Journal of International Economics 71 (2007) 434–447
www.elsevier.com/locate/econbase

Why parallel trade may raise producers' profits

Horst Raff [a,*], Nicolas Schmitt [b,c]

[a] *Institut für Volkswirtschaftslehre, Christian-Albrechts-Universität zu Kiel, 24098 Kiel, Germany*
[b] *Department of Economics, 8888 University Drive, Simon Fraser University, Burnaby, Canada V5A 1S6*
[c] *Département d'économie politique, Université de Genève, CH-1211 Genève 4, Switzerland*

Received 5 October 2005; received in revised form 6 July 2006; accepted 10 August 2006

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 and higher manufacturer profit. 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.

Keywords: Parallel trade; Distribution

JEL classification: F12

1. Introduction

The purpose of this paper is to demonstrate 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, copyright or patent owner, generally the manufacturer himself. Moreover, we show that under these circumstances parallel trade—sometimes also referred to as gray markets—is typically welfare improving.

* Corresponding author. Tel.: +49 431 880 1582; fax: +49 431 880 3150.
E-mail address: raff@econ-theory.uni-kiel.de (H. Raff).

0022-1996/$ - see front matter
doi:10.1016/j.jinteco.2006.08.001

These results arise when four conditions are met. The first two are about the nature of the product: retailers have to 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 have to be different across markets with positive probability, and different states of demand have to affect the quantity demanded rather than consumers' willingness to pay for the products.

There are several markets for which these conditions are met. In the markets for automobiles, motorcycles and clothes, manufacturers have to introduce new models frequently. In the toy market, products 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 basis and exploiting scale economies often requires significant lags between production and sale. In other words, orders have to 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 he 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 down to zero, since inventories are essentially sunk investments. But if they anticipate a possible loss, retailers will be reluctant to order large inventories. Banning parallel trade may thus be detrimental to the manufacturer's profitability. Allowing parallel trade, on the other hand, may provide a simple mechanism to keep retail prices from falling dramatically when the state of the demand turns out to be low and thus give retailers an incentive to place larger orders than they otherwise would. This is not a trivial issue, since parallel trade provides additional sales opportunities for those retailers that can export but, at the same time, lowers those of retailers that face competition from imports. We hence must show that parallel trade leads to larger total orders. The four conditions mentioned above make sure that this is indeed the case and that parallel trade raises the manufacturer's profit over a wide range of parameters. Importantly, our reasoning does not rely on an insurance argument—we assume throughout that retailers are *risk-neutral*.

There is considerable anecdotal evidence that manufacturers benefit from parallel trade at least in some markets.[1] This view is held by several 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

[1] Data on parallel trade are notoriously hard to come by, simply because trade statistics do not distinguish between authorized and unauthorized intermediaries (Maskus, 2000). In the US, parallel imports were estimated to be worth $7–10 billions in the mid-1980s (Cespedes et al., 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) and to even 25% in the UK motorcycle market (House of Commons, 1999).

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, p. 7).[2]

The North American automobile market is a good case to keep in mind. Retailers selling new automobiles must sign a contract with manufacturers forbidding them to re-sell these cars in other countries. Yet over 200,000 vehicles intended for the Canadian market were resold south of the border in 2001 (up from 16,000 in 1996; see 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 suggests that they are not particularly hurt by this trade.[6]

The existing literature on parallel trade has mainly focused on the issue of price discrimination (Malueg and Schwartz, 1994; Richardson, 2002; Maskus and Chen, 2002). This literature generally finds that manufacturers want to avoid parallel trade, because it interferes with their ability to segment markets and set different prices in different countries.[7] The welfare effects of allowing parallel trade in these models are 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. However, if there are significant (permanent or random) differences in the price elasticity of demand across markets, 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 prohibiting parallel trade to practice international price discrimination and allowing parallel trade to get retailers to order more inventory. This may explain why the North

[2] See also Cespedes et al. (1988), Computer Reseller News (2001), and House of Commons (1999).

[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 thinner, 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] Other examples of parallel trade arise from the fact that particular models are not distributed in one country (see Automotive News, 2004a; World IT Report, 2003), or that manufacturers deliberately oversupply certain markets to keep a presence there. For instance, Mercedes–Benz wanting to keep a presence in Barbados systematically ships too many cars there that are typically re-exported to the UK. According to House of Commons (1999), this implies that the manufacturer consents to parallel trade.

