ELSEVIER Journal of International Economics 37 (1994) 167–195

Journal of INTERNATIONAL ECONOMICS

Parallel imports, demand dispersion, and international price discrimination☆

David A. Malueg[a], Marius Schwartz[b,c,*]

[a]Department of Economics, Tulane University, New Orleans, LA 70118-5692, USA
[b]Department of Economics, Georgetown University, Washington, DC 20057-1045, USA
[c]Antitrust Division, U.S. Department of Justice, Washington, DC 20530, USA

Received October 1992, revised version received December 1993

Abstract

Parallel imports, goods imported by unauthorized resellers, are advocated worldwide for undermining international price discrimination. For a continuum of markets, we find that uniform pricing by a monopolist yields lower global welfare than third-degree discrimination if demand dispersion across markets is 'large': though uniform pricing avoids output misallocation, too many markets go unserved. Mixed systems, permitting discrimination across but not within designated groups of markets, yield significantly higher welfare than uniform pricing or unrestricted multimarket discrimination, and can Pareto dominate uniform pricing. Thus, while parallel imports might benefit some countries, our results weaken the (multilateral) case for allowing them.

Key words: Parallel imports; International price discrimination

JEL classification: D42; F12; F13; L42

* Corresponding author.

☆ The views expressed in this paper do not purport to represent those of the U.S. Department of Justice. For helpful comments we thank Jonathan Eaton, Gary Hufbauer, Serge Moresi, Greg Werden, two referees and an editor of this Journal, and seminar participants at Georgetown University and Pennsylvania State University. We also thank Ann Plamondon and Eugene Poirier for generous assistance.

SSDI 0022-1996(94)01322-J

1. Introduction

Parallel imports, or 'gray-market' imports, are genuine products – not counterfeits – imported by unauthorized resellers. A common situation is where one firm owns the national trademarks in several countries, each trademark conferring the exclusive distribution right in that country, but another party obtains the product in one country (typically from wholesalers rather than the trademark holder) and diverts it to another country without the authorization of the trademark holder. There are numerous variations on this scenario, and the legal treatment of parallel imports can differ across the various settings.[1] The basic economic question, however, is the same: Should a producer be entitled to enforce exclusive-distribution territories internationally?

Accurate data on parallel imports are limited, because the business is inherently rather secretive. Still, the phenomenon appears important. Parallel imports into the United States increased dramatically with the dollar's rapid appreciation in the early 1980s, and by the mid-1980s were estimated at \$7–10 billion, or 2–3 percent of the U.S. import bill. Moreover, they were disproportionately concentrated in particular products – typically name-brand consumer goods such as cosmetics and fragrances, luxury automobiles, and cameras – products in which they accounted for 15–20 percent of all sales [Chard and Mellor (1989), *Business Week* (1985, 1988)]. Although parallel imports into the United States appear to have peaked in the mid-1980s, they remain significant both in the United States and worldwide. For example, parallel imports of pharmaceuticals alone within the EC were estimated at over \$500 million in 1990, and projected to grow rapidly [REMIT (1992)]; a surge of parallel imports of pharmaceuticals from Mexico to the United States also is predicted if the

[1] The products may be covered by other intellectual property rights (IPRs) – in the parallel imports context trademarks are the most common IPRs, then copyrights, and occasionally patents – or by no IPRs. The authorized national distributors may be commonly owned as above (the same firm holds the various national trademarks), linked by licensing, or entirely independent. And the imported goods are usually produced abroad, but sometimes are produced domestically, exported, then reimported. For details on the legal treatment of parallel imports in these various cases in the United States and the EC, see Hawk (1991).

North American Free Grade Agreement materializes [*Drug Store News* (1993)].[2]

The attention parallel imports command further attests to their importance. In the United States, extensive litigation continues, and both opponents and proponents have repeatedly tried to enact federal legislation (see Hearings on S. 626, 1990). In the EC, a recent controversy centers on the role of national exclusive-distribution territories in sustaining different car prices between member states [*Financial Times* (1992)]. And Japan's Fair Trade Commission recently issued enforcement guidelines focusing prominently on parallel imports [JFTC (1991)].

Generally, policies worldwide firmly support parallel imports.[3] This support stems largely from a belief that parallel imports are driven not so much by free riding on promotional efforts of authorized distributors, but by international price discrimination. Why has the price discrimination view generated such strong support for parallel imports, whereas use of exclusive territories within a country typically elicits much less hostility? We see two likely explanations. First, the scope for price discrimination is probably greater internationally: disparities in demand elasticities across countries are likely to be greater than across regions within a country, most obviously because of the greater differences in per capita incomes between countries

[2] More broadly, articles found in a LEXIS–NEXIS search of the trade press, January 1989–May 1993, reveal the following. Parallel imports continue in many of the same products as in the mid-1980s: consumer electronics, automobiles, spirits, watches, cosmetics and fragrances (where their estimated worldwide market share is 20–30 percent). They have extended to additional consumer goods including haircare products, athletic sneakers, camcorders, and personal computers. They also have been reported in industrial products such as semiconductor chips and construction equipment. A survey of U.S. exporters to Asia [Palia and Keown (1991)] offers a clue to the pervasiveness of parallel imports: of 141 respondents that used sole import distributors (55 percent of all respondents), 41 percent reported problems with parallel imports in the past five years.

[3] The U.S. Supreme Court's decision in *K mart Corp. v. Cartier, Inc.*, 486 U.S. 281 (1988), upheld the Customs Service's policy of not excluding parallel imports of trademarked goods whenever the U.S. trademark holder and the trademark holder in the country where the parallel imports originated are commonly owned or controlled. The EC, in a stance that remains controversial, insists that unimpeded parallel imports among member states are central to promoting a common market [Hawk (1991)]. It treats as unlawful, under the competition provisions and free movement of goods provisions of the Treaty of Rome, the use of exclusive territories that overlap with national borders. The JFTC's (1991) guidelines arguably go furthest, prohibiting actions against parallel imports not only on Japanese soil but also abroad.

than regionally within countries.[4] A priori, therefore, international price discrimination is a plausible explanation for parallel imports.

Second, relative to uniform pricing, geographic price discrimination within a country typically harms consumers in some regions but benefits others. In contrast, countries facing relatively high prices under international price discrimination tend to ignore the gains from discrimination to consumers in low-price countries. Thus, high-price countries generally perceive that their welfare will be higher under uniform pricing.[5] Such perceptions account for the support for parallel imports in many countries, including the United States, Japan, and Australia.[6]

This paper departs from the traditional debate by analyzing parallel imports from the perspective of *world* welfare (the sum of profit and consumers' surplus in all countries) rather than national welfare. National welfare is an inadequate criterion for designing international trading rules, since quid pro quos among countries to compensate losers are generally possible in multilateral negotiations (perhaps via concessions in other areas). Our change of focus to global welfare, therefore, is relevant for evaluating multilateral approaches to the question of parallel imports (see also the discussion in section 5).

In our analysis we shall abstract from free-rider explanations for parallel imports – even though parallel imports *are* partly explained by free riding (see section 2). We choose to focus solely on the price-discrimination explanation purely as a modeling strategy: the price-discrimination scenario is probably the 'best case' for allowing parallel imports, and our theme is that the case for parallel imports is tenuous even then.

We consider a monopolist producer of a final good, facing markets with

[4] The following figures illustrate these differences in 1990. The quintile of the U.S. population represented by the states with the greatest per capita incomes had a per capita income approximately 1.5 times as great as did the quintile of the population in the poorest states [see U.S. Bureau of the Census (1992, Tables 25 and 687)]. In contrast, for the EC countries, instead of the states in the United States, the comparable ratio is about 2.3 [see U.S. Bureau of the Census (1992, Table 25 1370)]. Worldwide these income differences were significantly greater. Among World Bank member countries, the richest quintile (based on world population) had a per capita GNP approximately 75–85 times as great as in the poorest quintile [World Bank (1992, Table 1)]. Even after adjusting for purchasing power parity, as done in the United Nations' International Comparison Program, this ratio lay in the range 15–20 [World Bank (1992, Table 30)].

[5] We will indicate in section 5, however, why these perceptions need not be correct.

[6] The EC is a special case; as noted earlier, price discrimination within the EC is condemned for allegedly delaying economic integration, not for its distributional impact.

different, known demands.[7] If legally permitted to curb parallel imports, the monopolist can charge a different monopoly price in each market. We call this *complete discrimination*, to distinguish it from mixed systems (discussed shortly), which permit discrimination between but not within designated groups of markets. (Observe that complete discrimination is not perfect discrimination, since demand in each market is downward sloping.) If prohibited from curbing parallel imports, the monopolist is constrained by the threat of arbitrage to set a uniform price.

This formulation embodies several simplifying assumptions: arbitrage by end users is not feasible; consumers view parallel imports as perfect substitutes for authorized goods; and, if legally permitted, parallel imports would be in perfectly elastic supply. These assumptions abstract from some interesting issues;[8] our goal, however, is to develop a tractable and transparent model focused on the discrimination question.

