

Parallel Imports and Price Controls*

by

Gene M. Grossman
Princeton University

and

Edwin L.-C. Lai
Federal Reserve Bank of Dallas

Revised: August 2007

Abstract

Price controls create opportunities for international arbitrage. Many have argued that such arbitrage, if tolerated, will undermine intellectual property rights and dull the incentives for investment in research-intensive industries such as pharmaceuticals. We challenge this orthodox view and show, to the contrary, that the pace of innovation often is faster in a world with international exhaustion of intellectual property rights than in one with national exhaustion. The key to our conclusion is to recognize that governments will make different choices of price controls when parallel imports are allowed by their trade partners than they will when they are not.

JEL Classification: O34, F13

Keywords: pharmaceuticals, reimportation, intellectual property, TRIPs

*The authors are grateful to the U.S. National Science Foundation (grant nos. SES 0211748 and SES 0451712), the Hong Kong Research Grants Council (project no. CityU 1476/05H), the Department of Economics and Finance at City University of Hong Kong, and the International Economics Section at Princeton University for financial support. We thank Pol Antràs, Joan Costa-Font, Joseph Harrington, Stefan Szymanski, Tommaso Valletti and two anonymous referees for comments on an earlier draft. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or any other organization.

1 Introduction

Parallel trade occurs when a good protected by a patent, copyright, or trademark, having been legally purchased in one country, is exported to another without the authorization of the local owner of the intellectual property rights in the importing market (see Maskus, 2000b). Parallel trade represents a form of arbitrage whereby a legitimate product is shipped from the market intended by the intellectual property rights holder to another where it commands a higher price. Like other forms of arbitrage, parallel imports (also known as “grey-market imports” or “reimports”) respond to international price differences that result from retailer price discrimination, vertical pricing restraints, or national differences in government price controls.

The propriety of parallel trade is a matter of intense policy debate in a number of countries and in the World Trade Organization (WTO). At present, WTO provisions allow member countries to establish their own rules for the “exhaustion” of intellectual property rights (IPR).¹ If a country opts for *national exhaustion* of IPR, a rights holder there may exclude parallel imports, because intellectual property rights continue until such a time as a protected product is first sold in that market. If a country instead chooses *international exhaustion* of IPR, parallel imports cannot be blocked, because the rights of the patent, copyright or trademark holder expire when a protected product is sold anywhere in the world. The United States practices national exhaustion for patents and copyrights, but permits parallel imports of trademarked goods unless the trademark owner can show that the imports are of different quality from goods sold locally or otherwise might cause confusion for consumers. The European Union provides for regional exhaustion of IPR whereby goods circulate freely within the trading bloc but parallel imports are banned from non-member countries. Japanese commercial law permits parallel imports except when such trade is explicitly excluded by contract provisions or when the original sale is made subject to foreign price controls.²

Public debate about parallel imports has been especially heated in the area of prescription drugs. In the United States, where consumers, public health officials, and politicians have become increasingly concerned about the high and rising cost of medicine, bills that would introduce international exhaustion of patent rights for prescription drugs have been intro-

¹Article 6 of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) states that:

For the purposes of dispute settlement under this Agreement, subject to the provisions of Articles 3 and 4, nothing in this Agreement shall be used to address the issue of the exhaustion of intellectual property rights.

²Different national regimes of IPR exhaustion are described in more detail in Maskus (2000a, 2000b) and Arfwedson (2004).

duced in one or both houses of Congress in each of the last three sessions. In fact, in 2000, Congress passed a bill to permit reimportation of medicines. Although President Clinton signed the bill into law, his administration ultimately declined to implement it, citing concerns about consumer safety. New legislation was introduced to Congress in 2004 that would have forced deregulation of parallel imports of pharmaceuticals. In each case, the impetus for Congressional action came from public pressures to step up imports from Canada, where regulations and price controls have generated prices for prescription drugs significantly lower than those across the border.³ Despite the continuing legal impediments to parallel imports, reimportation has been a growing source of pharmaceuticals supply in the United States due to increased personal trafficking and the proliferation of Internet purchases. By one estimate, parallel imports of prescription drugs from Canada amounted to \$1.1 billion in 2004, or about 0.5% of the U.S. market (Cambridge Pharma Consultancy, 2004).

Parallel trade in pharmaceuticals features prominently in Europe as well, where it is the source of ongoing controversy. Differences in price regulations have resulted in significant variation in pharmaceutical prices across member countries of the European Union and the European Free Trade Association (Arfwedson, 2004; Kanavos and Costa-Font, 2005). Kanavos and Costa-Font, for example, report price differences for many products and country pairs of between 100% and 300%. Prices are highest in the countries with free or relatively free prices, which are Germany, the United Kingdom, Sweden, the Netherlands, and Denmark. These countries all have witnessed substantial growth in parallel importing, with reimports accounting for between seven and twenty percent of their national expenditures on prescription drugs in 2002. The parallel exporters include Spain, France, Greece, Italy, and Portugal, all of which are countries with controlled prices or explicit price caps (Ganslandt and Maskus, 2004).

Opponents of parallel trade—of which there are many—are concerned that such trade undermines manufacturers' intellectual property rights.⁴ The prevailing wisdom, expressed for example by Barfield and Groombridge (1998,1999), Chard and Mellor (1989), Danzon (1997, 1998), and Danzon and Towse (2003), is that parallel trade impedes the ability of research-intensive firms such as those in the pharmaceuticals industry to reap an adequate return on their investments in new technologies. Their arguments seem to find support in more formal analyses, which offer only a few minor caveats. For example, Li and Maskus

³Maskus (2001) compares prices for similar doses of 18 drugs and finds lower prices in Canada than in the United States in 17 cases. The average price for these drugs in Canada was 63 percent of that in the United States.

⁴Aside from any potential dynamic effects, parallel trade that undermines third-degree price discrimination can reduce world welfare by reallocating sales from markets with low demand elasticities to markets with high demand elasticities; see Malueg and Schwartz (1994).

(2006) develop a model in which parallel trade creates competition in the home market between a manufacturer and its own distributor in a foreign market. They find that the distortions associated with parallel imports reduce the manufacturer's incentive to invest in cost reducing innovation at a prior stage. Szymanski and Valletti (2005) model an industry with vertical product differentiation and examine how parallel trade affects a firm's decision of whether to supply high quality products. They conclude that such trade diminishes the firm's incentive to invest in quality and may reduce welfare in both countries. In a related model, Valletti (2006) shows that parallel trade reduces investment when differential pricing is based purely on differences in price elasticities, but increases investment when differential pricing results from idiosyncratic costs of serving the two markets (that the manufacturer must bear, but an arbitrager can avoid). Finally, Rey (2003) examines parallel importing that may result from differential price regulation, as is most germane in the case of pharmaceuticals. He argues that such trade impedes a country's ability to accept high local prices in order to promote R&D when its trade partner prefers to set lower prices. As a result, world investment in technology is lower in his model when parallel trade is allowed than when it is banned.

In this paper, we challenge the prevailing wisdom for the case (such as in pharmaceuticals) in which parallel trade is induced by different national price controls. We argue that the existing policy discussions and formal modeling overlook an important effect of national policy regarding the exhaustion of IPR. Whereas the papers in the literature that discuss parallel imports of pharmaceuticals in the presence of price regulation compare incentives for R&D in scenarios with and without parallel imports *for given levels of controlled prices*, we emphasize that a government will face different incentives in regulating prices in the presence or absence of arbitrage possibilities. This is so for two reasons. First, the admissibility of parallel trade introduces the possibility that a manufacturer will eschew low-price sales in the foreign market in order to mitigate or avoid re-importation. When arbitrage is impossible, the manufacturer is willing to export at any price above marginal production cost. But when the potential for arbitrage exists, the manufacturer may earn higher profits by selling only in the unregulated (or high-price) market than by serving both markets at the lower, foreign-controlled price. Accordingly, a switch from a regime of national exhaustion to one of international exhaustion can induce an increase in the controlled price as the foreign government seeks to ensure that its consumers are adequately served.