[7] 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), and Cosac (2002) show that parallel imports may also benefit manufacturers when they are perceived to be different from the authorized products.

American automobile manufacturers insist on a no-parallel-trade clause in their contracts with dealers, but do not always enforce it.[8]

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 clauses in their contracts with 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 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 model. Section 3 uses a more general model to examine the alignment of the manufacturer's and retailers' incentives and to show when parallel trade raises the manufacturer's profit. Section 4 concludes.

2. A simple model

Consider a *risk-neutral* monopoly manufacturer selling to a continuum of *risk-neutral* retailers in two different countries, denoted A and B. The manufacturer sets wholesale prices w_A and w_B. Given these prices, competitive retailers in each country order inventory. Retailers must deal with two key characteristics of the market: first, they have to 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 re-exported or inventory is left unsold. Demand in country A is deterministic and given by the 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 trade in our model occurs even if the manufacturer does not price-discriminate between markets. In fact, a monopolist selling directly to consumers would set the same retail price in the two markets. Second, parallel trade may go in both directions, even if one country has a larger market than the other and/or one country has deterministic demand.[9]

[8] 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. Today the volume of parallel trade in this market has slowed down considerably (Automotive News, 2004b).

[9] Our model could also be set up as a sequential game, in which 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. Retailers then order inventory, each taking the quantity of the other retailers as given (Cournot competition). After observing the true realization of demand, retailers simultaneously announce retail prices. The subgame perfect equilibrium of this game converges to the results we obtain as the number of retailers goes to infinity. See Tirole (1988), ch. 5 and the references cited there for the relevant convergence results.

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 the equilibrium is the same as the solution to the flexible price game analyzed by Deneckere et al. (1997).

Consider first country B. Competitive retailers have every incentive to put whatever volume they have ordered on the market at whatever price the market will support, as the product has no value (or is prohibitively costly to store) once the demand period has ended. This means that the retail price may drop to zero when demand turns out to be low. How much retailers order is determined by the condition that expected retail profit has to equal zero:

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

Condition (1) gives us the total quantity ordered by the retailers, x_B, as a function of w_B and parameters λ and θ. Maximizing $\Pi_B^{npt} = w_B x_B(w_B, \lambda, \theta)$, the manufacturer sets w_B such that his expected marginal revenue equals marginal cost (i.e., zero). Ex-post, the quantity ordered by the retailers is "too small" if demand turns out to be high, and "too large" if demand turns out to be low. An additional consideration of the manufacturer is the discontinuity in Eq. (1) that arises because the retail price may drop to zero in the low-demand state. We show in the Appendix that for a given λ, the manufacturer maximizes profit by setting a low wholesale price $\hat{w}_B = \frac{1-\lambda}{2}$ if $\theta < \frac{1-\lambda}{2-\lambda}$, and a high wholesale price $\hat{w}_B = \frac{1}{2}$ if $\frac{1-\lambda}{2-\lambda} \leq \theta$. The reason for the low wholesale price in the first scenario is that the manufacturer wants retailers to order enough inventory in case demand turns out to be high; he accepts that the retail price drops to zero in the low-demand state.

In country A the solution is much simpler than in B, since there is no uncertainty. It is straightforward to show that $\hat{p}_A = \hat{w}_A = 1/2$, $\hat{q}_A = \hat{x}_A = 1/2$ and that the manufacturer's profit is $\hat{\Pi}_A^{npt} = 1/4$. As a result, the manufacturer's total expected profit is

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

2.2. Parallel trade

We now consider the case where the manufacturer allows the retailers to engage in parallel trade. The direct effect is to provide retailers in B with an opportunity to sell unsold inventories in A when demand is low. But this would mean more competition for retailers in A. Hence, the possibility of having to compete with parallel imports from B implies that retailers in A also face random (residual) demand and may end up re-exporting part of their inventory when demand in B 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. It should be clear that when demand in B is low ($j=l$), parallel trade goes from B to A ($m^l>0$), whereas when demand in B is high ($j=h$), parallel trade originates in A ($m^h<0$). Sales in A in the presence of parallel trade are equal to x_A+m^j, and sales in B are given by x_B-m^j. The expected retail profit in B is then equal to

