncertainty and Price Maintenance: Markdowns as Destructive Competition' *American Economic Review 87, 4, 619-41.*

[8] Financial Post (2001), 'Wanted: Canadian Cars', November 3, 2001

[9] Gallini, Nancy and Aiden Hollis (1999), 'A Contractual Approach to the Gray Market', *International Review of Law and Economics,* 19(1), 1-21.

[10] Ganslandt, Mattias and Keith Maskus (2004), 'Parallel Imports and Pricing of Pharmaceuticla Products: Evidence from the European Union', *Journal of Health Economics*, forthcoming.

[11] House of Commons (1999), *Trademarks, Fakes and Consumers,* Eight Report, Trade and Industry Committee. http://www.parliament.the -stationery-office.co.uk/ pa/cm199899/cmselect/ cmtrdind/380/38007.htm

[12] Knox, Daniel and Martin Richardson (2002), 'Trade Policy and Parallel Imports', *European Journal of Political Economy* 19, 133-51.

[13] Lipner, Seth. 1990. *The Legal and Economic Aspects of Gray Market Goods.* London: Quorum Books.

[14] Maskus, Keith and Yongmin Chen (2002), 'Parallel Imports in a Model of Vertical Distribution: Theory, Evidence and Policy', *Pacific Economic Review.*

[15] Maskus, Keith. 2000. Parallel Imports. *World Economy* 23, 9. 1269-84.

[16] Malueg, D.A. and Marius Schwartz (1994), 'Parallel Imports, Demand Dispersion and International Price Discrimination', *Journal of International Economics* 37, 167-96.

[17] NERA (1999), 'The Economic Consequences of the Choice of a Regime of Exhaustion in the Area of Trademarks', Final report for DGXV of the European Commission, February.

[18] Richardson, Martin (2002), 'An Elementary Proposition Concerning Parallel Imports', *Journal of International Economics* 56, 1, 233-45.