It is well known that, in general, overall welfare can be higher under uniform-price monopoly or under discrimination. Demand dispersion, however, is likely to be greater internationally than between regions within a country, and we conjecture that high dispersion makes welfare higher under discrimination. This conjecture holds in familiar two-market examples with non-increasing marginal cost [see, for example, Tirole (1988) or Hausman and MacKie-Mason (1988)]: for large enough dispersion, a uniform-price monopolist would serve only the high-demand market, but under discrimination would add the second market and not raise price in the first. The two-market case, however, conceals a potential tradeoff that can arise with more than two markets. In those markets that would continue to be served under uniform pricing, increased dispersion could increase the loss from discrimination, because of the greater scope for misallocating output; this effect must be weighed against the greater number of markets served under discrimination. The conjecture that dispersion favors discrimination from the standpoint of global welfare therefore remains to be verified in a model with more than two markets.

Section 2 of the paper reviews the debate over the causes of parallel

[7] Focusing on a final rather than intermediate good is motivated by the fact, noted earlier, that parallel imports are concentrated in consumer goods. Regarding the monopoly assumption, a perfectly competitive model would be inappropriate, as most parallel imports involve differentiated products, characterized by high fixed costs (e.g. R&D, advertising, and marketing expenses). The modeling choice is between monopoly and imperfect competition, and we opt to study monopoly so as to focus on price discrimination free of strategic complications.

[8] For example, parallel imports are somewhat differentiated from authorized goods (in packaging, warranty coverage, and other dimensions), and issues of consumer confusion and damage to product reputations have been raised in the debate (see Hearings on S. 626).

imports, concluding that discrimination has played a part. Section 3 presents a model with a continuum of markets and confirms the conjecture that for 'high' dispersion, discrimination yields higher welfare – because of the powerful effect of serving more markets. Section 4 considers 'mixed systems'. Markets are placed into designated groups (e.g. by international agreement), with different prices allowed between but not within groups; such groupings may preserve the benefits of discrimination from serving more markets, but limit the misallocation effects. We present one such mixed system that yields higher welfare than complete discrimination and Pareto dominates uniform pricing. This system is shown to be the optimal mixed system with no 'holes' – if two markets are in a group, so are all markets that have marginal valuations (at any quantity) lying between the two; but mixed systems with holes can yield even greater welfare. Section 5 concludes that *some* price discrimination probably would increase global welfare, and discusses countries' incentives to curb parallel imports to sustain discrimination given the distributional issues involved.

2. The debate over parallel imports

Simplifying somewhat, opponents argue that parallel imports are profitable mainly because parallel traders free ride on investments of authorized distributors at various levels of the distribution chain, e.g. on national advertising by the authorized importer or on local advertising, display, or other services provided by wholesalers or retailers [Lexecon (1985), COPIAT (1986), DeMuth (1990)]. Parallel importers can obtain goods at prices that do not reflect many of these costs by purchasing abroad sufficiently high up the distribution chain (e.g. from wholesalers) for sale to a different market. They are then able to undercut the domestic authorized distributors, which do incur the expenses needed to cultivate and maintain local demand and reputation for the product. Opponents argue that such free riding disrupts a supplier's overall marketing plan, discourages various investments, and is generally inefficient.

Under the free-rider hypothesis, parallel trade does not rely on price differentials or other international asymmetries. Indeed, it can profitably occur in both directions, by the same logic as in the reciprocal 'dumping' model of Brander and Krugman (1983).[9] In contrast, proponents argue that parallel imports are primarily an arbitrage response to international price discrimination that a supplier tries to sustain via exclusive distribution

[9] Note, however, that the Brander–Krugman model implies price discrimination in favor of the foreign market – the reverse of what is alleged by supporters of parallel imports.

territories over different national markets. Proponents laud parallel imports as undermining such discrimination.

It is difficult to determine empirically whether the free-riding or discrimination hypothesis predominates, even in a single industry, as the hypotheses are not mutually exclusive. The difficulties stem from both the paucity of appropriate data and the subtlety of the testable implications. For instance, retail price differentials between countries could be due to differences in distribution costs and not price discrimination by a manufacturer; potentially more informative are the manufacturer's export prices to different markets, but these are not generally available. Observing parallel imports flowing only in one direction also does not prove discrimination; it could merely reflect differences in demand attributable to different levels of 'free-ridable' distributor investments in the two countries.[10] These caveats notwithstanding, there is some suggestive evidence regarding the causes of parallel imports.

Generally, the type of goods in which parallel imports are concentrated – name-brand consumer goods, entailing heavy promotional investments – suggests free riding. But the timing of parallel imports points to an arbitrage explanation. Parallel imports have generally surged as a country's exchange rate appreciated, suggesting 'incomplete pass-through' – that import prices in the destination currency were not reduced in the same proportion as the appreciation of that currency, thereby creating scope for arbitrage. An extensive literature documents that, indeed, incomplete pass-through is quite common [see Knetter (1989), Marston (1990) and Kasa (1992), and the references they cite]. As noted by several authors [Tarr (1985), Dornbusch (1987), Krugman (1987), Feenstra (1989)], for plausible demand conditions in the foreign market, incomplete pass-through can be an equilibrium response by an imperfectly competitive firm or by a monopolist to the new demand conditions caused by the appreciation of the destination currency when the exporting firm's costs are sticky in its own currency.

Neither piece of evidence is conclusive. The concentration of parallel imports in upscale consumer goods might alternatively arise because such goods are highly differentiated, and the associated market power encourages considerable price discrimination. Also, the co-movement of parallel imports with the importing country's exchange rate can alternatively be due to increased scope for free riding. If a manufacturer's export price to the United States reflects marketing costs incurred in dollars, a dollar appreciation can expand the gap between the manufacturer's prices to the United States and to other markets purely on cost grounds; this increased gap can

[10] Such differential expenses can, in turn, reflect differences in costs of promotion across markets due to factor prices, or different levels of promotion reflecting different brand-positioning strategies.

further attract free-rider-based parallel imports [Tarr (1985), DeMuth (1990)]. Finally, even if incomplete pass-through is not cost-based, it need not imply price discrimination. It could instead reflect adjustment costs and various dynamic considerations, especially when exchange rate movements might be transitory or unexpected [Hilke (1988), Froot and Klemperer (1989), Kasa (1992)].

For the United States in the early 1980s, Hilke (1988) concludes that while no single explanation of parallel imports was completely adequate, incomplete pass-through (due to either conscious discrimination or adjustment lags) was more persuasive than was the free-rider hypothesis. Hilke's inference rests largely on the co-movement of parallel imports with the dollar, which we noted is also consistent with a free-rider explanation. However, for some products for which data were available, Tarr (1985) finds that the differentials in the manufacturers' prices to the United States versus their domestic markets exceed plausible estimates of the differential marketing costs.[11] He concludes that free riding was an important factor in some industries, such as perfumes, but that discrimination was a substantial factor in other industries, including German automobiles, Japanese cameras, ski equipment, and champagne.

More generally, there is widespread evidence that manufacturers engage in international price discrimination. Knetter (1989) interprets his findings of incomplete pass-through by German exporters from 1977 to 1985 in these terms. Marston (1990) detects incomplete pass-through by Japanese exporters and, significantly, finds that this behavior is not primarily explained by adjustment lags. Cross-sectional comparisons also find price discrimination internationally, e.g. in luxury automobiles (Mertens and Ginsburg (1985)], pharmaceuticals [Schut and Van Bergeijk (1986)], and books [Prices Surveillance Authority (1989)].

Our reading of the above findings is that price discrimination is *a* factor in explaining parallel imports (though not necessarily the major factor). The support for parallel imports on these grounds therefore merits closer scrutiny.

[11] For example, Tarr notes that despite a 40 percent appreciation of the dollar against the mark from 1980 to 1984, Mercedes-Benz's U.S. prices remained constant in dollars, implying a 40 percent increase in terms of marks. Assuming that Mercedes-Benz's expenses on U.S. marketing were 20 percent of its total costs (production accounting for 80 percent), and that in the United States these costs must be paid in dollars, the differential in marketing costs changed by only 8 percent.

3. Uniform pricing vs. complete discrimination

3.1. The basic model

We consider a monopolist with zero marginal cost. The monopolist faces a continuum of markets, the continuum having total mass of 1. Each market can be viewed as a country. The inverse demand function in market a is $p(q) = a(1-q)$. Thus, demand functions are linear with equal horizontal intercepts but different vertical intercepts (choke prices), a; note that at any price, higher demands have lower elasticities. This demand structure emerges, for example, if demanders have an identical particular utility function separable in income and the monopolist's product, but have different incomes, hence different marginal utilities of income.[12] The intercept a is uniformly distributed over $[1-x, 1+x]$, where the parameter $x \in [0, 1]$ measures demand dispersion for the continuum.

The assumption that demands are linear aids tractability. Its main purpose, however, is to ensure a welfare tradeoff between discrimination and uniform pricing. Discrimination has the advantage of opening up new markets. We want to allow uniform pricing to be *potentially* superior. The simplest way of ensuring this is with linear demands. If a set of linear-demand markets is served under both pricing schemes (each market purchases positive output), then total output from those markets would be the same under the two regimes;[13] whenever total output is no higher under discrimination than under uniform pricing, welfare will be lower under discrimination (if demands are continuous and strictly decreasing) – because discrimination fails to equate marginal valuations.[14] The assumption of linear demands therefore is relatively favorable for uniform pricing: if under uniform pricing the monopolist drops no markets, uniform pricing yields higher welfare than discrimination.