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

Parallel trade equalizes retail prices across countries. Hence

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

It follows that the volume of parallel trade is

$$m^l = \frac{x_B-\theta x_A}{1+\theta} \text{ or } m^h = \frac{x_B-x_A}{2}. \tag{4}$$

Substituting Eq. (4) into Eq. (3) and setting expected retail profit to zero, we obtain

$$w_B = 1-\frac{1+\lambda+\theta(1-\lambda)}{2(1+\theta)}(x_A+x_B). \tag{5}$$

Since parallel trade equalizes retail prices in the two countries, it follows that the equivalent of Eq. (5) has to hold for A and that $w_A=w_B=w$. As a result, we can only determine the total volume of orders that maximizes the manufacturer's overall profit, $\Pi^{pt} = w(x_A+x_B)$.

It is easy to show that this overall profit is maximized for $\hat{w} = \frac{1}{2}$. This results in a total shipment of $\hat{X} \equiv \hat{x}_A + \hat{x}_B = \frac{1+\theta}{(1+\lambda+\theta(1-\lambda))}$ and a manufacturer's overall profit of

$$\hat{\Pi}^{pt} = \frac{1+\theta}{2(1+\lambda+\theta(1-\lambda))}. \tag{6}$$

Comparing Eq. (2) and Eq. (6), 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* ($\theta>\frac{1-\lambda}{3-\lambda}$) *and decreases otherwise.*

Fig. 1 illustrates the manufacturer's expected equilibrium profit as a function of θ with and without parallel trade. To see what is going on, suppose that $\theta\geq\frac{1-\lambda}{2-\lambda}$, so that parallel trade raises expected profit. In this case, the retail price does not drop to zero in the low-demand state, and wholesale prices are the same in A and B with and without parallel trade. Thus the increase in profit clearly comes from the fact that retailers overall order more inventory when parallel trade is authorized.[10] If $\theta<\frac{1-\lambda}{3-\lambda}$, demand in B in the low-demand state is very low and the effect of parallel trade on the overall level of inventory is reversed. While still giving an incentive to retailers in B to order more, retailers in A who expect to be swamped with parallel imports when demand in B is low order so much less that the total level of inventory falls with parallel trade. Hence, the manufacturer wants to prohibit parallel trade in this case.

[10] The same effect is at work if $\frac{1-\lambda}{3-\lambda}\leq\theta<\frac{1-\lambda}{2-\lambda}$: parallel trade gives retailers an incentive to order more inventory. Without parallel trade, the manufacturer would have to reduce the wholesale price to achieve this.

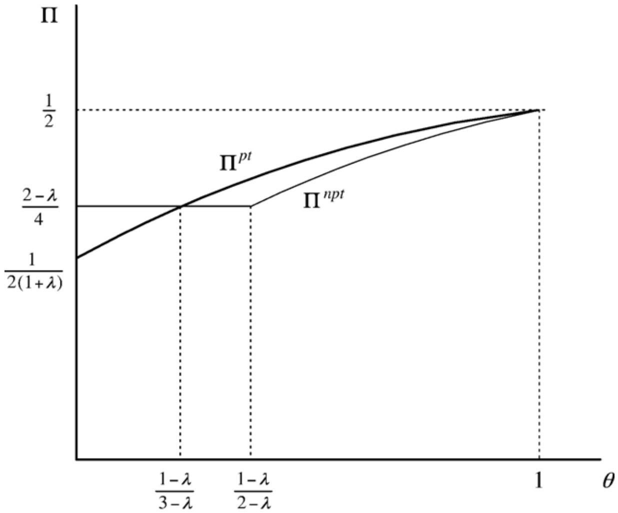

Fig. 1. Manufacturer's total profit.

2.3. Welfare

We now investigate the welfare effects of allowing parallel trade, starting with the effect on consumer surplus. When parallel trade is not allowed, consumer surplus in country A is $\mathrm{CS}_A^{npt} = \frac{1}{8}$ and the expected consumer surplus in B (see the Appendix for details) is

$$\mathrm{CS}_B^{npt} = \begin{cases} \dfrac{4\lambda\theta + 1 - \lambda}{8} & \text{if } \theta < \dfrac{1-\lambda}{2-\lambda} \\ \dfrac{\theta}{8(\theta(1-\lambda)+\lambda)} & \text{if } \dfrac{1-\lambda}{2-\lambda} \leq \theta \end{cases}$$

Notice that the discontinuity in w_B at $\theta = \dfrac{1-\lambda}{2-\lambda}$ causes a discontinuity in the retail price and hence in consumer surplus in B.