Second, the admissibility of parallel trade mitigates the opportunity for one government to free-ride on the protection of IPR granted by another. As we have shown in Grossman and Lai (2004), national policies to protect IPR often are strategic substitutes in a two-country (or many-country) policy-setting game; the greater is the protection afforded by one country, the less will be the optimal level of protection perceived by the other. In a world without

parallel trade in which one government allows manufacturers to price freely in its market (or enforces a price ceiling at a relatively high level), the other government can hold the line on prices so as to benefit its local consumers and nonetheless enjoy the fruits of relatively high rates of innovation. But when arbitrage can occur, a low price cap in one country dulls incentives for investment worldwide. The government of a less-innovative country faces a different trade-off in setting its regulated prices in the presence or absence of arbitrage. As we will show, the deregulation of parallel imports in a highly-innovative economy always induces the government of its less-innovative trade partner to loosen its price controls.⁵ In our analysis, the presumptive outcome is one in which deregulation of parallel imports generates both an increase in consumer surplus in the innovative country and an *increase* in the world pace of innovation. Thus, the more innovative country may face no trade-off at all between static and dynamic gains in its choice of regime for exhaustion of IPR. We find that typically the more innovative country gains from parallel trade, while the less innovative country loses.⁶

The starting point for our analysis is the two-country model of ongoing innovation and trade that we developed in Grossman and Lai (2004) to study incentives that governments have to protect IPR in an open, world economy. In this model, firms devote resources to inventing horizontally differentiated goods with finite product lives. The greater is the global protection of IPR, as reflected in the duration of patents and the enforcement rate for live patents, the greater is the incentive for product development. The model yields a steady-state equilibrium in which the measure of differentiated products is constant. This measure responds to the policy regime. Here, we take the global regime for protection of patent rights as given, but introduce the possibility of price controls. That is, we assume that new pharmaceutical products receive similar patent protection in all countries (as indeed is dictated by the terms of the TRIPS agreement), but that existing international agreements do not prevent a country from regulating domestic prices of prescription drugs. We compare prices, profits and rates of innovation in a regime in which the high-price country allows parallel imports with those that arise when such trade is disallowed. We do so first under the assumption that the innovative “North” allows its firms to price freely and that no innovation takes place in “South.”⁷ Subsequently, we show that the main insights carry over to the case

⁵Indeed, Kanavos and Costa-Font (2005, p.766) cite policy changes of this sort in some European countries in response to the growth of intra-European parallel trade in pharmaceutical goods. For example, they note that Portugal and Italy have in recent years moved their regulated prices for reimbursed medicines closer to average European prices, while France too has tended to allow freer pricing of highly innovative products.

⁶Pecorino (2002) makes a related though different point in a static model of sales and exports by a single firm. In his model, export prices are determined as a Nash bargain between the manufacturer and the foreign government. The possibility of reimportation strengthens the bargaining position of the manufacturer by lending credibility to the threat that he will not serve the foreign market. Accordingly, the monopolist’s profits can be higher when parallel imports are allowed than when they are not.

⁷Whereas the geographic designations seem apt for Europe, the North American reader should not be

in which both countries regulate prices in a Nash equilibrium policy game, and also to the case in which innovation occurs in both countries provided that North conducts sufficiently more R&D than South.

We will argue that by allowing parallel imports, the North dulls the incentives in the South for free-riding on its protection of IPR. This argument does not require that the South be a single, large economy, as we initially assume. In fact, our results extend readily to a setting with several or many countries in the South. The size of a typical country in the South enters our analysis in several ways. If such a country is small (for example, because there are many such countries), there will be fewer consumers there to benefit from faster world innovation. But there will also be fewer consumers to be harmed by high local prices. These two effects of country size offset one another in our model, and so they do not affect a government's incentives in setting its price ceilings. The size of a country in the South can also affect the leverage that its governments has over world innovation and the risk it faces that its local market will not be served. The policies of a small country have little effect on R&D investment in a world without parallel trade, because each firm captures only a tiny share of its profits in such a market. Therefore, in a regime of national exhaustion of IPR in the North, the free-rider problem quickly grows more severe as the number of countries in the South expands. In contrast, when arbitrage is possible, a country that strictly controls prices impinges upon profit opportunities not only in its own small market, but also in other, larger markets to which the goods can be shipped. Accordingly, even a small country has considerable leverage over world innovation in such a setting. Note too that innovating firms may well be tempted to cut off supply from a very small market when goods can be transhipped from there to the North, but such firms have no reason to withhold supply from small markets in the absence of parallel trade. For these reasons, we find that the North's preference for international exhaustion of IPR extends to a world with multiple symmetric Southern countries and even to a world in which the typical such country is vanishingly small.

The remainder of the paper is organized as follows. In the next section, we adapt the Grossman-Lai model of ongoing innovation to allow for price controls. We develop the model under the assumption that all innovation takes place in the North and that prices are regulated only in the South. In Section 3, we derive and discuss our main results, which compare prices, profits and rates of innovation under alternative regimes for exhaustion of IPR in which North does and does not allow parallel imports. Section 4 addresses circumstances in which North also may impose price controls and in which firms in South have the capacity to innovate and Section 5 extends the analysis to a world with several or many countries in the South. The final section summarizes our findings.

confused by the fact that, in the context of our model, the United States is "North" and Canada is "South."

2 A Model of Ongoing Innovation

We consider an industry such as pharmaceuticals in which firms innovate by introducing new varieties of a differentiated product, all of which have finite economic lives. Our approach builds on Grossman and Lai (2004).

The world economy has two sectors, one that produces a homogeneous good and the other that produces a continuum of differentiated products. The designs for differentiated products are the outgrowth of investments in R&D. Each new product has a useful economic life of length $\bar{\tau}$, which means that it can provide utility to consumers for a period of that duration. When $\bar{\tau}$ years have elapsed from the time of invention, the product ceases to be of economic value.

There are two countries, North and South. Innovative products are consumed in both countries whereas (for now) all innovation takes place in North. In country j , $j = N$ (for North) or $j = S$ (for South) there are M_j consumers who demand the differentiated products of the innovating industry. The typical such consumer maximizes an intertemporal utility function of the form

$$U_j(t) = \int_t^\infty u_j(z) e^{-\rho(z-t)} dz \quad (1)$$

where

$$u_j(z) = y_j(z) + \int_0^{n(z)} h[x_j(i, z)] di, \quad (2)$$

$y_j(z)$ is consumption of the homogeneous good at time z , $x_j(i, z)$ is consumption of the differentiated product i at time z , and $n(z)$ is the number (measure) of differentiated products invented before time z that remains economically viable at z . As in Grossman and Lai (2004), we will refer to M_j as the “size of the market” in country j .

A consumer with the preferences represented by (2) maximizes utility by purchasing differentiated product i at time z up to the point where its marginal utility, $h'[x_j(i, z)]$ is equal to the local price of the good. After the consumer makes all of his optimal purchases of the economically viable differentiated products available at time z , he devotes residual spending (which we take to be strictly positive) to the homogeneous good. With spending on the homogeneous good assumed positive at every moment, (1) implies that the interest rate must be constant and equal to ρ . Then each household attains indirect utility equal to the present discounted value of lifetime spending plus the present discounted value of consumer surplus captured on purchases of differentiated products.

New varieties of the differentiated product result from a research activity that requires labor and a sector-specific factor that we call “research capital.” We take $\phi_j(z) = \frac{1}{\beta} K_j^{1-\beta} L_{Rj}(z)^\beta$, where $\phi_j(z)$ is the flow of new products developed in country j at time z , K_j is the (fixed)

stock of research capital in country j , and $L_{Rj}(z)$ is the (effective) labor devoted to research in country j at time z .⁸ To capture a setting in which research capacity is highly concentrated in one country, we assume (for now) that $K_N > 0$ while $K_S = 0$. Note that the time derivative of $n(z)$ is $\dot{n}(z) = \phi_N(z) - \phi_N(z - \bar{\tau})$, because products invented at time $z - \bar{\tau}$ become obsolete at time z . In steady state, ϕ_N is constant and so $\dot{n}(z) = 0$.

Our earlier paper focused on the determination of patent lengths and enforcement rates in non-cooperative and cooperative international regimes of intellectual property protection. Here, we take the patent regime as given and assume that patent lengths have been harmonized by some international agreement such as TRIPS. Every inventor is awarded a patent of length $\tau < \bar{\tau}$ in both countries. We assume that patents are perfectly enforced.⁹ By taking the terms of patent protection as given, we can focus most sharply on price controls and parallel trade as instruments of innovation policy.

Once a variety has been invented, it can be produced in North with c units of labor per unit of output. We take North's wage as numeraire, so c also represents the marginal cost of production there. A firm that holds a live patent for an economically viable good has the sole right to produce that good and to make the first sale at home or abroad. Whether it can also prevent reimportation of goods that it initially sells in South depends on the prevailing regime for exhaustion of IPR. If North adopts national exhaustion, the patent holder can prevent parallel imports of units it sells to consumers (or distributors) in South. If North instead chooses international exhaustion, parallel imports cannot be barred. For simplicity, we ignore transportation and marketing costs. Therefore, price arbitrage is perfect in a regime that permits parallel trade.

Firms that hold live patents set prices for (initial) sales in each market. They do so after North has decided its regime for exhaustion of IPR and after the two governments have announced their pricing policies, if any. If the price that a firm charges for sales in a country exceeds the legal price ceiling there, no distributor or consumer will be able to purchase its product. If North adopts international exhaustion and a patent holder asks a higher price there than that for sales in South, it will lose the Northern market to parallel traders. Accordingly, a patent holder must charge a price for sales in each country no greater than the local price ceiling or its sales there will be zero, and it must charge identical prices in the two markets under international exhaustion or its sales in the higher-priced market

⁸In Grossman and Lai (2004), we allow for a constant elasticity of substitution between labor and research capital that need not be equal to one. We find that most qualitative results do not depend on the size of this elasticity, as long as it does not exceed two.

⁹In Grossman and Lai (2004), where we allowed for imperfect enforcement, we showed how one could construct a single measure of strength of IPR protection that reflects both duration of patents and rates of enforcement. Here, we ignore the complications associated with imperfect enforcement for simplicity.

will be zero.