Consider first the equilibrium under complete discrimination. Given our

[12] The following discussion is based on Tirole (1988, ch. 2). Let y denote consumption of the numeraire good and q consumption of the monopolist's good, and suppose the consumer has utility function $V(y, q) = u(y) + q - (1/2)q^2$. If a consumer has income I and the monopolist's price per unit of the good is p, then the inverse demand function, in implicit form, is $p(q) = (1-q)/(u'(I-pq))$, which is 0 at $q = 1$ for all I, and is increasing in I, assuming u is concave. For pq 'small', $u'(I-pq)$ approximately equals $u'(I)$, so the inverse demand function is approximately linear. This is simply the inverse demand function of the model, with the vertical intercept a corresponding to $1/u'(I)$; thus, higher values of a correspond to higher incomes. For tractability, our basic model assumes that a is uniformly distributed. Subsection 3.3 below discusses skewed distribution.

[13] See the discussion in section 4 following Lemma 1.

[14] This welfare result holds for any marginal cost function [see Schwartz (1990)].

zero cost and 'rotating demands' assumptions, for all values of the dispersion parameter x the monopolist serves all markets, sets equal outputs in all markets, and charges a price that increases with type: $q(a) = 1/2$, $p(a) = a(1 - q(a)) = a/2$. Total output thus remains unchanged as x varies. Because a is distributed uniformly over $[1-x, 1+x]$, the mean price is 1/2. The monopolist's profit is therefore $\Pi^{\mathrm{d}} = 1/4$. For linear demands, consumer surplus in each market is one-half of profit. Thus, for all $x \in [0, 1]$, complete discrimination (hereinafter 'discrimination' for brevity) yields profit, consumer surplus, and welfare of $\Pi^{\mathrm{d}} = 1/4$, $S^{\mathrm{d}} = 1/8$, and $W^{\mathrm{d}} = 3/8$, respectively.

Under uniform pricing, the aggregate demand function facing the monopolist consists of two segments, depending on whether, at the given price, all markets are served. In case 1, price falls in the range $0 \leqslant p \leqslant 1 - x$, and all markets are served.[15] Case 2 has price in the range $1 - x < p \leqslant 1 + x$; in this case, only types $a \in (p, 1 + x]$ are served. Letting $b = \max\{p, 1 - x\}$, aggregate demand is therefore

$$
\begin{aligned}
Q(p) &= \frac{1}{2x} \int_{b}^{1+x} \left(1 - \frac{p}{a}\right) \mathrm{d}a \\
&= \begin{cases} 1 - \dfrac{p}{2x} \log\left(\dfrac{1+x}{1-x}\right), & \text{if } 0 \leqslant p \leqslant 1 - x\,, \\[2ex] \dfrac{1+x}{2} - \dfrac{p}{2x}\left(1 + \log\left(\dfrac{1+x}{p}\right)\right), & \text{if } 1 - x < p \leqslant 1 + x\,. \end{cases}
\end{aligned}
$$

It is straightforward to verify that the demand function $Q(p)$ is continuous and strictly decreasing everywhere. For prices below $1 - x$, the function is linear, since all the (linear-demand) markets are served; for prices above $1 - x$, it is strictly convex, because lowering price increases output at an increasing rate due to the newly served markets.

The relevant equilibrium expressions, derived in Malueg and Schwartz (1993), are reported in Table 1. In Table 1, y is the unique number greater than 1 satisfying $y = 1 + 2\log(y)$: $y \approx 3.5128$. The monopolist chooses to serve all markets for 'low dispersion', $x \leqslant x^*$, and drop some for $x > x^*$, where $x^* \equiv (y - 1)/(y + 1) \approx 0.5568$.

3.2. Welfare comparisons

We are interested in comparing the two pricing regimes for different values of x. Recall that output, profit, and consumer surplus – and therefore

[15] In the case where $p = 1 - x$, the market with $a = 1 - x$ purchases $q = 0$, but this market is itself of measure zero. In this case we continue to say that (almost) all markets are served.

Table 1
Equilibrium under uniform pricing and discrimination

	Case 1 $0 \leq x \leq x^*$	Case 2 $x^* < x \leq 1$	Discrimination
p	$p_a = \dfrac{x}{\log\left(\dfrac{1+x}{1-x}\right)}$	$p_s = \dfrac{1+x}{y}$	$p(a) = \dfrac{a}{2}$
$Q(p)$	$\dfrac{1}{2}$	$\left(\dfrac{y-1}{y}\right)\left(\dfrac{1+x}{4x}\right)$	$\dfrac{1}{2}$
Π	$\dfrac{p_a}{2}$	$\left(\dfrac{y-1}{4x}\right)p_s^2$	$\dfrac{1}{4}$
S	$\dfrac{1}{2} - \dfrac{3}{4}p_a$	$\dfrac{(y-1)(y-2)}{8x}p_s^2$	$\dfrac{1}{8}$
W	$\dfrac{1}{2} - \dfrac{1}{4}p_a$	$\dfrac{y(y-1)}{8x}p_s^2$	$\dfrac{3}{8}$

also welfare – all are constant under discrimination. (This is a special feature of zero cost and our 'rotating demands', as discussed in subsection 3.3 below.)

We now turn to uniform pricing. Fig. 1 shows that the uniform price decreases with x until x^* and increases thereafter. The price decreases initially because at $p < 1 - x$, increases in x raise the elasticity of aggregate demand. (Increasing x is a mean-preserving spread for the inverse but not for the direct demand functions, and the increase of elasticity in low-demand markets outweighs the decrease in high-demand markets.) For $x > x^*$, however, the monopolist sets $p > 1 - x$. In this range, increases in x lead to the dropping of high-elasticity (low a) markets, thereby decreasing the elasticity of aggregate demand and inducing the monopolist to raise price. Table 1 shows that total output $Q(p)$ is constant at 1/2 for $x \in [0, x^*]$ but decreases with x for $x \in (x^*, 1]$.[16]

Fig. 2 graphs consumer surplus, profit, and welfare under the two regimes. Clearly profit is always lower under uniform pricing, but the rankings of consumer surplus and welfare depend on x. Consider first the range over which the monopolist serves all markets, $x \in [0, x^*]$. As x increases in this range, output remains at 1/2. Had the allocation of this

[16] Output remains at 1/2 initially because for $x < x^*$ the monopolist serves all markets and, hence, operates in the linear segment of aggregate demand. As x increases, the linear segment pivots counterclockwise about the horizontal intercept of 1, hence the corresponding marginal revenue curve always cuts the monopolist's zero marginal cost as output of 1/2. For $x > x^*$, the monopolist operates in the strictly convex portion of aggregate demand, and the output at which the corresponding marginal revenue equals zero decreases with x.

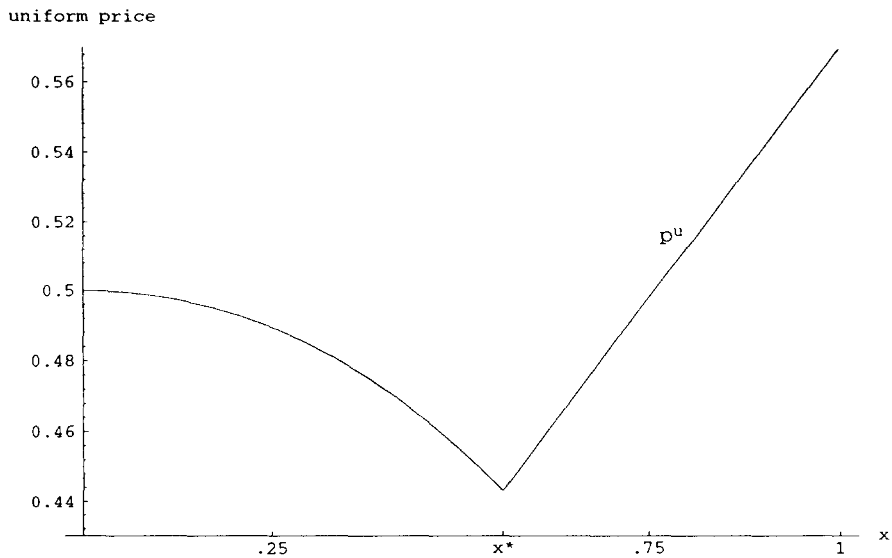

Fig. 1. The monopolist's uniform price, p^u.

output across markets also remained constant (as it does under discrimination), welfare would have remained constant, since the mean inverse demand curve remains unchanged. But under uniform pricing, increased x causes reshuffling of output from low-demand markets to high-demand ones, which equates marginal valuations and thus increases welfare. In contrast, profit falls because output is constant but price is cut. The welfare gain as x increases thus reflects increased consumer surplus; hence, consumer surplus is higher than under discrimination.