In the presence of parallel trade expected consumer surplus in A and B, respectively, is given by (see the Appendix):

$$\mathrm{CS}_A^{pt} = \frac{4\lambda + (1-\lambda)(1+\theta)^2}{8(1+\lambda+\theta(1-\lambda))^2} \text{ and } \mathrm{CS}_B^{pt} = \frac{4\lambda\theta + (1-\lambda)(1+\theta)^2}{8(1+\lambda+\theta(1-\lambda))^2}.$$

A comparison of expected consumer surplus in the two countries with and without parallel trade 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.*

If $\theta < \frac{1-\lambda}{2-\lambda}$, the manufacturer sets a low wholesale price in B when he does not allow parallel trade; in addition, the retail price in B is equal to zero if demand is low. Allowing parallel trade leads to a higher wholesale price and eliminates the possibility of having a zero retail price. Hence, for these low values of θ, parallel trade can only hurt consumers in B. However, parallel trade also induces retailers to place larger orders, which is good for consumers in both countries. It is this effect that explains why parallel trade always raises expected consumer surplus in A and also raises it in B for $\theta \geq \frac{1-\lambda}{2-\lambda}$.

Finally we turn to world social welfare, $W^j, j=\{pt, npt\}$, which consists of the sum of consumer surplus and the manufacturer's profit (the retailers expect zero profit): $W^j = \mathrm{CS}_A^j + \mathrm{CS}_B^j + \hat{\Pi}j$. We find:[11]

Result 3. *Parallel trade raises expected world welfare if $\theta \geq \frac{1-\lambda}{2-\lambda}$ and reduces it otherwise.*

Fig. 2 illustrates Result 3 and highlights the discontinuity in W^{npt} at $\theta = \frac{1-\lambda}{2-\lambda}$ that is behind it. Since CS_A^{npt} and $\hat{\Pi}^{npt}$ are continuous in θ, the discontinuity in W^{npt} has to come from the fact that CS_B^{npt} is discontinuous. In fact, the discontinuity in CS_B^{npt} is so large as to determine entirely what happens to world welfare at this point. It is easy to see that world welfare has to increase with parallel trade for $\theta \geq \frac{1-\lambda}{2-\lambda}$: the manufacturer's expected profit and expected consumer surplus in both countries are greater than without parallel trade, since the wholesale prices are the same and retailers order more inventory. For $\theta < \frac{1-\lambda}{2-\lambda}$, expected consumer surplus in A is greater with parallel trade for all θ, and profit with parallel trade is greater than profit without parallel trade, at least as long as $\theta > \frac{1-\lambda}{2-\lambda}$. Still, world welfare decreases with parallel trade, since CS_B^{pt} is so much lower than CS_B^{npt} for the reason discussed in connection with Result 2.

The key insight is that the interests of consumers and the manufacturer are often aligned when parallel trade is possible. In fact, 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. This result is in sharp contrast with the conclusions of the literature viewing 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

To obtain a more general result concerning the effect of parallel trade on the manufacturer's profit, consider the following 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

$$\begin{array}{ll} \int_{\theta^-}^{\theta^+} p_A(x_A + m) x_A f(\theta) \mathrm{d}\theta & \text{if } m \geq 0 \\ \int_{\theta^-}^{\theta^+} \{p_A(x_A + m)(x_A + m) + p_B(x_B - m, \theta)(-m)\} f(\theta) \mathrm{d}\theta & \text{if } m < 0, \end{array} \tag{7}$$

[11] Formal proofs of the Results are available from the authors upon request.

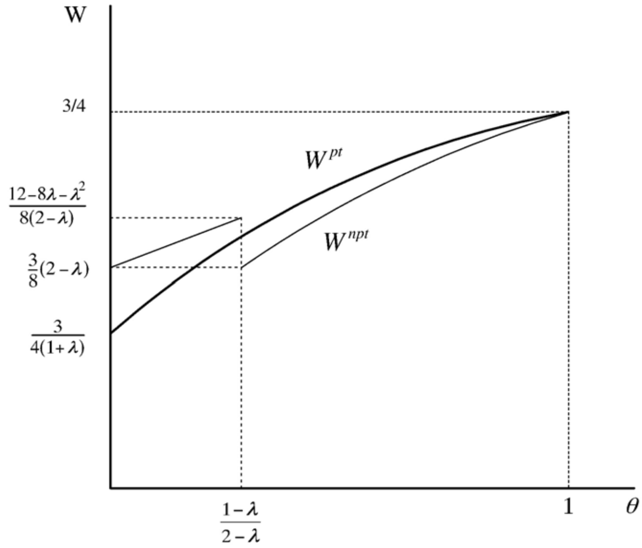

Fig. 2. World social welfare.