We denote by $\pi(p)$ the profits that a patent holder makes per consumer when it charges the price p in some market and serves all local demand at that price. Note that consumers in both countries have identical demands, so these profits are the same (as a function of price) for sales in North and South. Similarly, $C(p)$ is the surplus that a consumer in either country enjoys when the differentiated product i is available at price p . Let $C_C = C(c)$ be the surplus per consumer when a differentiated product is available at the competitive price c , and let $C_M = C(p_M)$ be the surplus per consumer when the patent holder charges the unconstrained monopoly price p_M . As usual, the monopoly price is such that the markup over marginal cost as a fraction of price is equal to the inverse demand elasticity.¹⁰

Whenever a patent expires, the formerly-patented good can be produced by firms anywhere in the world at a unit cost of c . That is, we assume that any wage gap that exists between North and South reflects a difference in the productivity of labor in all of its uses. This, and the tradability of the numeraire good, allows us to measure Southern labor in efficiency units, with the wage per efficiency unit in South equal to that in North. Then the production cost for unpatented goods is the same in both places.¹¹

We describe now the steady-state equilibrium for a given policy regime. In a steady state, the measure of differentiated products is constant. The products invented at time z replace the products that become obsolete at z , which are those that were invented at $z - \bar{\tau}$. The total number of viable products is $n = \phi_N \bar{\tau}$, with $n_M = \phi_N \tau$ of these under patent protection at any moment, and the remaining $n_C = \phi_N (\bar{\tau} - \tau)$ competitively priced. The patented goods carry a price p_N in North and p_S in South. These prices will reflect the policy regime in force, including the height of any price ceilings and whether North allows parallel imports. A Northern monopolist earns a flow of profits $M_N \pi(p_N) + M_S \pi(p_S)$ for a period of length τ , the present discounted value of which is

$$v = [M_N \pi(p_N) + M_S \pi(p_S)] T , \quad (3)$$

where $T \equiv (1 - e^{-\rho\tau}) / \rho$ is the present value of a flow of one dollar from time zero to time τ . Firms in North deploy labor in R&D up to the point where the marginal value product

¹⁰We assume that the monopoly price is finite, which requires a demand elasticity greater than one at the relative point on the demand curve. For example, it would be sufficient that $-xh''(x)/h'(x) < 1$ for all x . We do not make the latter assumption as we did in Grossman and Lai (2004), because we wish to consider linear demands as a special case.

¹¹More formally, let a_j and ca_j be the unit labor requirement in country j for producing the numeraire good and a differentiated good, respectively. Then, by the normalization $a_N = 1$, we have $w_N = 1$, $w_S = a_N/a_S$, and one efficiency unit of Southern labor is equal to a_S/a_N units of Southern labor. The price of an unpatented good, which is sold at cost under perfect competition, is equal to c in both countries. This assumption of equal production costs is not needed for our results, but it simplifies the exposition.

in this activity is equal to the wage, or

$$v(K_N/L_{RN})^{1-\beta} = 1. \quad (4)$$

Then, given this allocation of labor, the flow of new innovations is $\phi_N = \frac{1}{\beta}K_N^{1-\beta}L_{RN}^\beta$, from which the steady-state measure of differentiated products can be computed.

We assume that Northern firms are owned by households in North and, without further loss of generality, that investments in R&D are financed by aggregate savings there.¹² Spending in North, E_N , is the difference between national income Y_N and national savings, or between national income and national investment. National income comprises labor income, the return to research capital, and the flow of monopoly profits paid as dividends, so $Y_N = L_N + rK_N + n_M[M_N\pi(p_N) + M_S\pi(p_S)]$, where r is the return to research capital and L_N is the aggregate labor supply in North. Investment is devoted entirely to R&D, the cost of which is $L_{RN} + rK_N$. Therefore,

$$E_N = L_N - L_{RN} + n_M[M_N\pi(p_N) + M_S\pi(p_S)]. \quad (5)$$

In South, all income derives from wages and all is spent, so $E_S = L_S$, where L_S is South's labor supply measured in efficiency units.

We are ready now to describe the policy environment and the objective functions of the two governments. North's government faces an initial decision about its policy for exhaustion of IPR. Our analysis focuses on the relative merits of its alternative options.¹³ We assume that the choice of IPR exhaustion is a long-term commitment; i.e., the government decides once and for all whether it will tolerate or ban parallel imports. Subsequently, at each moment $z > 0$, the governments set their price regulations for goods invented at time z . A government can commit at time z to a price ceiling that will apply during the economic life of any good i invented at time z . Without this ability to pre-commit, both governments would be tempted to enact price ceilings at the competitive level c for every new good the moment after it is invented and, foreseeing this, firms would have no incentive to invest in R&D. The ability to pre-commit can be sustained, as usual, by an appropriate reputation mechanism.¹⁴ We

¹²With quasi-linear preferences, our analysis would be unaffected if we were to introduce the possibility of international borrowing and lending. The constant interest rate equal to the discount rate makes consumers indifferent to the time profile of their spending.

¹³We could allow South's government to make a simultaneous choice of policy regarding parallel imports. However, inasmuch as prices of differentiated products are higher in North than South in all equilibria that we consider with national exhaustion, there is no incentive for parallel trade to flow from North to South, and thus South's policy toward parallel imports is irrelevant.

¹⁴Our assumption that the government sets its pricing ceiling before firms make their R&D decisions suggests that the government has more incentive to commit than the firms, or is better able to do so. We believe the

assume for the time being that North imposes no price controls or, equivalently, that its price ceiling for every good equals or exceeds the monopoly price p_M .

While in principle South's government could choose to set different price ceilings for goods invented at different times (or for different goods invented at the same time), it has no reason to do so. Therefore, we shall solve for a single price ceiling that is optimal for South for all goods and at all times, which however depends on what parallel trade regime is in effect. At time zero and thenceforth, the government of South chooses a price ceiling to maximize the discounted value of the total consumer surplus that its households achieve on goods invented subsequently. Considering that the number of Southern consumers is M_S , that innovation will be constant over time at rate ϕ_N , and that each newly invented good spends the first τ years of its economic life as a patented good with (endogenous) price p_S and the remaining $\bar{\tau} - \tau$ years as an unpatented good with price c , the South government's objective is to maximize the present discounted value of the sum of spending and consumer surplus,

$$W_S = \frac{L_S}{\rho} + \frac{M_S \phi_N}{\rho} [TC(p_S) + (\bar{T} - T) C_C] , \quad (6)$$

where $\bar{T} = (1 - e^{-\rho \bar{\tau}}) / \rho$ is defined analogously to T .

In North, aggregate welfare is the present value of spending plus consumer surplus, and spending is augmented by profit income but diminished by the savings that finance R&D; see (5). Accordingly, the government of North will evaluate the alternative trade regimes with an eye toward maximizing

$$W_N = \frac{L_N - L_{RN}}{\rho} + \frac{\phi_N}{\rho} [TM_N C(p_N) + (\bar{T} - T) M_N C_C + T \Pi_N] , \quad (7)$$

where $\Pi_N = M_N \pi(p_N) + M_S \pi(p_S)$ is the global profit flow earned by a typical Northern patent holder. The differences between (6) and (7) reflect that (i) the sizes of the two markets differ, (ii) profits accrue to firms in North but not to firms in South, and (iii) North devotes resources to R&D whereas South does not. In other words, as in Grossman and Lai (2004), the governments have different interests because the countries have different capacities for research and differently sized markets for innovative products.

assumption is plausible in our context, because the government's policies affect many potential investors at each moment in time and even more investors over time. In contrast, an individual firm makes a single R&D decision at a given moment and may invest for only a limited period. Accordingly, the government has ample incentive to invest in a reputation for not expropriating IPR via price controls, whereas the individual firms have no reason to commit to high levels of investment and little to gain by committing not to invest. We emphasize, however, that our timing assumption is crucial to our argument. Without pre-commitment by the Southern government, international exhaustion of IPR would kill all incentives for innovation. Lack of pre-commitment ability in North would imply zero innovation also under national exhaustion of IPR.

In the next section, we discuss the optimal price control in South in a regime with national exhaustion of IPR (no parallel imports) and in a regime with international exhaustion. Then we compare the aggregate welfare in North under the alternative regimes, to shed light on its decision whether or not to allow parallel trade. We also discuss how parallel trade affects welfare in South.

3 Price Controls with and without Parallel Trade

Throughout this section, we assume that the government of North does not impose price controls and firms in South do not innovate. When North prohibits parallel imports, South can limit the prices of differentiated products in its market without any effect on the prices that these products carry in North. Such restraints will benefit Southern consumers and will not cause any loss of income for Southern households. However, they will reduce the incentives for innovation in North to an extent that depends on the size of the Southern market. We derive the optimality conditions for the South's price control in such circumstances and then turn to the more subtle calculus that arises when parallel imports are tolerated in the North. Then the government of South must recognize that Northern firms would prefer at some prices to sacrifice sales in the Southern market in order to curtail arbitrage. And, if patent holders set prices for initial sales in South that fall within the bounds of the price controls, the opportunities for arbitrage will affect prices in the North and thus the profits that the innovative firms can earn at home. After we derive the optimal price controls in the presence of arbitrage, we compare the welfare levels that result in each country under the alternative parallel trade regimes.