Consider next the range over which the monopolist drops some markets, $x \in (x^*, 1]$. As x increases beyond x^*, output drops for two reasons: the pure effect of dispersion, and the induced change in price. Holding price constant, increasing dispersion would reduce output because more markets are dropped (their demands become zero at the old price). The fact that more markets are dropped, in turn, leads the monopolist to re-optimize and raise price, further reducing output. Profit continues to fall as x increases beyond x^* (albeit more slowly than for $x < x^*$), but now the reduction in output and increase in price cause consumer surplus also to fall (in contrast to the range $x < x^*$). Nevertheless, consumer surplus remains higher than under discrimination. Total welfare is subject to two opposing effects as x

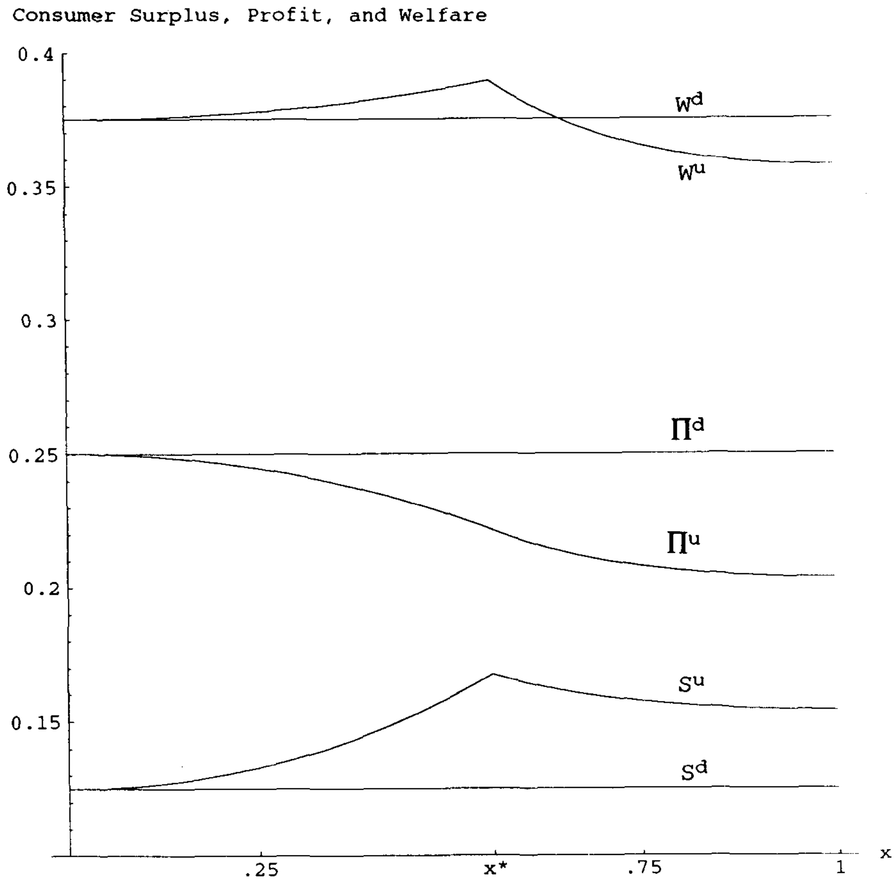

Fig. 2. Consumer surplus, profit, and welfare.

increases: improved output allocation across markets (as in the range $x < x^*$), but decreasing total output. The latter effect dominates for all $x > x^*$. Thus, once the dispersion in demands is large enough that the monopolist drops some markets under uniform pricing, any further increase in dispersion reduces welfare.

Fig. 3 shows the ratio of welfare under discrimination to welfare under uniform pricing, W^d/W^u. It starts at 1 (for $x = 0$), falls to a minimum of about 0.965 (at $x^* \approx 0.56$), and rises to a maximum of about 1.048. When uniform pricing serves all markets ($x < x^*$), total output is the same as under discrimination, hence increased demand dispersion lowers W^d/W^u, due to the increased misallocation of output under discrimination. When some markets are dropped ($x^* < x \leq 1$), welfare under uniform pricing decreases

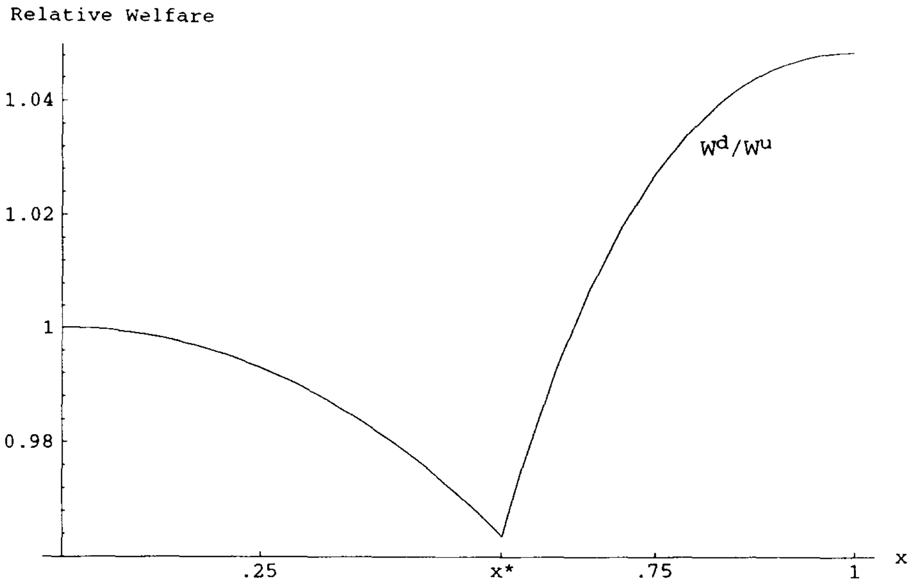

Fig. 3. Relative welfare and dispersion, W^d/W^u.

with x due to the fall in output, but remains constant under discrimination; eventually discrimination dominates.[17] Proposition 1 summarizes these findings:

Proposition 1. For zero demand dispersion, $W^d/W^u = 1$. For 'small' demand dispersion ($x < x^*$), all markets are served under both discrimination and uniform pricing; in this range, W^d/W^u decreases monotonically with dispersion. For 'large' dispersion ($x > x^*$), some markets are dropped under uniform pricing; in this range, W^d/W^u increases monotonically with dispersion and exceeds 1 when dispersion is sufficiently large.

In our model discrimination can yield anywhere from about a 4 percent loss in welfare to a 5 percent gain, relative to uniform pricing.[18] These magnitudes are modest by the standards of the rent-seeking literature. However, compared with standard estimates of 'triangle' losses from

[17] To see analytically that discrimination eventually dominates, observe that at $x = 1$, $W^u = (y-1)/2y \approx 0.3577$, whereas $W^d = 0.375$. From Table 1, $W^u = W^d$ for $x = x_e$, which is the solution to $[(y-1)(1+x)^2]/[xy] = 3$. Therefore the two regimes yield equal welfare for dispersion level $x_e = \{(y+1) - [3y(4-y)]^{1/2}\}/2(y-1) \approx 0.646$.

[18] Interestingly for policy, discrimination yields the largest welfare gain when demand dispersion is largest; it is also when demand dispersion is largest that the monopolist has the greatest incentive to engage in discrimination.

allocative inefficiency, such as Harberger's for monopoly or others' for trade barriers, these numbers are significant.[19] Moreover, the magnitudes can be significantly larger when the distribution of markets is not uniform but skewed (see subsection 3.3 below).

It is instructive to explore further why increased dispersion eventually makes uniform pricing inferior to discrimination. As noted, increasing x beyond x^* causes output under uniform pricing to drop for two reasons: more markets would be dropped at the old price and, faced with the truncated distribution, the monopolist raises price to those markets still served. To assess the effect of the latter, we computed the hypothetical welfare level $W^{\mathrm{u}}(x \mid p = 1 - x^*)$ for $x \in [x^*, 1]$, where $1 - x^*$ (≈ 0.4432) is the monopolist's price when dispersion is at the maximum level consistent with the monopolist choosing to serve all markets.

Remark 1. At demand dispersion x^* the monopolist would charge a uniform price $p = 1 - x^*$ and just serve all markets. Holding the uniform price fixed at $1 - x^*$, for all $x > x^*$ welfare at this uniform price, $W^{\mathrm{u}}(x|p = 1 - x^*)$, price would be higher than under discrimination, despite the dropping of markets under uniform pricing.

Remark 1 and Proposition 1 show that increased dispersion eventually makes uniform pricing inferior to discrimination not because of the 'impact effect' of dropping of markets, but because dropping markets leads the monopolist to reoptimize and raise price.

3.3. Robustness

Here we discuss the robustness of the welfare patterns described in Proposition 1. We consider three variations of the basic model: positive marginal cost, a family of 'parallel' rather than 'rotating' demands, and rotating demands with a skewed rather than uniform distribution of markets.