and that in B, ER_B^r, as

$$\begin{cases} \int_{\theta^-}^{\theta^+} \{p_B(x_B - m, \theta)(x_B - m) + p_A(x_A + m)m\} f(\theta)\mathrm{d}\theta & \text{if } m \geq 0 \\ \int_{\theta^-}^{\theta^+} p_B(x_B - m, \theta)x_B f(\theta)\mathrm{d}\theta & \text{if } m < 0. \end{cases} \tag{8}$$

Since the expected retail profit, $E\pi_i^r = \mathrm{ER}_i^r - w_i x_i$, has to equal zero for $i=A$, B, we can write the manufacturer's problem of maximizing his expected profit, $\prod = w_A x_A + w_B x_B$, as

$$\max_{x_A, x_B} \int_{\theta^-}^{\theta^+} \{p_A(x_A + m)(x_A + m) + p_B(x_B - m, \theta)(x_B - m)\} f(\theta)\mathrm{d}\theta. \tag{9}$$

The volume of parallel trade, $m = m(x_A, x_B, \theta)$, is equal to the flow of goods that equalizes retail prices across the two countries ex-post for each level of x_A, x_B and θ. It is implicitly defined by

$$p_A(x_A + m) = p_B(x_B - m, \theta) \tag{10}$$

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 Eq. (9) with respect to m must be zero at $m=0$:

$$\int_{\theta^-}^{\theta^+} \{p_A(x_A^*) + x_A^* p'_A(x_A^*) - p_B(x_B^*, \theta) - x_B^* p'_B(x_B^*, \theta)\} f(\theta)\mathrm{d}\theta = 0. \tag{11}$$

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,

$$p_A(x_A^*) + x_A^* p_A'(x_A^*) = \int_{\theta^-}^{\theta^+} \{p_B(x_B^*, \theta) + x_B^* p_B'(x_B^*, \theta)\} f(\theta) \mathrm{d}\theta = 0. \tag{12}$$

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

$$\mathrm{MR}_A^* \equiv (p_A(x_A^*) + x_A^* p_A'(x_A^*)) = 0 \gtrless p_B(x_B^*, \theta) + x_B^* p_B'(x_B^*, \theta) \equiv \mathrm{MR}_B^*(\theta). \tag{13}$$

In this case, an ex-post reallocation of output between markets may raise the manufacturer's expected profit. This can best be seen if we suppose that the manufacturer could handle the distribution of goods himself. For realizations of θ for which $\mathrm{MR}_B^*(\theta) > 0$, 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 $\mathrm{MR}_B^*(\theta) < 0$, the manufacturer would not reallocate goods to A, but sell only enough units in B to make marginal revenue there 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 to reallocate 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 profits 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 profits for this value of θ fall.

Now take expectations over θ. If for given wholesale prices the expected aggregate retail profits in A and B rise when parallel trade is allowed, orders by retailers have to rise to keep the zero–profit constraint satisfied, and hence the manufacturer's expected profit has to increase. By contrast, if the expected aggregate retail profits fall, the manufacturer's expected profit decreases. 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 the retail price falls to zero without parallel trade, then allowing parallel trade raises the manufacturer's expected profit. (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.*

Proof. In case (a), 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

$$p_A\left(1 - \frac{1}{\varepsilon(p_A)}\right) = p_B\left(1 - \frac{1}{\varepsilon(p_B)}\right), \tag{14}$$

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 strictly positive 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 $p_B\left(\frac{x_B^*}{\tilde{\theta}}\right) = 0$, so that the retail price is zero for $\theta \in [\theta^-, \tilde{\theta}]$.[12] We then have to modify Eq. (12) as follows:

$$p_A(x_A^*) + x_A^* p_A'(x_A^*) = \int_{\tilde{\theta}}^{\theta^+} \left\{ p_B\left(\frac{x_B^*}{\theta}\right) + \frac{x_B^*}{\theta} p_B'\left(\frac{x_B^*}{\theta}\right) \right\} f(\theta)\mathrm{d}\theta = 0. \tag{15}$$

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 profits 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 profits will increase with parallel trade.