3.1 Optimal Price Ceiling with National Exhaustion

Suppose that North prohibits reimportation of patented goods. With no price controls in North, each producer of a differentiated product charges the monopoly price p_M there for as long as its patent remains alive. In South, the monopolists can make positive profits at any price above c and will charge the highest price allowed under the price control, up to a maximum of p_M . Thus, South's government can choose the local price between c and p_M that maximizes W_S in (6), recognizing that the price determines both local consumer surplus $M_S C(p_S)$ and the flow of global profits per patent, $M_S \pi(p_S) + M_N \pi(p_M)$. The latter determines L_{RN} , which in turn determines the rate of innovation, via (4).

If the South sets a price control p_S above the marginal cost c , then the first-order condition

implies

$$\frac{d\phi_N}{dv} \frac{dv^{ne}}{dp_S} [TC(p_S) + (\bar{T} - T) C_C] + \phi_N TC'(p_S) = 0 \quad (8)$$

or, using (3), (4), $\phi_N = \frac{1}{\beta} \left(K_N^{1-\beta} L_{RN}^\beta \right)$ and the fact that $C'(p_S) = -x(p_S)$,

$$\frac{\beta}{1-\beta} \left[C(p_S) + \left(\frac{\bar{T} - T}{T} \right) C_C \right] \left[\frac{\pi'(p_S)}{x(p_S)} \right] = \pi(p_S) + m\pi(p_M), \quad (9)$$

where $v^{ne} = [M_S\pi(p_S) + M_N\pi(p_M)]T$ is the value of a patent with national exhaustion of IPR in North, $x(\cdot)$ is the demand per consumer for the representative differentiated product as a function of its price, and $m \equiv M_N/M_S$ is the relative size of the Northern market. For m large, the right-hand side of (9) exceeds the left-hand side for all values of p_S above c . Then, South's government sets its price ceiling p_S^{ne} equal to c , where p_S^{ne} denotes the price ceiling in a regime with national exhaustion. Otherwise, p_S^{ne} is the value of p_S that solves (9).

The first-order condition (9) expresses the tradeoff between the benefit to South of the extra product diversity that results from greater innovation and the loss in surplus that Southern consumers suffer as a result of higher prices there. The term $\beta/(1-\beta)$ represents the elasticity of innovation ϕ_N with respect to the value of a patent v , and thus can be interpreted as the supply elasticity of R&D.¹⁵ The government of South sets a more lenient price control (higher ceiling) when innovation is more responsive to profits, when product life is long relative to the duration of patents, and when the South's market is relatively large.¹⁶

3.2 Optimal Price Ceiling with International Exhaustion

Now suppose that North permits parallel imports. If the price of a differentiated product in North were to exceed the (regulated) price of the same good in South, an arbitrage opportunity would present itself. We have assumed that shipping and distribution are costless, which means that no such arbitrage opportunities can persist in equilibrium. With international exhaustion, either a differentiated product must carry the same price in both markets or the price charged by patent holders for sales in the regulated market must be so high as to preclude sales there.

Under the threat of such arbitrage, a Northern patent holder faces a choice. It can

¹⁵With $\phi_j = \frac{1}{\beta} K_j^{1-\beta} L_{Rj}^\beta$ and $v (\partial\phi_j/\partial L_{Rj}) = 1$,

$$\frac{\partial\phi_j}{\partial v} \frac{v}{\phi_j} = \frac{\beta}{1-\beta}.$$

¹⁶These claims rely on the second-order conditions being satisfied, which we assume to be the case.

set a price for sales in South that exceeds the price ceiling, thereby effectively cutting off supply from that region and eliminating any source of reimports. Or it can accept the competition from arbitrageurs and sell in both markets at a price less than or equal to the South's controlled price.¹⁷ The former strategy allows the firm to earn maximal profits in the Northern market. The latter strategy affords (possibly) positive profits in South at the expense of diminished profits at home.

We define \tilde{p} implicitly by

$$(M_N + M_S) \pi(\tilde{p}) = M_N \pi(p_M) ; \quad (10)$$

this is the price at which the two strategies yield equal profits. If the government of South sets a price ceiling below \tilde{p} , a Northern patent holder earns higher profits by selling at home at p_M and eschewing all sales in South than it does by selling in both markets at the maximum price allowed by South. It can achieve these profits, for example, by setting the price p_M for all sales. If the South's price ceiling instead exceeds \tilde{p} , the patent holder prefers to sell in both markets at the highest price allowed in South. It cannot be optimal for South's government to allow its market for differentiated products to be foreclosed. Accordingly, the government of South chooses its price ceiling p_S subject to the constraint that $p_S \geq \tilde{p}$.

Let us ignore the constraint for a moment. Then the first-order condition for the optimal choice of the South's price ceiling is given by

$$\frac{d\phi_N}{dv} \frac{dv^{ie}}{dp_S} [TC(p_S) + (\bar{T} - T) C_C] + \phi_N TC'(p_S) = 0 ,$$

which, with $v^{ie} = [(M_S + M_N)\pi(p_S)]T$, implies that

$$\frac{\beta}{1 - \beta} \left[C(p^*) + \left(\frac{\bar{T} - T}{T} \right) C_C \right] \left[\frac{\pi'(p^*)}{x(p^*)} \right] = \pi(p^*) , \quad (11)$$

¹⁷Joseph Harrington points out that, in addition to setting a price for its sales to the South, a monopolist might wish to limit the quantity that it sells there. If a firm could direct sales only to those Southern consumers who value the product more than the monopoly price, then it would certainly wish to do so. However, the firm cannot legally charge a high price that would force these consumers to reveal themselves, and it will have no means to target sales to the high-valuation consumers without knowing who they are. Moreover, every resident of the South has positive demand at the controlled Southern price, because they can all resell the product at a profit in the North. When firms lacks the ability to identify consumers who will not engage in arbitrage, the profitability of limiting total sales to the South will depend upon the rationing mechanism that is used to allocate those sales.

One simple assumption supports our formulation, however. Suppose that the monopolists do not sell directly to consumers but instead sell to competitive retailers who have zero cost and no demand for the product for personal use. Perfect enforcement of the price ceiling implies that the the retailers will prefer to sell their entire allotment in the North as long as the price there exceeds the controlled price in the South. In such circumstances, a Northern firm maximizes profits either by setting a price in excess of the ceiling and selling nothing to the South, or else by setting the regulated price and satisfying all demand by Southern retailers.

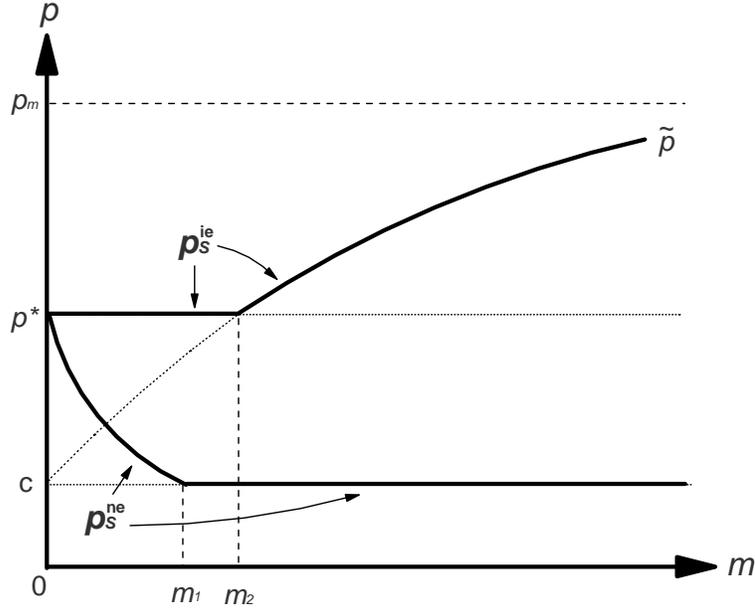


Figure 1: Price Ceilings under Alternative Parallel Trade Regimes

where v^{ie} is the value of a patent with international exhaustion (and thus uniform pricing) and p^* is the price ceiling that South would enforce in a regime with parallel trade if it were subject to no constraints.