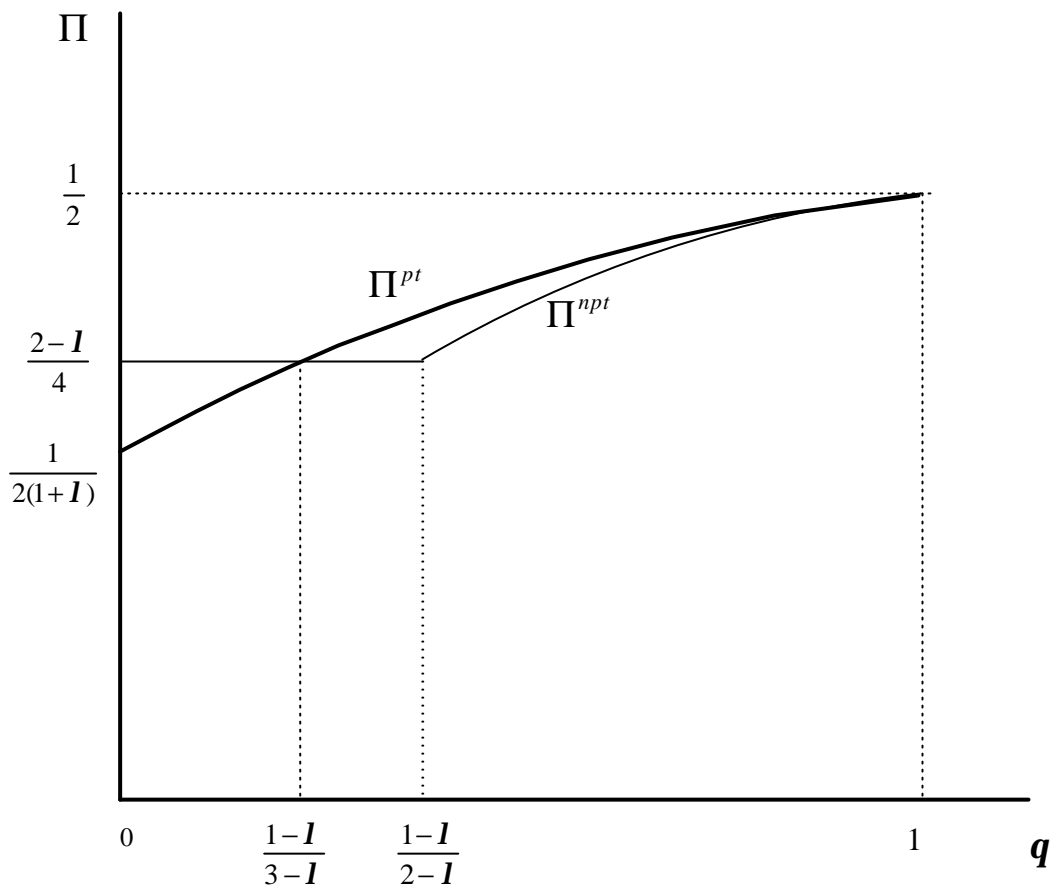[19] World IT Report (2003), '30 Percent of Malaysian Cells Purchased Through Grey Market', December 28.
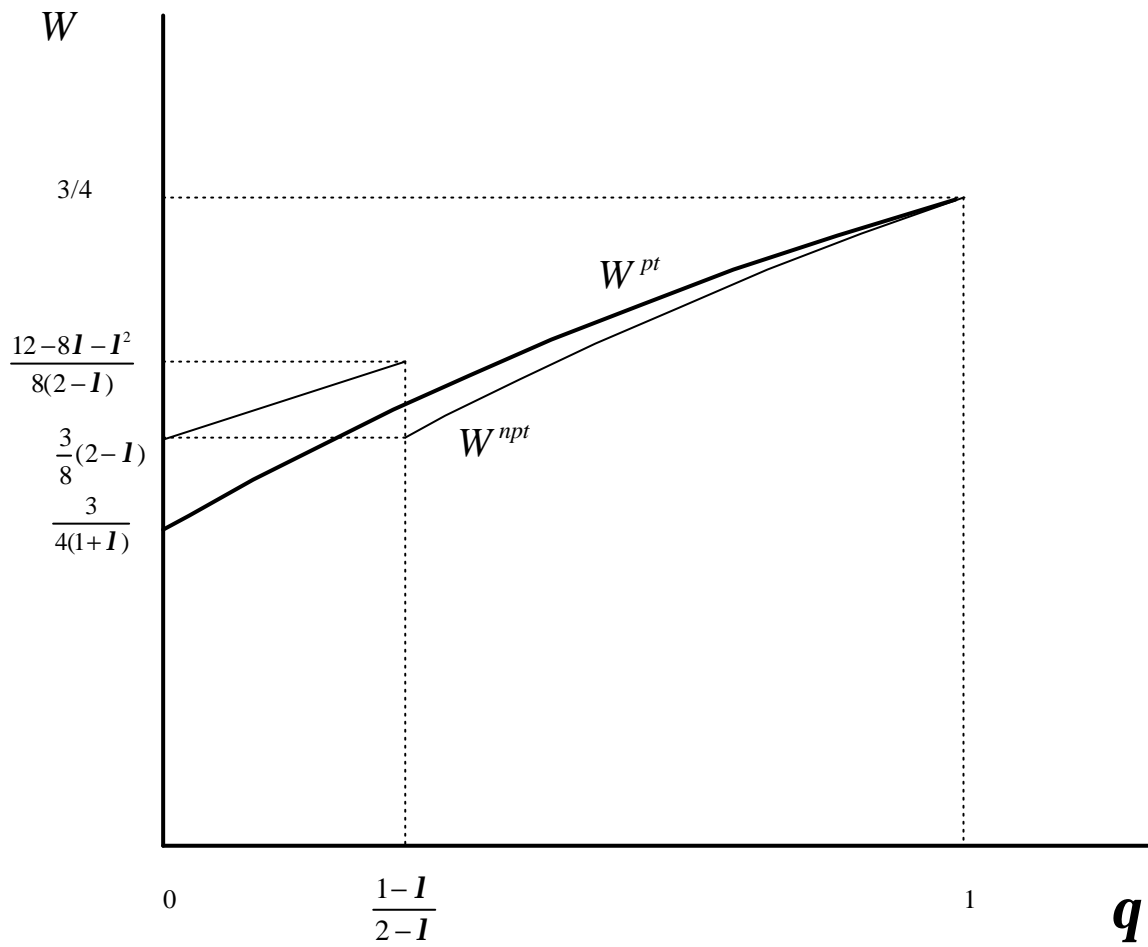
Figure 1: Manufacturer's total profit

Figure 2: World Social Welfare