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

$$p(x_A)\left(1 - \frac{1}{\varepsilon(x_A)}\right) = p(x_B)\left(1 - \frac{1}{\varepsilon(x_B)}\right), \tag{16}$$

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 profits. □

The proposition shows that the form of demand uncertainty plays a crucial role in determining whether the manufacturer likes parallel trade or not.[13] 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: ship goods ex-post to the market where demand is highest so that it pays to hold 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 adjusts ex-post to the realized willingness to pay. Without parallel trade, perfect competition between retailers ensures that this adjustment takes place. Parallel trade, however,

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

[13] The proposition does not depend on whether the retailing sector is perfectly competitive or not. Consider the case of a single retailer in each country. If the realization of θ produces different retail prices, the marginal revenue of the retailer located in the low-price market on the first unit shipped to the high price market is equal to the retail price in that market (and thus necessarily higher than the marginal revenue of selling this unit in his home market). Not surprisingly this means that the flow of parallel trade goes in the same direction as in the presence of a competitive retailing sector. More importantly, this implies that parallel trade raises (lowers) the manufacturer's expected profit with case-a (case-b) demand shocks. The same line of argument can be applied if we have oligopolisitic retailers in each country: we only need to reinterpret marginal revenue as residual marginal revenue.

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.

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. This result arises when four conditions are met: first, retailers must place orders before they know the state of the demand; second, there is a positive probability that states of demand differ across markets; third, the products have little value at the end of the demand period (or equivalently, it is costly to maintain them as inventories); and fourth, we have the right type of demand shock. In particular, we should expect to see parallel trade 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 markets, as well as the electronics market to name just a few. In all these markets, we expect parallel imports to change direction or at least to vary widely in volumes from year to year. This is the case, for instance, in the North American automobile market where the volume of parallel trade between Canada and the US is today a fraction of what it was a few years ago (Automotive News, 2004b). The same is true in the US market for earth-moving equipment where, as a result of the Asian crisis, the gray market share rose from 4.7% in 1997 to 19.5% in 1998 (Business Week, 1998).

It is important to note that there are other possible mechanisms that would allow a manufacturer to raise his profit in the face of demand uncertainty. 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 his 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.

Acknowledgements

We thank Jonathan Eaton, Frank Stähler, an anonymous referee and seminar participants for their useful comments. The authors acknowledge the financial support from the Social Sciences and Humanities Research Council of Canada and from the Transcoop Program of the Alexander-von-Humboldt Foundation and the German Ministry of Education.

Appendix A

A.1. Prices and quantities without parallel trade

If $\max\left\{0, 1-\frac{x_B}{\theta}\right\} = 0$, we obtain from Eq. (1) $x_B = 1-\frac{w_B}{1-\lambda}$, and the solution to the manufacturer's profit maximization problem is $\hat{w}_B = \frac{1-\lambda}{2}$. This implies an order volume of $\hat{x}_B = \frac{1}{2}$, a retail price in the high-demand state of $\hat{p}_B^h = \frac{1}{2}$, and a low-demand retail price of $\hat{p}_B^l = 0$. In the high-demand state, retailers sell their entire inventory, i.e., $q_B^h = \frac{1}{2}$. Realized sales in the low-demand state are equal to $\hat{q}_B^l = \min\left\{\theta, \frac{1}{2}\right\}$.

If $\max\left\{0, 1-\frac{x_B}{\theta}\right\} = 1-\frac{x_B}{\theta}$, Eq. (1) yields $x_B = \frac{\theta(1-w_B)}{\theta(1-\lambda)+\lambda}$, and the manufacturer's expected profit is maximized for $\hat{w}_B = \frac{1}{2}$. This implies that $\hat{x}_B = \hat{q}_B = \frac{\theta}{2(\theta(1-\lambda)+\lambda)}$, $\hat{p}_B^h = \frac{2(\theta(1-\lambda)+\lambda)-\theta}{2(\theta(1-\lambda)+\lambda)}$ and $\hat{p}_B^l = \frac{2(\theta(1-\lambda)+\lambda)-1}{2(\theta(1-\lambda)+\lambda)}$. Comparing profits across the two cases, we find that the manufacturer chooses $\hat{w}_B = \frac{1-\lambda}{2}$ if $\theta<\frac{1-\lambda}{2-\lambda}$ and $\hat{w}_B = \frac{1}{2}$ otherwise.