If $p^* > \tilde{p}$, the constraint that dictates a price ceiling high enough to ensure positive sales in South does not bind. Then the government optimally sets a maximum price of p^* . Otherwise, the government sets the ceiling at \tilde{p} , the lowest level that generates positive supply. In other words, the optimal price ceiling for South in a regime in which North permits parallel imports is given by

$$p_S^{ie} = \max(\tilde{p}, p^*) . \quad (12)$$

3.3 Comparing Prices and Profits in the Alternative Regimes

In Figure 1, we depict p_S^{ne} , p^* and \tilde{p} as functions of m , the relative size of the Northern market. The price ceiling with national exhaustion p_S^{ne} is a decreasing function of m until it reaches c at some finite level of m , which we denote by m_1 . South's government would never set a price ceiling below c , because patent holders would not serve the Southern market at prices below marginal production cost. Accordingly, $p_S^{ne} = c$ for all $m \geq m_1$. Notice that $p^* = p_S^{ne}$ when $m = 0$ inasmuch as the incentives facing the Southern government are the same in the

two regimes when the Northern market is negligibly small. However, p^* is independent of m under international exhaustion; this is because arbitrage renders the effect of price controls on the incentives for innovation independent of the geographic distribution of world demand. Finally, (10) implies that $\tilde{p} = c$ at $m = 0$ and that \tilde{p} is an increasing function of m for all $m > 0$. This is because patent holders will sell in South at any price equal to or exceeding marginal cost when the Northern market is negligibly small, but require an ever higher price to sell there (under the threat of arbitrage) as the Northern market grows relatively larger. Let m_2 be the level of m at which $p^* = \tilde{p}$.

The figure depicts p_S^{ne} and p_S^{ie} in bold. Evidently, we have

Proposition 1 *Suppose that South does not innovate and North imposes no price controls. Then $p_S^{ie} > p_S^{ne}$ for all $m > 0$.*

Intuitively, the prospect of arbitrage mitigates the ability of South to free ride on the intellectual property protection afforded by North. With more at stake in terms of the adverse effect of a price ceiling on the innovation rate, the South will accept higher domestic prices. And the gap between p_S^{ie} and p_S^{ne} is all the greater when, in the presence of parallel trade, the South is forced to set a price ceiling even higher than its preferred (unconstrained) choice to give patent holders reason to serve its market.

Now we are ready to compare the options facing North's government. We begin with $m < m_2$, so that $p^* > \tilde{p}$ and thus $p_S^{ie} = p^*$. Let us define the function

$$\Omega(p) \equiv \frac{\beta}{1-\beta} \left[C(p) + \left(\frac{\bar{T} - T}{T} \right) C_C \right] \left[\frac{\pi'(p)}{x(p)} \right].$$

Notice that the left-hand side of (9) is $\Omega(p_S^{ne})$ and the left-hand side of (11) is $\Omega(p^*)$. For some demand functions (more on this later), the following condition holds.

Condition 1 *Either $\Omega'' < 0$ or $|\Omega''/\Omega'| < |\pi''/\pi'|$ for all $p \in [c, p_M]$.*

In the appendix we prove the following lemma.

Lemma 1 *Suppose that South does not innovate, North imposes no price controls, $m < m_1$, and Condition 1 is satisfied. Then $(M_N + M_S) \pi(p^*) > M_N \pi(p_M) + M_S \pi(p_S^{ne})$.*

The consequent of Lemma 1 states that aggregate profits for a patent holder are greater when there is international exhaustion of IPR in North and South sets its price control at p^* than they are when there is national exhaustion and South sets its price control at $p_S^{ne} > c$. Figure 2, which depicts the case of a linear $\Omega(p)$ curve, may help us to understand why this

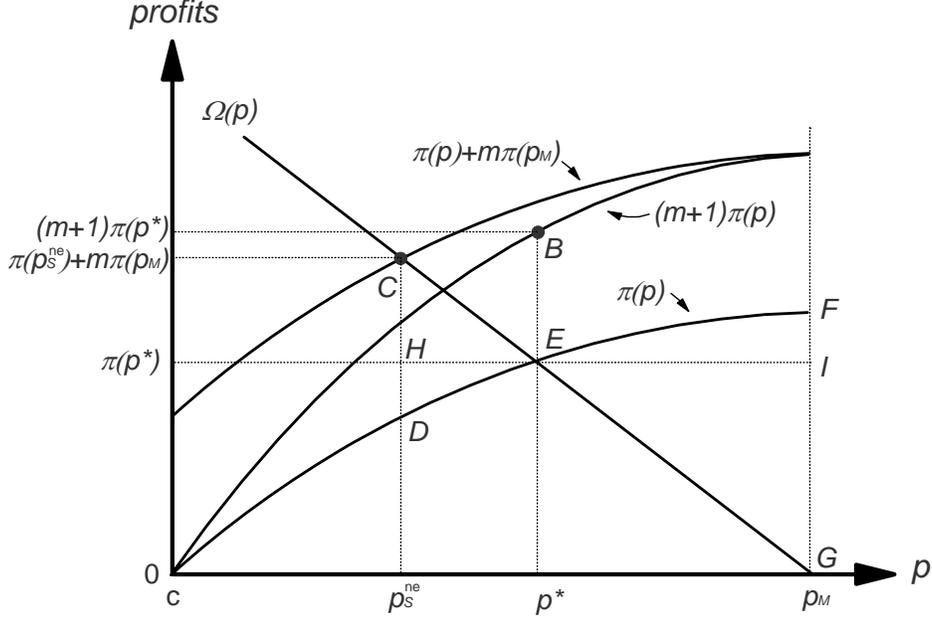


Figure 2: Global Profits under Alternative Parallel Trade Regimes

is so. The concave $\pi(p)$ curve and the linear $\Omega(p)$ intersect at point E , which determines p^* . The figure also shows $(m+1)\pi(p)$ and $\pi(p) + m\pi(p_M)$. The intersection of $\Omega(p)$ and $\pi(p) + m\pi(p_M)$ at the point labelled C gives the price ceiling with national exhaustion, p_S^{ne} . Note that $(M_N + M_S)\pi(p^*) > M_N\pi(p_M) + M_S\pi(p_S^{ne})$ if and only if $(m+1)\pi(p^*) > \pi(p_S^{ne}) + m\pi(p_M)$; i.e., if and only if the length of HD exceeds m times the length of FI . But $m = [\Omega(p_S^{ne}) - \pi(p_S^{ne})] / \pi(p_M)$ by (9), and so is represented in the figure by the ratio of the length of CD to the length of FG . Therefore, for the length of HD to exceed m times the length of FI , we need the ratio of HD to CD to exceed the ratio of FI to FG . But this is true, because the triangle CEH is similar to the triangle EIG , while the curve DEF is concave. If the $\Omega(p)$ curve passing through G and E instead were concave, the point C would lie to the southwest of the position shown, which would strengthen the argument. However, if the $\Omega(p)$ curve passing through G and E were convex, point C would lie to the northeast of the position shown. In this case, the length of HD would exceed m times the length of FI if and only if the absolute value of the curvature of $\Omega(p)$ were less than that of the curvature of $\pi(p)$, as provided in the second part of Condition 1.

Next we will show that, under Condition 1, $m_1 < m_2$, as depicted in Figure 1. Recall that m_1 is the smallest value of m for which $p_S^{ne} = c$ (South's government sets a price ceiling at the marginal production cost in the absence of parallel imports), while m_2 is

the smallest value of m for which $\tilde{p} \geq p^*$ (the constraint on South's government to ensure positive supply in a regime with parallel imports binds). Suppose that $m_1 > m_2$. Then $(1 + m_2)\pi(p^*)$ exceeds $m_2\pi(p_M) + \pi[p_S^{ne}(m_2)]$ by Lemma 1, which in turn exceeds $m_2\pi(p_M)$ by the fact that $p_S^{ne}(m_2) > c$. But $m_2\pi(p_M) = (1 + m_2)\pi[\tilde{p}(m_2)]$ by the definition of \tilde{p} . So $(1 + m_2)\pi(p^*) > (1 + m_2)\pi[\tilde{p}(m_2)]$, which implies that $p^* > \tilde{p}$ at m_2 . But this contradicts the definition of m_2 .

For m between m_1 and m_2 , South's government sets a price ceiling of $p_S^{ne} = c$ in a regime with national exhaustion of IPR in North and a price ceiling of $p_S^{ie} = p^*$ in a regime with international exhaustion in North. With national exhaustion, aggregate profits per patent are $M_N\pi(p_M)$. With international exhaustion, profits are $(M_N + M_S)\pi(p^*)$, which exceeds $(M_N + M_S)\pi(\tilde{p})$ because $p^* > \tilde{p}$. But $(M_N + M_S)\pi(\tilde{p}) = M_N\pi(p_M)$ by the definition of \tilde{p} . It follows that global profits per patent are higher with international exhaustion than with national exhaustion for this range of values of m as well.

Finally, when $m \geq m_2$, $p_S^{ne} = c$ and $p_S^{ie} = \tilde{p}$. In either case, global profits are equal to $M_N\pi(p_M)$. We conclude that, under Condition 1, the value of a patent when North tolerates parallel imports is no less than the value of a patent when it bars such imports, and patents actually are more valuable with parallel trade whenever $m < m_2$. Innovation responds positively to the value of a patent. It follows that innovation is at least as great, and for small m is greater, in a regime with legal parallel trade.