Positive cost. It is difficult to solve explicitly for the uniform price as a function of constant marginal cost, c, and dispersion when only some markets are served. However, for maximum dispersion in this model, $x = 1$, we find that $W^{\mathrm{d}}/W^{\mathrm{u}} > 1$ for all values of cost c that we checked between 0

[19] Note also that the empirical literature often reports welfare gains as a percentage of GNP, which excludes consumer surplus; doing so in our model, by computing the welfare gain as a percentage only of revenue – which in the model coincides with profit – would increase the above maximum percentage gain from discrimination substantially, from about 4.8 percent to about 7.8 percent (since for $x = 1$, profit and consumer surplus under uniform pricing are respectively about 0.204 and 0.155, while welfare under discrimination is 0.375).

and 1. Thus, W^d/W^u follows the pattern of Proposition 1 also for any positive cost $c \in [0, 1)$.[20]

Parallel demands. We also consider an alternative family of inverse demands, $p = a - q$, where the intercept a is uniformly distributed over $[1 - x, 1 + x]$ and $x \in [0, 1]$. Thus, demands would be parallel shifts of each other rather than rotations as earlier.[21] Under both pricing regimes, dispersion affects the individual variables of interest quite differently than with 'rotating demands' [see Malueg and Schwartz (1993)]. However, the ratio W^d/W^u behaves in the same way as described in Proposition 1.

Skewed distributions of demand. A more important extension is to allow for demand distributions to be skewed rather than uniform, since actual distributions of per capita incomes are skewed towards low incomes. We consider the basic model with zero cost and rotating demands (with dispersion parameter x), but now assume that the distribution of intercepts a is given by $f(a \mid t, x) = k(t, x)(a - (1 - x))^t$, where $t > -1$, $a \in (1 - x, 1 + x)$ and where, given t and x, $k(t, x)$ is a parameter that makes the density integrate to 1. This density simply results from a linear transformation of a beta distribution on the interval $(0, 1)$ to the interval $(1 - x, 1 + x)$. The uniform distribution case discussed above corresponds to $t = 0$. For $t > 0$ the distribution of intercepts is skewed toward the right (higher intercepts), for $t < 0$ it is skewed towards the left. As t approaches -1 (from above) the distribution of intercepts becomes more concentrated at the lower end. Thus, compared with $t = 0$, negative values of t might better approximate demand distributions in a world where per capita incomes are skewed toward lower incomes. The following discussion is based on results found using numerical methods to solve this model [see Malueg and Schwartz (1993)].

Consider values $x < 1$ such that, when the distribution of intercepts is uniform ($t = 0$), $W^d/W^u > 1$ (here the monopolist drops some markets under uniform pricing). Now consider increasing skewness toward low demands by

[20] We know that $W^d/W^u = 1$ at $x = 0$ and decreases initially (since all markets are then served for small x, hence output is the same under the two regimes but discrimination misallocates it). After some intermediate level of x, output becomes lower under uniform pricing, as markets are dropped, and W^d/W^u increases, since we verified that it eventually exceeds 1 (for the case of maximum dispersion, $x = 1$). Note that, unlike the zero-cost case, total output under both regimes now decreases with x over the entire range $[0, 1]$, but this difference does not affect the relative welfare patterns (though W^d/W^u, while it remains above 1, does decrease with cost at the maximum dispersion level of $x = 1$).

[21] Such parallel demands can emerge, for example, by assuming the same conditions that gave rise to our rotating demands (see subsection 3.1 above) but adding the assumption that the markets with higher per capita incomes also have larger populations of consumers. We focus on the rotating demands in the paper because it is higher per capita income and not different populations of identical consumers that generates the difference in demand elasticities.

reducing t toward -1. As skewness (to the left) increases, the monopolist lowers its uniform price. However, for modest levels of skewness the price falls slower than markets are shifted to the low end of the distribution, leading to a reduction in the proportion of markets served under uniform pricing and to an increase in W^d/W^u. As t nears -1, the distribution becomes concentrated around $1-x$, and the monopolist's profit function becomes double-peaked, achieving one local maximum at a price above $1-x$ (serving only some markets) and the second at a price below $1-x$ (serving all markets). At some threshold level of skewness t^*, these two local maxima yield equal profit. As t falls below t^*, the monopolist discontinuously drops its price below $1-x$ and serves all markets, causing welfare under uniform pricing to exceed welfare under discrimination. The threshold level of skewness, t^*, decreases with the dispersion level, x. For $x=1$, the monopolist would never serve all markets (since doing so would require setting price to zero), and W^d/W^u increases monotonically as t approaches -1.

Skewness also increases the range of W^d/W^u, compared with the uniform-distribution case. For example, when $x=1$, $W^d/W^u = 1.048$ for $t=0$ but exceeds 1.17 as t approaches -1. For values of x slightly less than 1, W^d/W^u falls below 0.80 for very high levels of skewness (t near -1, where the monopolist now serves all markets); but for 'moderate' skewness, t just above t^*, we have $W^d/W^u > 1.20$. Summarizing, skewness increases the potential difference in welfare under discrimination and uniform pricing and need not erode the advantage of discrimination.[22]

[22] Given the highly stylized nature of our model, we are hesitant to lean on it too heavily for predicting that global welfare would be higher under complete discrimination than under uniform pricing. In particular, we cannot estimate the key parameters x and t^* from actual distributions of per capita income, since our demand parameter a is the inverse of the marginal utility of income, $1/u'(y)$, not income itself, and we did not specify a functional form for $u(y)$ (we only assumed concavity, see footnote 12). Still, one can hazard an inference from the observation that, for many products subject to parallel imports, some countries are not served.

One explanation is that, even if discrimination is feasible (because the threat of parallel imports in practice is not too strong), serving such markets is unprofitable as their 'choke prices' lie below the *positive* marginal cost of production. This scenario resembles the case of positive cost ($c>0$) and maximum dispersion ($x=1$), reported above, insofar as some markets are always dropped. In that case ($x=1$, $c>0$) we found, for all $c>0$ that we checked, that $W^d > W^u$ when the distribution is uniform. This welfare ranking would likely extend to skewed distributions, because we also found that, for $x=1$, W^d/W^u monotonically increased as t fell from 0 to -1. Thus, if the explanation for some markets being dropped is that discrimination is feasible but demand dispersion is simply too high, then complete discrimination would yield higher global welfare than uniform pricing.

An alternative explanation is that the threat of parallel imports is strong enough to force the seller close to uniform pricing, and under uniform pricing it is unprofitable to serve all markets. This corresponds to the case of 'moderate' skewness ($t^* < t < 0$, as opposed to $-1 < t < t^*$), in which case again the model predicts $W^d > W^u$.

4. Mixed systems

So far we have considered two polar regimes: uniform pricing, whereby all markets are charged the same price, and complete discrimination, whereby each market receives its own price. For policy purposes, one can envisage intermediate regimes in which the monopolist is permitted to engaged in some but not complete multimarket discrimination. The idea behind such regimes is to allow enough discrimination to ensure that all markets are served, but only that much discrimination, so as to limit the harmful output-misallocation effect.

Specifically, we consider *mixed systems*: markets are placed into groups, with all markets in a group required to receive the same price but prices are allowed to differ across groups. Mixed systems might be attainable through multilateral agreements that allow parallel imports within but not between groups of countries (see also section 5). Arbitrage within groups would then push toward uniform pricing without the need for cumbersome hands-on price regulation. (Section 5 discusses some practical problems with implementing such systems.) We analyze mixed systems for the basic model of section 3, though some of the results generalize.

Definition. A *countable partition* of $[1-x,\ 1+x]$ is a countable collection $\mathcal{P} = \{P_i\}_i$ of (Lebesgue measurable) sets such that (i) $\cup_i P_i = [1-x,\ 1+x]$ and (ii) for any i, j with $i \neq j$, $P_i \cap P_j = \emptyset$.

Definition. A *mixed system* is a countable partition of $[1-x,\ 1+x]$, $\mathcal{P} = \{P_i\}_i$, together with the requirement that, for each i, all markets in P_i are charged the same price. In mixed system $\mathcal{P}$, group P_i is said to be *fully served* if $\mu\{a \in P_i \mid q(p_i \mid a) > 0\} = \mu(P_i)$, where p_i denotes the monopolist's profit-maximizing price charged to group P_i, $q(p_i \mid a)$ denotes the demand of market a at price p_i, and μ denotes the Lebesgue measure.

Alternatively, group P_i is fully served at price p_i if almost all group members have choke prices above p_i. Lemma 1 provides intuition for several ensuing results and extensions. This lemma essentially covers all collections of linear demand functions (including the 'parallel' demands and 'rotating' demands with non-uniform distributions of markets discussed earlier).