A.2. Consumer surplus

The expected consumer surplus in B is

$$\mathrm{CS}_B = \lambda(1-\hat{p}_B)\frac{\hat{q}_B}{2} + (1-\lambda)(1-\hat{p}_B)\frac{\hat{q}_B}{2}. \tag{17}$$

In the absence of parallel trade, the specific level of expected consumer surplus depends on whether the retail price drops to zero when the state of demand is low. Using the respective prices and quantities in Eq. (17) yields CS_B^{npt}.

In the case of parallel trade we computed the total volume of orders, $\hat{X} = \frac{1+\theta}{(1+\lambda+\theta(1-\lambda))}$, but could not determine the volume of orders from each country and the volume of parallel trade. To compute consumer surplus for each country, however, all we need is individual country sales which can be found in the following way. Since parallel trade equalizes retail prices, we have $1-\frac{\hat{q}_B}{\theta} = 1-\hat{q}_A$ and thus $\hat{q}_A = \frac{\hat{q}_B}{\theta}$ if demand in B is low, and $\hat{q}_A=\hat{q}_B$ if it is high. In addition, parallel trade ensures that total sales across the two countries must be equal to total orders (i.e., $\hat{q}_A+\hat{q}_B=\hat{X}$). Combining these conditions, we obtain

$$\hat{q}_B = \frac{\theta\hat{X}}{1+\theta} \text{ and } \hat{q}_B = \frac{\hat{X}}{2}$$

if demand in B is low, respectively high. Once we have computed retail prices in B (or in A), we can use Eq. (17) to obtain CS_A^{pt} (once subscripts B are replaced by A) and CS_B^{pt}.

References

Ahmadi, R., Yang, R., 2000. Parallel imports: challenges from unauthorized distribution channels. Marketing Science 19, 279–294.

Automotive News, 2002. Gray market yields gold. 76, Issue 5972, February 25.

Automotive News, 2004a. No matter what color, that Lexus in Asia quite often is gray. 78, Issue 6078, February 2.

Automotive News, 2004b. Once-sizzling Canadian import market loses steam. 79, Issue 6109, August 30.

Business Week, 1998. The earth is shifting under heavy equipment. Issue 3572, April 6.

Cespedes, F., Corey, E.R., Rangan, V.K., 1988. Gray markets: causes and cures. Harvard Business Review 75–82 (July–August).

Computer Reseller News, 2001. Vendors aim to squash gray market. Issue 963, September 17.
Cosac,, T., 2002. Vertical restraints and parallel imports with differentiated products. Unpublished PhD Thesis, Simon Fraser University.
Deneckere, R., Marvel, H.P., Peck, J., 1997. Demand uncertainty and price maintenance: markdowns as destructive competition. American Economic Review 87, 619–641.
Financial Post, 2001. Wanted: Canadian cars. November 3, 2001.
Gallini, N., Hollis, A., 1999. A contractual approach to the gray market. International Review of Law and Economics 19, 1–21.
House of Commons, 1999. Trademarks, fakes and consumers. Eighth Report, Trade and Industry Committee. http://www.parliament.the-stationery-office.co.uk/pa/cm199899/cmselect/ cmtrdind/380/38007.htm.
Knox, D., Richardson, M., 2002. Trade policy and parallel imports. European Journal of Political Economy 19, 133–151.
Lipner, S., 1990. The legal and economic aspects of gray market goods. Quorum Books, London.
Malueg, D.A., Schwartz, M., 1994. Parallel imports, demand dispersion and international price discrimination. Journal of International Economics 37, 167–196.
Maskus, K., 2000. Parallel imports. World Economy 23, 1269–1284.
Maskus, K., Chen, Y., 2002. Parallel imports in a model of vertical distribution: theory, evidence and policy. Pacific Economic Review 7, 319–334.
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.
Richardson, M., 2002. An elementary proposition concerning parallel imports. Journal of International Economics 56, 233–245.
Tirole, J., 1988. The theory of industrial organization. MIT Press, Cambridge.
World IT Report, 2003. 30 percent of Malaysian cells purchased through grey market. December 28.