We are now ready to address the North's choice of parallel trade regime and to evaluate the implications of this choice for welfare in South. Let W_N^j denote North's aggregate welfare (as defined in (7)) in regime j , for $j = \{ne, ie\}$. We have

Proposition 2 *Suppose that South does not innovate, North imposes no price controls, and Condition 1 is satisfied. Then $W_N^{ie} > W_N^{ne}$.*

Proof. We have already seen that the innovation rate ϕ_N is at least as great in a regime with international exhaustion in North as in a regime with national exhaustion. Also, $p_N^{ne} = p_M$ while $p_N^{ie} = \max[p^*, \tilde{p}] < p_M$, where p_N^j is the price of a patented differentiated product in North in regime j . So $C(p_N^{ie}) > C(p_N^{ne})$. Finally, note the term $\phi_N T\Pi_N - L_{RN}$ that appears in W_N . This is aggregate profits from all patents minus the cost of labor used in research. But $T\Pi_N = v$, the value of a patent, and the L_{RN} that is chosen in each regime maximizes $\phi_N v - L_{RN}$; see (4). It follows that $\phi_N^{ie} T\Pi_N^{ie} - L_{RN}^{ie} \geq \phi_N^{ne} T\Pi_N^{ne} - L_{RN}^{ne}$, because $\max_{L_{RN}} (\phi_N v - L_{RN})$ is increasing in v and the value of a patent is at least as great in a regime with international exhaustion as in a regime with national exhaustion. ■

Whereas North fares better by choosing international exhaustion (when Condition 1 is satisfied), this choice adversely impacts welfare in South. We record this observation in

Proposition 3 *Suppose that South does not innovate, North imposes no price controls, and Condition 1 is satisfied. Then $W_S^{ie} < W_S^{ne}$.*

Proof. With international exhaustion, the world price of a patented good is p_S^{ie} . In a regime with national exhaustion, it is feasible for South's government to set its price ceiling for patented goods at p_S^{ie} . This would generate the same consumer surplus in South as in a regime with international exhaustion and, because $p_S^{ie} < p_M$, more innovation and greater product diversity. Thus, by setting a price ceiling of p_S^{ie} in a regime with national exhaustion, South could achieve greater welfare than W_S^{ie} . Its optimal choice of p_S^{ne} must yield even greater welfare. ■

This leads us to ask how likely it is that Condition 1 will be satisfied. The expression for Condition 1 in terms of the primitive parameters of the model is rather complicated. But we can readily examine a pair of special cases that yield familiar demand functions.

Constant Elasticity Demand

Suppose that $h(x) = a^{1/\varepsilon} \left(\frac{\varepsilon}{\varepsilon-1}\right) x^{(\varepsilon-1)/\varepsilon}$ for $a > 0$, $\varepsilon > 1$. Then the demand function is $x(p) = ap^{-\varepsilon}$, which has a constant-elasticity form. The monopoly price in this case is a mark-up over marginal cost, or $p_M = c\varepsilon/(\varepsilon - 1)$. We calculate that

$$\Omega'(p) = -a \left(\frac{\beta}{1-\beta}\right) p^{-\varepsilon} \left[1 - \frac{\varepsilon(p-c)}{p}\right] - a \left(\frac{\beta}{1-\beta}\right) \frac{c\varepsilon}{p^2} \left[\frac{p^{1-\varepsilon}}{\varepsilon-1} + \frac{c^{1-\varepsilon}}{\varepsilon-1} \left(\frac{\bar{T}}{T} - 1\right)\right],$$

which is negative for all $p \in [c, p_M]$. Now we define $Z \equiv \Omega''(p)\pi'(p) - \Omega'(p)\pi''(p)$. and calculate

$$Z = -a^2 c\varepsilon p^{-2(2+\varepsilon)} \left(\frac{\beta}{1-\beta}\right) \left\{ p^2 + c^{1-\varepsilon} p^\varepsilon [c\varepsilon - (\varepsilon - 2)p] \left(\frac{\bar{T}}{T} - 1\right) \right\},$$

which is negative for all $p \in [c, c\varepsilon/(\varepsilon - 1)]$. Since $\Omega'(p) < 0$ and $\pi'(p) > 0$ for all $p \in [c, p_M]$, $Z < 0$ implies $|\Omega''/\Omega'| < |\pi''/\pi'|$ and Condition 1 is satisfied. Notice that this is true for all values of $\varepsilon > 1$.

Linear Demand

Suppose that $h(x) = A + ax/b - x^2/2b$ for $a > 0$, $b > 0$, and $a > bc$. This last parameter restriction ensures positive demand when a good is competitively priced. Then the demand function $x(p) = a - bp$ is linear and the monopoly price is $p_M = (a + bc)/2b$. We find that

$$\Omega'(p) = -\frac{1}{2} \left(\frac{\beta}{1-\beta}\right) \left[3a + bc - 4bp + \frac{(a-bc)^3 \left(\frac{\bar{T}}{T} - 1\right)}{(a-bp)^2} \right],$$

which again is negative for all $p \in [c, p_M]$. Now we calculate

$$\left| \frac{\Omega''(p)}{\Omega'(p)} \right| - \left| \frac{\pi''(p)}{\pi'(p)} \right| = \frac{2b - \left[\frac{(a-bc)^3 \left(\frac{\bar{T}}{T} - 1 \right) b}{(a-bp)^3} \right]}{\left(\frac{3a}{2} + \frac{bc}{2} - 2bp \right) + \left[\frac{(a-bc)^3 \left(\frac{\bar{T}}{T} - 1 \right)}{2(a-bp)^2} \right]} - \frac{2b}{a - 2bp + bc} ,$$

which is negative for all $p \in [c, (a + bc) / 2b]$. Therefore, with linear demands too, Condition 1 is satisfied.

We conclude from these examples that Condition 1 is not unreasonable inasmuch as it always is satisfied for two families of commonly-used demand functions. When the condition is satisfied, the orthodox view that allowing parallel imports would retard innovation and harm a country that specializes in research is false. Once we recognize that this country's trade partners will regulate prices differently according to the prevailing policy on exhaustion of IPR, we find that the policy change induced by a switch to international exhaustion can easily be large enough to allow the innovating country to enjoy lower prices, greater profits, faster innovation, and higher welfare.

4 Extensions

In this section, we relax two of the simplifying assumptions that were used in the preceding analysis to highlight our main points. First, we allow for positive research capacity and thus active innovation activity in South. Second, we allow for the possibility that North may institute price controls of its own.

4.1 Innovation in South

In Section 3, we emphasized the asymmetries in research capacities by assuming that $K_S = 0$. Now we allow for $K_S > 0$, which, with the Cobb-Douglas technology for product development, ensures positive innovation in South. As before, we assume that any wage gap that exists between North and South matches a gap in labor productivity in all activities (including research). Then L_{RS} in the technology for product development in South measures the labor deployed in Southern research in efficiency units.

When South is endowed with research capital, the objective function for country j becomes

$$W_j = \frac{1}{\rho} \left\{ L_j - L_{Rj} + (\phi_N + \phi_S) M_j [TC(p_j) + (\bar{T} - T) C_C] + \phi_j T \Pi_j \right\} ,$$

where $\Pi_j = M_N\pi(p_N) + M_S\pi(p_S)$ is the global profit earned by a typical producer in country j of a patented good when prices of such products in North and South are p_N and p_S , respectively. The expression for welfare reflects that households in country j derive consumer surplus from differentiated products produced anywhere in the world, but they derive profit income only from producers in their own country. Notice that $\Pi_S = \Pi_N$ under our assumption of wage equalization in efficiency units, and thus the value of a patent v is the same in both countries.

Firms in South hire (efficiency) labor to perform research up to the point where the marginal value product is equal to the wage, or $v(K_S/L_{RS})^{1-\beta} = 1$. Together with (4), the analogous condition for equilibrium investment in North, and the production function for innovation, $\phi_j = \frac{1}{\beta}L_{Rj}^\beta K_j^{1-\beta}$, this implies $\phi_N/\phi_S = K_N/K_S$. We let $\mu_j = K_j/(K_N + K_S)$ denote the fraction of the world stock of research capital located in country j . Then μ_j also is the fraction of world innovation that takes place in country j ; i.e., $\mu_j = \phi_j/(\phi_N + \phi_S)$.

Still assuming that North does not regulate prices, South's government chooses p_S in a regime with national exhaustion of IPR to maximize W_S . The first-order condition that applies when $p_S^{ne} > c$ is

$$\frac{d}{dp_S}\{M_S(\phi_N + \phi_S)[TC(p_S) + (\bar{T} - T)C_C] + v\phi_S - L_{RS}\} = 0$$

or¹⁸

$$\left(\frac{\beta}{1-\beta}\right) \frac{C(p_S) + \left(\frac{\bar{T}-T}{T}\right)C_C}{\frac{x(p_S)}{\pi'(p_S)} - \mu_S} = \pi(p_S) + m\pi(p_M) \quad (13)$$

Note the similarity with (9), to which (13) reduces when $\mu_S = 0$.