Lemma 1. Suppose the monopolist has constant marginal cost $c \geqslant 0$. Consider the family of demand functions $Q(p|a) = \alpha(a) - \beta(a)p$, where $a \in [0, 1]$, and α and β are (measurable) functions from $[0, 1]$ to $\boldsymbol{R}_+$. Suppose index a is distributed according to the cumulative distribution function F. Let $\mathcal{P} = \{P_i\}_i$ be a mixed system for this family of demands. If group P_i is fully *served under $\mathcal{P}$, then the total output supplied to group P_i under mixed system $\mathcal{P}$ equals the sum to the outputs that group P_i's member markets would have received under complete discrimination.*

Malueg and Schwartz (1993) provide proofs of Lemma 1 and all subsequent results. Lemma 1 can be understood as follows. For uniform prices at which all markets in group i are served, group i's aggregate demand is simply the horizontal sum of the demands of all groups i's markets. Because this aggregate demand segment is linear, so is its corresponding marginal revenue curve, $MR_i(Q_i)$. Therefore, in a mixed system that fully serves group i, total output to group i is determined where $MR_i(Q_i)$ cuts the monopolist's marginal cost (here zero). Under discrimination, total output to these same markets is determined where the horizontal sum of all individual marginal revenue curves cuts marginal cost. But for linear demands, when all markets are served, the horizontal sum of marginal revenue curves coincides with $MR_i(Q_i)$; total output therefore is the same if the mixed system fully serves group i.[23]

4.1. No holes

We say that a mixed system has *no holes* if each member of the associated partition is an interval. We say that a mixed system contains *holes* if at least one member of the associated partition is not an interval. For example, some relatively high-demand markets might be grouped together with relatively low demands, with the medium-demand markets placed in a different group.

Consider the following mixed system with no holes, involving 'recursive' divisions into groups:

$$I_0 = \{0\}\ , \qquad I_1 = \left[\frac{t}{y}, t\right], \qquad I_n = \left[\frac{t}{y^n}, \frac{t}{y^{n-1}}\right),$$
$$\text{where } t = 1 + x \text{ and } n = 2, 3, \ldots . \tag{1}$$

Recall from Table 1 that for dispersion levels that induce the monopolist to drop some markets ($x > x^*$), the monopolist's profit-maximizing uniform price is t/y. Group I_1 thus consists of those markets that the monopolist

[23] The above logic shows that Lemma 1 would hold for any marginal cost function if the mixed system consisted of only one group. This is the well-known result mentioned in subsection 3.1: if uniform pricing fully serves a set of linear-demand markets, then total output is the same as under discrimination. In a mixed system with multiple groups, however, our added assumption of constant marginal cost is needed to allow the monopolist to optimize separately for each group.

would have served if constrained to charge a uniform price; faced with system (1), the monopolist clearly would charge this group t/y. For the other markets there are two possibilities, depending on the value of x: either the monopolist would set a second price that serves all these remaining markets, or it would set t/y^2 (t/y^2 because once types I_1 are excluded the new 'top' demand has vertical intercept t/y, and, by a simple rescaling argument, the profit-maximizing price to any *uniform* distribution of markets – if some markets would still be dropped at the lower price – is $1/y$ times the top intercept). In the latter case, the remaining markets not served at price t/y^2 would ultimately be served at the price(s) chosen for the lower group(s). The system (1) thus ensures that all markets are served. (Some intervals would contain no consumers, except in the case of full spread, $x = 1$.)

Proposition 2. (a) *If demand dispersion is small enough that the monopolist would serve all markets under uniform pricing ($x \leqslant x^*$), then uniform pricing and the mixed system* (1) *are identical and yield higher welfare than discrimination.* (b) *If demand dispersion is large enough that the monopolist would drop some markets under uniform pricing ($x > x^*$), then the mixed system* (1) *yields higher welfare than discrimination and is a Pareto improvement over uniform pricing.*

Part (a) of Proposition 2 is straightforward. The mixed system (1) boils down to uniform pricing for dispersion in the range $x \leqslant x^*$, because all markets then fall in the top group. Uniform pricing then dominates discrimination because with all markets served under uniform pricing, total output is the same under the two regimes (Lemma 1). And discrimination misallocates this output, whereas uniform pricing allocates it optimally.

Turn to part (b). Clearly the mixed system is a Pareto improvement over uniform pricing: those markets that would be served under uniform pricing are still served and at the same price, but the mixed system serves additional markets. Now compare the mixed system with discrimination. Consider the markets in any group i (that has strictly positive measure). Because all markets are served under the mixed system, total output to the markets in group i is the same as they would collectively receive under discrimination (Lemma 1). Because all markets in group i are charged a uniform price under the mixed system, misallocation of output among these markets is avoided; thus, the welfare to group i is strictly higher than under discrimination [Schwartz (1990)].

Observe that, if marginal cost is constant, then for any distribution (not just uniform) of markets with linear demands, one can recursively construct a mixed system similar to (1) that will satisfy Proposition 2. (That welfare would be higher than under complete discrimination follows by application of Lemma 1.) Moreover, for any distribution of (possibly non-linear) demands, as long as uniform pricing would not serve all markets, there exists a mixed system that Pareto dominates uniform pricing; e.g. form two groups, one consisting of those markets served under uniform pricing and the other consisting of all remaining markets. This observation is interesting for policy purposes because, as noted earlier, prevailing sentiment worldwide favors uniform pricing.

For our basic model with a uniform distribution of 'rotating-demand' markets, we can show the following.

Proposition 3. The recursive system (1) *maximizes welfare in the class of mixed systems with no holes.*

The intuition behind Proposition 3 is roughly that discrimination can increase welfare only if it leads the monopolist to increase total output. The proof shows that in designing groups, the top group should then be made as large as possible consistent with it being fully served. The lower boundary of the top group is therefore t/y. The same argument is repeated for the remaining markets (recall that the distribution of markets is uniform).

Example. To indicate the potential welfare gain from mixed systems, consider the case of maximum dispersion, $x = 1$ (the distribution of demand intercepts is uniform over $[0, 2]$). For this case, the recursive system (1) would require an infinite number of groups for the monopolist to choose to serve all markets. Let $W(n)$ denote welfare under this recursive system for groups I_1 through I_n and observe that $W(1)$ is just welfare under uniform pricing (the monopolist would charge $p = 2/y$ and drop all lower markets). It can be shown, for all $n \geqslant 1$, that

$$W(n) = \frac{y}{2(y+1)}\left(1 - \frac{1}{y^{2n}}\right) \quad \text{and} \quad \frac{W(n)}{W(1)} = \left(\frac{y^2}{y^2-1}\right)\left(1 - \frac{1}{y^{2n}}\right).$$

Because $y \approx 3.5128$, $W(\infty)/W(1) \approx 1.0882$, that is, the recursive system (1) increases welfare by approximately 8.82 percent relative to uniform pricing. (The gain relative to discrimination is about 3.84 percent.) This gain is significant. Even marginal-cost pricing would yield welfare of only 0.5, or a gain relative to uniform pricing of under 40 percent; and marginal-cost pricing is infeasible for many of the products experiencing parallel imports, given that these products often entail substantial fixed costs. Interestingly, the bulk of the gain from the mixed system is achieved with a small number of groups; with $n = 2$ the gain is already 8.10 percent, and with $n = 3$ it is 8.76 percent.

4.2. Holes

Proposition 3 considers only mixed systems without holes. Surprisingly, perhaps, welfare can be further increased by mixed systems that contain holes.

Proposition 4. If demand dispersion is large enough that the monopolist would drop markets under uniform pricing ($x > x^*$), then the recursive system (1) *is not optimal among the class of mixed systems that allows for holes.*

Starting with the optimal mixed system with no holes, (1), we show that moving a small set of markets (s_1 below) from its initial group to the group below it and creatng a 'hole' will increase welfare, due to improved allocation of total output.

Define the following sets of markets:

$$s_1 = (t - \varepsilon, t] , \qquad s_2 = \left[\frac{t}{y}, t - \varepsilon\right] , \qquad s_3 = \left[z, \frac{t}{y}\right) ,$$

where $\varepsilon > 0$ is 'small', $t = 1 + x$, and $z = \max\{1 - x,\ t/y^2\}$. Consider the partition of $[z,\ 1 + x]$ without holes, $\{A, B\}$, and the partition with holes, $\{A', B'\}$, where

$$A = s_1 \cup s_2 , \qquad B = s_3 \quad \text{and} \quad A' = s_2 , \qquad B' = s_1 \cup s_3 .$$

The sets A and B correspond to the top two groups of markets in system (1). There are two cases to consider, depending on the level of demand dispersion, x. If $1 - x > t/y^2$, the monopolist charges group A the price $p_A = t/y$ and group B a price $p_B \in (t/y^2,\ 1 - x)$. Group B is fully served with *slack* – even if the price to that group rose slightly, group B would still be fully served. If $1 - x \leq t/y^2$, the prices are $p_A = t/y$ and $p_B = t/y^2$; group B is fully served, but with *no slack*. Now compare $\{A, B\}$ with $\{A', B'\}$.

Here we address the case of slack, in which total output remains unchanged after the move of s_1; Malueg and Schwartz (1993) prove that welfare increases also in the case of no slack, where total output decreases after the move. Suppose the monopolist is faced instead with the partition with holes, $\{A', B'\}$; this simply involves moving the top markets s_1. The monopolist's prices would satisfy the inequalities

$$p_B < p_{B'} < p_{A'} < p_A ,$$

where the first and third inequalities hold because the highest demand set s_1 is included in B' and A but not in B and A', and the second because this transferred set s_1 is sufficiently 'small'. The key to understanding the change in welfare is to determine what happens to total output and to its allocation among various markets.