We can similarly derive South's choice of price control that applies under international exhaustion of IPR when the constraint that $p_S^{ie} \geq \tilde{p}$ does not bind. The analog to equation (11) above is

$$\left(\frac{\beta}{1-\beta}\right) \frac{C(p^*) + \left(\frac{\bar{T}-T}{T}\right)C_C}{\frac{x(p^*)}{\pi'(p^*)} - (m+1)\mu_S} = \pi(p^*) . \quad (14)$$

We define

$$\Lambda(p) \equiv \left(\frac{\beta}{1-\beta}\right) \frac{C(p) + \left(\frac{\bar{T}-T}{T}\right)C_C}{\frac{x(p)}{\pi'(p)} - \mu_S} .$$

Then (13) can be written as $\Lambda(p_S) = \pi(p_S) + m\pi(p_M)$, while the value of p^* that solves (14)

¹⁸To derive (13), we use the fact that $dv^{ne}/dp_S = M_S T \pi'(p_S)$, $dC(p_S)/dp_S = -x(p_S)$ and $v(d\phi_S/dp_S) = dL_{RS}/dp_S$.

must be bigger than the p that solves $\Lambda(p) = \pi(p)$; call the latter p^{**} .

We now introduce a modified version of Condition 1, namely

Condition 2 *Either $\Lambda'' < 0$ or $|\Lambda''/\Lambda'| < |\pi''/\pi'|$ for all $p \in [c, p_M)$.*

We can use exactly the same strategy and arguments as before to prove that if Condition 2 is satisfied and $p_S^{ne} > c$, then $(m+1)\pi(p^{**}) > m\pi(p_M) + \pi(p_S^{ne})$. But since $p^* > p^{**}$, this is sufficient to establish that $(m+1)\pi(p^*) > m\pi(p_M) + \pi(p_S^{ne})$. In other words, with Condition 2 and $p_S^{ne} > c$, global profits per firm and the world innovation rate are greater with international exhaustion of IPR in North than with national exhaustion. It is also true as before that when $p_S^{ne} = c$, global profits per firm are at least as high with international exhaustion as with national exhaustion. And North enjoys higher consumer surplus with parallel trade. This establishes

Proposition 4 *Suppose that North imposes no price controls and Condition 2 is satisfied. Then $W_N^{ie} > W_N^{ne}$.*

Note, however, that Condition 2 may be more stringent than Condition 1. We compute for example that if the demand function $x(p)$ is linear, Condition 2 is satisfied for all parameters b , c and $a > bc$ when $\mu_S < .38$. If the demand function has constant elasticity ε then Condition 2 holds when $\mu_S < 1/2$ and both ε and \bar{T}/T are sufficiently large. For example, if $\mu_S = 0.2$, Condition 2 is satisfied if $3.1\bar{T} + \varepsilon(1.9T - 2.3\bar{T}) < 1.9T$ or if $\varepsilon = 2$ and $\bar{T}/T > 1.3$.¹⁹ Of course, Condition 2 must be satisfied for sufficiently small (but positive) μ_S . And while Condition 2 is sufficient for $W_N^{ie} > W_N^{ne}$, it is not a necessary condition for this outcome.

4.2 Price Controls in North

Up until now, we have assumed that North does not impose any price controls of its own. This assumption can be rigorously justified if administering price controls imposes fixed costs, and if these costs fall in a range such that maintaining a regulatory program is worthwhile for South but not for North. Now we shall examine a policy environment in which the fixed costs of maintaining price controls, while not literally zero, are arbitrarily small.²⁰

¹⁹The details of these calculations are available from the authors upon request.

²⁰We make the assumption of small administrative costs to eliminate some unreasonable Nash equilibria of the regulation game that can otherwise arise when parallel imports are allowed. When regulatory schemes are costless, each country might impose the same low price ceiling and while both countries would prefer to see higher prices of patented goods, neither has an incentive to raise its ceiling as long as the other's ceiling is unchanged.

To simplify the exposition, we assume once again that all research capacity is concentrated in North. First consider the regulation game in a regime in which North does not allow parallel imports. In this case, the best response by South to a price ceiling p_N in North is given by the solution to

$$\left(\frac{\beta}{1-\beta}\right) \frac{C(p_S) + \left(\frac{\bar{T}-T}{T}\right) C_C}{\frac{x(p_S)}{\pi'(p_S)}} = \pi(p_S) + m\pi(p_N), \quad (15)$$

if such a solution exists with $p_S > c$, and by $p_S^{ne} = c$ otherwise. In North, the maximization of W_N in (7) with respect to the choice of p_N implies

$$m \left(\frac{\beta}{1-\beta}\right) \frac{C(p_N) + \left(\frac{\bar{T}-T}{T}\right) C_C}{\frac{x(p_N)}{\pi'(p_N)} - 1} = \pi(p_S) + m\pi(p_N). \quad (16)$$

Notice that $p_N < p_M$ for all values of p_S .²¹ This is because a price ceiling in North set just below the monopoly price causes only a second-order reduction in profits and so a negligible loss of product diversity, but generates a first-order gain in consumer surplus. Note too that $p_N^{ne} > p_S^{ne}$ as long as $m \geq 1$, which reflects the fact that monopoly profits accrue to households in North but not to households in South and that a country with a larger market provides more incentives to innovators for the same increment in the price ceiling.

Now consider the game that emerges when parallel imports are permitted by North. We define p_N^* as the uniform global price ceiling most preferred by the government of North²² and p^* as the one most preferred by South, which once again is the value of p that solves (11). Note that $p_N^* > p^*$, because the two governments face similar incentives except that monopoly profits accrue to households in North. If $(M_N + M_S) \pi(p^*) > M_N \pi(p_M)$, then South will set a price ceiling of $p_S^{ie} = p^*$ and North will impose no price ceiling in the Nash equilibrium of the regulation game. In this case, South's government is unconstrained in its choice of price ceiling, and North's government has no reason to bear the (small) fixed cost of a regulatory program. If $(M_N + M_S) \pi(p^*) < M_N \pi(p_M)$ but \tilde{p} (as defined in [10]) is less than p_N^* , then $p_S^{ie} = \tilde{p}$ and again North has no price ceiling in the Nash equilibrium. In this case,

²¹If $p_N = p_M$, then $\pi'(p_N) = 0$ and the left-hand side of (16) equals zero, while the right-hand side of (16) is strictly positive.

²²That is,

$$p_N^* = \arg \max_p M_N \phi_N [TC(p) + (\bar{T} - T) C_c] + v \phi_N - L_{RN},$$

where $v = T(M_N + M_S) \pi(p)$, which implies that p_N^* is the solution to

$$\left(\frac{\beta}{1-\beta}\right) \frac{C(p_N^*) + \left(\frac{\bar{T}-T}{T}\right) C_C}{\frac{x(p_N^*)}{\pi'(p_N^*)} - \left(\frac{1}{m} + 1\right)} = \pi(p_N^*).$$

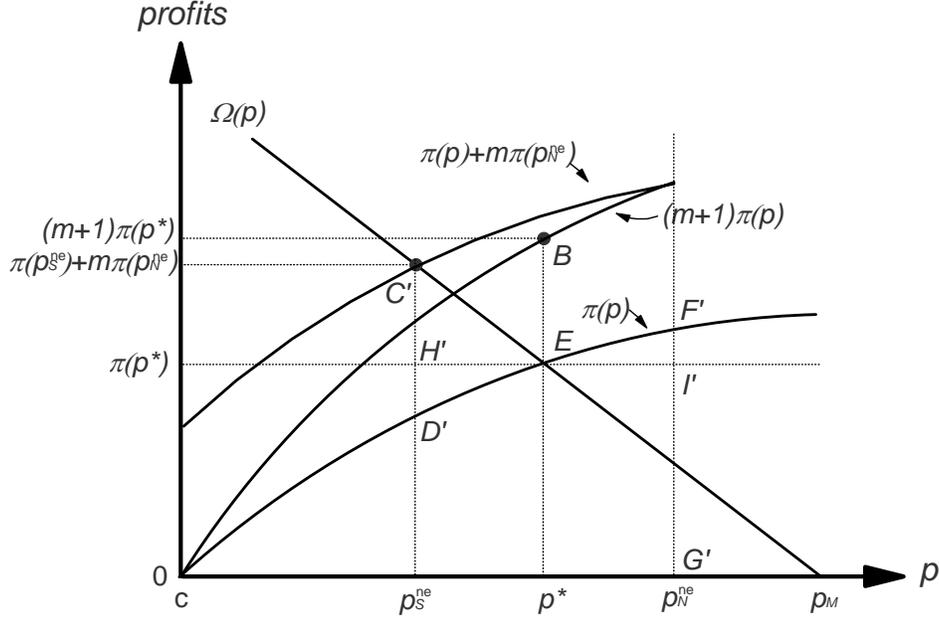


Figure 3: Profit Comparison with Northern Price Controls

South is constrained and it chooses \tilde{p} , but North has no reason to introduce a price control because it would like to see South choose an even higher value of p_S^{ie} than this. Finally, if $(M_N + M_S) \pi(p^*) < M_N \pi(p_M)$ and $\tilde{p} > p_N^*$, then North sets a price ceiling such that the constrained government of South will choose $p_S^{ie} = p_N^*$ in order that its market be served.²³

In the appendix, we prove the following proposition that compares global profits per firm, the values of a patent and rates of innovation in the alternative regimes.