Consider first the allocation of output. Price to s_1 falls; hence, its consumption increases: $p_{B'} < p_A \Rightarrow Q_1' > Q_1$. Similarly for s_2: $p_{A'} < p_A \Rightarrow Q_2' > Q_2$. For s_3, price rises; hence, consumption falls: $p_{B'} > p_B \Rightarrow Q_3' < Q_3$. Thus, output is reallocated from markets in s_3 to markets in s_1 and s_2. Denote by MV_1, MV_2, and MV_3 the initial marginal valuations in markets 1, 2, and 3, respectively, and by MV_1', MV_2', and MV_3' the equilibrium marginal valuations after the move of s_1. The marginal valuations in markets s_1 and s_3 differed initially but are now equated: $MV_3 = p_B < p_A = MV_1$ but $MV_3' = p_{B'} = MV_1'$. The marginal valuation in s_2 still exceeds that in s_3 but the gap has narrowed: $MV_2' - MV_3' = p_{A'} - p_{B'} < p_A - p_B = MV_2 - MV_3$. Therefore, the allocation of total output following the move of s_1 is more efficient.

Now consider the level of total output. All markets are served before *and* after the move of s_1. This follows because originally all markets are served with slack ($p_B < 1 - x$), so all markets also will be served after the move ($p_{B'} < 1 - x$) provided the transferred set s_1 is sufficiently 'small'. By Lemma 1, we therefore know that total output remains unchanged. Given the improved allocation, the move of s_1 (introducing a 'hole') must increase welfare.

Note that if initially group b were served with *no slack* ($1 - x \leq t/y^2$ and $p_B = t/y^2$), then total output would fall after the move of s_1, because $p_{B'} > p_B = t/y^2$. The proof that such a small move nevertheless would increase welfare therefore relies on the gain from reallocation of output outweighing the loss from the output reduction. Observe, however, that for any x, $x^* < x < 1$, only a finite number of the groups in (1) contain a positive measure of markets, and generically the bottom group will be served with slack. Shifting a small group of consumers from the second-from-the-bottom group to the bottom group will then increase welfare, by the same argument as in the example above involving the top groups I_1 and I_2. Moreover, that reasoning holds for any family of linear demands (not just 'rotating') and for any mixed system with no holes that fully serves with slack the lower of the two adjacent groups (because the reasoning relies only on Lemma 1, which holds for all linear demands).

Proposition 4, though not characterizing an optimal system, shows that the substantial welfare gain (relative to uniform pricing) achieved by the no-holes system (1) is a lower bound on the gains achievable by mixed systems.

5. Conclusion

This paper examined whether the strong policy support that parallel imports enjoy internationally is justified from the standpoint of world welfare, even assuming that parallel imports are caused entirely by international price discrimination, not free riding. An advantage of allowing at least some international price discrimination is that additional countries would likely be served. If parallel imports are prevented, at least between certain groups of countries, firms could offer lower prices to lower-demand (more elastic) countries without fear of the products resurfacing in high-price markets. Absent such (partial) segmentation, firms may well choose relatively high uniform prices, at which many low-demand countries are likely to go unserved.[24]

In a stylized model we compared world welfare under uniform pricing with that under complete discrimination (a different price in each country) and under mixed systems, which allows discrimination between groups of countries but not within a group. Whether mixed systems are feasible will depend on the possibility of tracing the original source of the goods. This problem of re-exportation, or 'trade defection' [Corden (1984)], is well known in the literature on free-trade areas. Its severity will vary across countries and industries. Where mixed systems are not feasible, complete discrimination may still be possible – because it requires curbing unauthorized imports regardless of origin. Our welfare comparison between complete discrimination and uniform pricing is relevant for such cases.

Comparing complete discrimination with uniform pricing, we found that when demand dispersion is large enough, welfare is higher under discrimination. The beneficial effect of higher output under discrimination from continuing to serve low-demand markets outweighs the misallocation effect of discrimination.

The welfare gains relative to uniform pricing can be higher still if suppliers are granted some but not complete discretion to price discriminate. We constructed a mixed system, (1), that yields higher welfare than complete discrimination and is a Pareto improvement over uniform pricing (whereas complete discrimination can yield higher welfare than uniform pricing but harms the high-demand markets). Interestingly, a small number of groups suffices to achieve the bulk of the welfare gain. Thus, one could imagine assigning countries into a few blocks based on per capita income – e.g. low, middle, and high – and allowing discrimination only between blocks. System

[24] A famous example of such behavior is the decision by Distillers Company Limited to drop its premium Red Label brand from the U.K. market in response to 1978 decisions by the European Commission and Court of Justice that Distillers could not discourage parallel exports from the lower-priced U.K. market to the Continent [Hawk (1991)]. Also in response to the EC's prohibitions of curbs on parallel trade, manufacturers are reportedly curtailing supplies of pharmaceuticals to low-price countries [REMIT (1992)]. Harder to measure, but potentially more important, are decisions by manufacturers not to enter certain low-price markets in the first place, for fear that parallel exports from these markets would undercut high prices in other countries.

(1) was shown to yield highest welfare in the class of systems with no holes. Surprisingly, perhaps, introducing holes – by grouping together some markets with low and high demands while grouping separately some intermediate ones – can increase welfare still further.[25]

Discrimination, however, affects not only total welfare but also *distribution* between countries (as well as within countries – producers vs. consumers). Even if complete discrimination increases world welfare relative to uniform pricing, it is likely to harm consumers in high-price countries [though it need not if marginal cost is decreasing; cf. Hausman and MacKie-Mason (1988)]. But manufacturers of products prone to parallel imports also are predominantly from richer (more industrialized) countries, and those manufacturers would gain from discrimination. Thus, permitting complete international price discrimination need not systematically reduce the national welfare of industrialized countries. Nevertheless, some rich countries are likely to lose: countries that are relatively under-represented in the manufacture of goods involved in parallel trade would capture little of the profit gain from discrimination but would pay relatively high discriminatory prices. Such countries, of which the United States likely is one (with the exception of pharmaceuticals, the United States is a relatively small producer of prominent parallel-traded goods, such as luxury cars, consumer electronics, and perfumes – see footnote 2), would lose from unilaterally curbing parallel imports so as to permit complete discrimination.

Rather than complete discrimination, countries may more easily agree on mixed systems such as (1), which mitigate the distributional impact of discrimination by permitting lower prices only to markets that otherwise would not be served. In our model, system (1) in fact yields a Pareto improvement over uniform pricing. In practice, of course, things will be messier. For example, the appropriate sorting of countries into groups will vary by industry (as national demands vary across products), but a country's policies and laws toward parallel imports may have to be relatively uniform across industries. Given such inflexibility, some rich countries might lose even from a regime that permits only limited discrimination between groups.

[25] As a practical matter, there may be political difficulties in implementing systems with 'holes', due to the distributional issues raised. Such systems group together high-demand countries with low-demand ones, and the welfare gain arises from reshuffling output from the latter to the former – for instance, putting the United States in a group with the developing countries in order to 'pull down' the price that the monopolist would charge the United States. In the absence of adequate compensation mechanisms (discussed further shortly), such groupings may be politically awkward; grouping together countries with similar demands, as in our recursive system, may be more palatable. On the other hand, the finding that holes can increase welfare suggests that if holes are introduced inadvertently, because a no-holes system such as (1) is not feasible, say due to geographical considerations, the welfare advantage of limited discrimination over uniform pricing need not be eroded.

If so, such countries could perhaps be compensated by poorer countries that stand to gain from discrimination in their favor. Such compensation need not be direct, but could instead take the form of offsetting concessions in multilateral negotiations over a range of trade issues; for example, in exchange for rich countries curbing parallel imports, poor countries might offer stronger protection for intellectual property rights.[26]

The basic idea that some international price discrimination might be beneficial has implications for policy toward parallel imports both within the EC and between less developed countries (LDCs) and industrialized ones (North–South trade). Within the EC demand conditions still vary considerably, because of differences in incomes and in national policies such as price controls and the strength of IPRs (hence the availability of counterfeit substitutes). Insisting as the EC does on unencumbered parallel imports that arbitrage national price differences may well lead manufacturers to sharply curtail sales to certain countries. This prospect suggests that *some* price discrimination should be permitted. As a rough cut, low-income countries such as Greece, Ireland, and Portugal might be grouped in a block separate from the richer countries, with parallel imports allowed within but not between blocks. The EC's unwavering support for parallel imports is therefore questionable.[27]

[26] Indeed, a U.S. proposal (May 1990) to the talks on Trade Related Aspects of Intellectual Property Rights (TRIPs) of the Uruguay Round of the GATT addressed, inter alia, traditional issues concerning the strength of IPRs (forced licensing, penalties for infringement, etc.) as well as the right of holders of copyrights and patents to prevent parallel imports. For reasons noted earlier, however, absent such compensation, some rich countries likely will resist curbing parallel imports.

[27] Simulation studies [e.g. Smith and Venables (1988), Mercenier (1994)] often find that *all* EC countries would benefit from moving to uniform pricing (by completing the market integration). These authors assume imperfectly competitive firms rather than monopoly, and attribute their results to a decrease in the average degree of market power when moving to uniform pricing. The rationale given by Smith and Venables (1988, p. 1522) is that a firm's market share is typically larger at home than in foreign markets (say due to preferences for home products); integration (uniform pricing) reduces market power by reducing concentration, since 'the relevant measure of concentration is for the EC as a whole'.

This argument, however, is incomplete. The aggregate EC shares mask a persisting preference for home products and therefore need not accurately proxy the change in market power. Consider a symmetric model with two countries, each with a single firm. Given preference for home products, with segmented markets each firm has a larger share of its home market and charges a higher price (at least under Smith and Venables' CES preferences). We conjecture that requiring a uniform price would lead each firm to lower its domestic price but raise its foreign price; moreover, the uniform price may well exceed the average discriminatory price, because the domestic market is relatively more important. Thus, requiring uniform pricing in oligopoly need not create a systematic tendency toward lower prices. Given Holmes' (1989) results that the comparison is generally ambiguous, we believe that the simulation findings above result from special assumptions, e.g. about the particular nature of product differentiation. For more on this point, see Haaland and Wooton (1992).

Potentially larger gains from international price discrimination could arise by permitting suppliers to offer lower prices to LDCs, whose demand elasticities are likely to be much higher than in industrialized countries due to vastly lower per capita incomes. There are signs that suppliers indeed would price favorably to LDCs. Schut and Van Bergeijk (1986) found that pharmaceuticals's prices internationally varied widely and were strongly positively correlated with per capital incomes. Their finding is corroborated by anecdotal evidence [see Drug Diversion (1985)] that some U.S. pharmaceutical companies sold their products for as little as one-quarter the domestic price when destined for export to LDCs (although the 'exports' were frequently diverted back to the U.S. market). Discounts of the same order are offered by some economics journals (e.g. *Econometrica*) to subscribers in LDCs.[28]

Offering lower prices to LDCs, in addition to its inherent advantage of 'opening markets', may also soften LDCs' reluctance to grant stronger protection of IPRs. Many LDCs fear that most of their consumers could not afford the high prices that would be charged for products embodying intellectual property if they did grant stronger IPRs [Diwan and Rodrik (1991)]. Preventing parallel trade between LDCs and richer countries, however, would help make it possible to offer selectively lower prices in LDCs.

In conclusion, our analysis casts doubt on the view that world welfare would be enhanced by encouraging unrestricted parallel imports in order to undermine price discrimination. Importantly, we have focused on price discrimination as the sole cause of parallel imports, abstracting entirely from other factors that exclusive territories are designed to combat, such as free riding and consumer confusion (when parallel imports differ from the authorized products targeted to the local market). Those other roles of exclusive territories are often efficient. Moreover, these efficiencies are likely to be at least as great in the international context as within countries, given that substantial country-specific investments are often required to introduce new products and that such investments are often best elicited by awarding sole-import distributorships. From the standpoint of global welfare, the case against parallel imports is correspondingly strengthened.

[28] To be sure, poor countries sometimes pay higher prices than rich ones. Such 'reversals' are partly attributable, however, to various government policies in LDCs that reduce competition in their markets, e.g. by discouraging foreign entry. Pricing reversals may also arise from choices by suppliers to target their products to the rich rather than the mass market in LDCs; but such choices may themselves be dictated by the inability to offer lower prices selectivity to LDCs if parallel exports from LDCs cannot be controlled. Thus, pricing reversals do not refute the argument that price discrimination, if feasible, likely would favor poorer countries.

References

Brander, J. and P. Krugman, 1983, A reciprocal dumping model of international trade, Journal of International Economics 15, 313–321.

Business Week, 1985, The $7 billion gray market: Where it stops, nobody knows, 15 April.

Business Week, 1988, A red-letter day for gray marketeers, 13 June.

Chard, J.S. and C.J. Mellor, 1989, Intellectual property rights and parallel imports, World Economy 12, 69–83.

Corden, W.M., 1984, The normative theory of international trade, in: Ronald W. Jones and Peter B. Kenen, eds., Handbook of international economics, Vol. 1 (North-Holland, Amsterdam).

COPIAT, 1986 Comments of the Coalition to Preserve the Integrity of American Trademarks (COPIAT submitted to U.S. Customs Service, 20 October 1986, in the Matter of Importations Bearing Recorded U.S. Trademarks, Solicitation of Public Comments on Gray Market Policy Options, 51 Fed. Reg. 22005, 17 June.

DeMuth C., 1990, Statement Before the Subcommittee on Patents, Copyrights and Trademarks Committee on the Judiciary, Hearing on S. 626, Senate Committee on the Judiciary, 101st Cong., 2nd Sess., 4 April.

Diwan, I. and D. Rodrik, 1991, Patents, appropriate technology, and north–south trade, Journal of International Economics 30, 27–47.

Dornbusch, R., 1987, Exchange rates and prices, American Economic Review 77, 93–106.

Drug Diversion, 1985, A staff report by the Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, U.S. House of Representatives, June 1985 (USGPO, Washington, DC).

Drug Store News, 1993, One if by land, two if by sea; Offshore mail order dispensaries dumping worries local drugstores, 26 April.

Feenstra, R.C., 1989, Symmetrical pass-through of tariffs and exchange rates under imperfect competition: An empirical test, Journal of International Economics 27, 25–45.

Financial Times, 1992, Sir Leon cracks the whip, 25 June.

Froot, K.A. and P.A. Klemperer, 1989, Exchange rate pass-through when market share matters, American Economic Review 4, 637–654.

Haaland, J.I. and I. Wooton, 1992, Market integration, competition, and welfare, in: A. Winters, ed., Trade flows and trade policy after '1992' (Cambridge University Press, Glasgow).

Hausman, J.A. and J. MacKie-Mason, 1988, Price discrimination and patent policy, RAND Journal of Economics 19, 253–265.

Hawk, B.E., 1991, United States, Common Market and international antitrust: A comparative guide, 2nd edn. (Prentice-Hall Law & Business, Englewood Cliffs, NJ).

Hearing On S. 626, 1990. Legislation to Amend the Lanham Trademark Act Regarding Gray Market Goods/ Hearing on S. 626 Before the Subcommittee on Patents, Copyrights and Trademarks, Senate Committee on the Judiciary, 101st Cong., 2nd Sess., 4 April.

Hilke, J.C., 1988, Free-trading or free-riding: An examination of the theories and available empirical evidence on gray market imports, World Competition 32, 75–91.

Holmes, T.J., 1989, The effects of third-degree price discrimination in oligopoly, American Economic Review 79, 244–250.

JFTC, 1991, The Antimonopoly Act Guidelines Concerning Distribution Systems and Business Practices, Executive Office, Japanese Fair Trade Commission, Tokyo, 11 July.

Kasa, K., 1992, Adjustment costs and pricing to market: Theory and evidence, Journal of International Economics 32, 1–30.

Knetter, M.M., 1989, Price discrimination by U.S. and German exporters, American Economic Review 79, 198–210.

Krugman, P., 1987, Pricing to market when the exchange rate changes, in: S.W. Arndt and J.D. Richardson, eds., Real-financial linkages among open economies (MIT Press, Cambridge, MA).

Lexecon Inc., 1985, The economics of gray-market imports (Washington DC).

Malueg, D.A. and M. Schwartz, 1993, Parallel imports, demand dispersion, and international price discrimination, Discussion Paper 93-17, Department of Economics, Georgetown University.

Marston, R.C., 1990, Pricing to market in Japanese manufacturing, Journal of International Economics 29, 217–236.

Mercenier, J., 1994, Can '1992' reduce unemployment in Europe? On welfare and employment effects of Europe's move to a single market, Journal of Policy Modeling, forthcoming.

Mertens, Y. and V. Ginsburgh, 1985, Product differentiation and price discrimination in the European Community: The case of automobiles, Journal of Industrial Economics 34, 151–166.

Palia, A.P. and C.F. Keown, 1991, Combating parallel importing: Views of U.S. exporters to the Asia-Pacific region, International Marketing Review 8, 47–56.

Prices Surveillance Authority 1989, Inquiry into book prices: Final report, Report No. 25, Sydney, Australia.

REMIT Consultants, London, 1992, Impediments to parallel trade in pharmaceuticals within the European Community, Final Report Prepared for DGIV of the European Commission (Office for Official Publications of the European Communities, Luxembourg).

Schut, F.T. and P.A.G. Van Bergeijk, 1986, International price discrimination: The pharmaceutical industry, World Development 14, 1141–1150.

Schwartz, M., 1990, Third-degree price discrimination and output: Generalizing a welfare result, American Economic Review 80, 1259–1262.

Smith, A. and A.J. Venables, 1988, Completing the internal market in the European Community: Some industry simulations, European Economic Review 32, 1501–1525.

Tarr, D.G., 1985, An economic analysis of gray market imports, mimeo, Federal Trade Commission, September.

Tirole, J., 1988, The theory of industrial organization (MIT Press, Cambridge, MA).

U.S. Bureau of the Census, 1992, Statistical abstract of the United States: 1992, 112th edition (Washington, DC).

World Bank, 1992, World development report 1992 (Oxford University Press, New York, NY).