Proposition 5 *Suppose that South does not innovate and Condition 1 is satisfied. Then $(M_N + M_S) \pi(p_S^{ie}) > M_N \pi(p_N^{ne}) + M_S \pi(p_S^{ne})$, $v^{ie} > v^{ne}$, and $\phi_N^{ie} > \phi_N^{ne}$.*

The proposition states that profits are higher, patents are more valuable, and innovation is faster when the North allows parallel imports than when it does not. The proof is tedious, because it requires separate consideration of five different cases depending on whether p_S^{ie} is equal to p^* , \tilde{p} , or p_N^* , whether p_S^{ne} equals or exceeds c , and whether p^* exceeds or falls short of p_N^{ne} . However, the reasoning is straightforward in several of these scenarios, which we describe now.

²³That is, North chooses p_N^{ie} so that $(M_N + M_S) \pi(p_N^{ie}) = M_N \pi(p_N^{ie})$.

First, suppose that $p^* > p_N^{ne}$; that is, an unconstrained South would choose a higher price ceiling under international exhaustion than North's Nash equilibrium price ceiling in the regulation game with national exhaustion. Since the prevailing world prices with international exhaustion are at least as high as p^* , and the prices in North and South with national exhaustion are both lower than p^* , global profits must be higher in the former regime than in the latter.

Next suppose that $p_S^{ie} = p^* < p_N^{ne}$ and $p_S^{ne} > c$; i.e., the constraint on South's government to ensure positive supply of its market does not bind in either regime. Then it chooses p^* under international exhaustion and a price above marginal cost under national exhaustion. Now the argument is very similar to that used in proving the analogous result that applies when North eschews price regulation. Figure 3, like Figure 2, depicts a linear $\Omega(p)$ curve. When there is national exhaustion of IPR in North, the Nash equilibrium of the regulation game has $\Omega(p_S^{ne}) = m\pi(p_N^{ne}) + \pi(p_S^{ne})$ which is represented by the point C' in the figure. South's choice of price ceiling with international exhaustion is found at E , where $\Omega(p^*) = \pi(p^*)$. The comparison of global profits in the two equilibria now requires a comparison of the length of $H'D'$ with m times the length of $F'I'$. But the relative size of the two markets m is represented in Figure 3 by the ratio of the length of $C'D'$ to the length of $F'G'$. So, the concavity of $\pi(p)$ ensures that $(m+1)\pi(p^*) > m\pi(p_N^{ne}) + \pi(p_S^{ne})$. And, if $\Omega(p)$ is concave, or convex with curvature less in absolute value than that of $\pi(p)$, we reach the same conclusion.

A third case arises when $p_S^{ie} = \min[\tilde{p}, p_N^*]$ and $p_S^{ne} = c$; i.e., South would like to set a lower price ceiling in both regimes, but in each case it is constrained by the need to ensure positive supply of its market. In this scenario, the profits in North under international exhaustion are at least as large as $\pi(p_N^*)$ and those in South are positive. When parallel imports are forbidden by North, profits there are $\pi(p_N^{ne})$ while those in South are nil. But p_N^* exceeds p_N^{ne} for much the same reason that p^* exceeds p_S^{ne} ; namely, North would choose a higher price were that price to apply to markets in both countries than the price it chooses when the price applies only at home. So patent holders earn higher profits in both countries under international exhaustion than under national exhaustion.

Although Proposition 5 gives conditions under which parallel imports promote innovation (by inducing changes in price regulation), these conditions do not guarantee a welfare gain for households in North. Contrary to the common wisdom that parallel imports would enhance consumer surplus but retard innovation, just the opposite can happen when North has price controls of its own in the absence of parallel trade, and the endogenous response of these controls is taken into account. If $p_S^{ie} = p^* < p_N^{ne}$, then consumer surplus in North is higher with international exhaustion than national exhaustion, and Northern welfare is unambiguously higher in the former regime than the latter. But if a ban on parallel trade leads North's

government to set a reasonably tight price control in the Nash equilibrium of the regulation game, then a removal of the ban could lead to a rise in prices for Northern consumers.²⁴ If so, the North still may benefit from parallel trade, but the realization of such a benefit would require that the gains from greater product variety exceed the losses from higher prices.

5 Several or Many Countries in the South

Until now, we have considered a two-country world with a single North in which firms invest in R&D and a single South with little or no such investment. We have argued that, by opening itself to parallel trade, the government of the North induces its counterpart in the South to set a higher price ceiling than otherwise. The South's choice of a high price ceiling reflects one of two considerations: either it wishes to ensure that its market is not cut off from innovative products or it seeks to preserve sufficient incentives for R&D. It may seem that the latter motivation rests on there being only one or a few countries in the South, because it requires that the government perceives a non-negligible influence of its own policy on the value of a patent. We address this issue by allowing for multiple Southern countries and show that, in fact, the conclusion does not require the presence of a large actor in the South.

Suppose that the South comprises n_S identical countries, each with a population of m_S consumers. Let $M_S = n_S m_S$ be the total population in the South and let $m = M_N/M_S$ be the relative size of the North, as before. We assume as in Section 3 that no innovation takes place in the South and that the North adopts no price controls, and we focus on symmetric Nash equilibria in which all the countries in the South set the same maximum price.²⁵

As before, we are interested in the North's choice between national exhaustion and international exhaustion of IPR. With several or many countries in the South, the relative attractiveness of these alternative regimes to North may hinge upon the permissibility of parallel imports in its various trade partners. One possible way for us to proceed would be to consider the simultaneous setting of policies toward parallel trade by the North and all Southern countries. However, analysis of the multi-stage game of choice of parallel trade regime followed by ongoing choices of price ceilings is complicated by the fact that pure-strategy equilibria do not exist in the price-control sub-games for all possible configurations

²⁴This could happen either because $p_S^{ie} = p^*$ and $p^* > p_N^{ne}$ or because $p_S^{ie} = \min[\tilde{p}, p_N^*] > p^*$ and $\min[\tilde{p}, p_N^*] > p_N^{ne}$.

²⁵A symmetric Nash equilibrium always exists in the settings that we consider. In some cases—especially when the number of countries in the South is small—there may also exist asymmetric equilibria. However, the existence of such equilibria seems to require that the demand functions have special properties. Since our point can be seen most clearly when the Southern governments behave similarly, and since we have little to say about equilibrium selection when multiple equilibria exist, we choose to focus on the symmetric equilibrium.

of parallel trade policies.²⁶ Rather than delve into the complexities introduced by mixed strategies, we choose instead to treat the policies of the South with regard to parallel imports as exogenous. We will consider in turn the attractiveness to the North of allowing parallel trade under the alternative assumptions that all Southern countries prohibit parallel imports and that all Southern countries tolerate such imports.

5.1 National Exhaustion of IPR in the South

Suppose the symmetric countries in the South all prohibit parallel imports. By doing so, each preserves autonomy in regulating its own internal prices without the risk that its decisions will be undermined by arbitrage. We consider in turn the alternative regimes available to the government of the North and then compare the two outcomes.

If the North also practices national exhaustion of IPR, each country in the South has the same first-order condition for its choice of a price ceiling (analogous to [8]), which for country i is given by

$$\frac{d\phi_N}{dv} \frac{dv^{ne}}{dp_i} [TC(p_i) + (\bar{T} - T) C_c] - \phi_N T x(p_i) = 0 .$$

Note that

$$\frac{d\phi_N}{dv} \frac{v}{\phi_N} = \frac{\beta}{1 - \beta}$$

is the elasticity of innovation supply with respect to the value of a patent and that

$$\frac{dv^{ne}}{dp_i} \Big|_{\mathbf{p}_{-i}} = T m_S \pi'(p_i) = T \frac{M_S}{n_S} \pi'(p_i)$$

is the effect of a change in the price ceiling in country i on the value of a patent, where \mathbf{p}_{-i} is the vector of prices in Southern countries other than i . With no parallel trade, each country affects world innovation in proportion to the size of its market. In a symmetric equilibrium, $p_i = p_S$ for all i , so $v^{ne} = T [M_S \pi(p_S) + M_N \pi(p_m)]$. Combining, as we have done before, gives

$$\frac{\beta}{1 - \beta} \left[C(p_S) + \left(\frac{\bar{T} - T}{T} \right) C_C \right] \left[\frac{\pi'(p_S)}{x(p_S)} \right] = n_S [\pi(p_S) + m\pi(p_m)] , \quad (17)$$

which is analogous to (9). The p_S^{ne} curve implicit in (17) again is downward sloping as a function of m and becomes horizontal at c . But notice that the right-hand side of (17) increases with n_S ; this captures the declining influence of each Southern country on the

²⁶In particular, we find that when $n_S - 1$ countries in the South practice national exhaustion of IPR, and the remaining country in the South plus the North practice international exhaustion of IPR, there exists no pure-strategy equilibrium in the second-stage game between the Southern countries. The described configuration of parallel trade policies arises naturally when considering a potential equilibrium in which all Southern countries choose national exhaustion and one contemplates a deviation to international exhaustion.

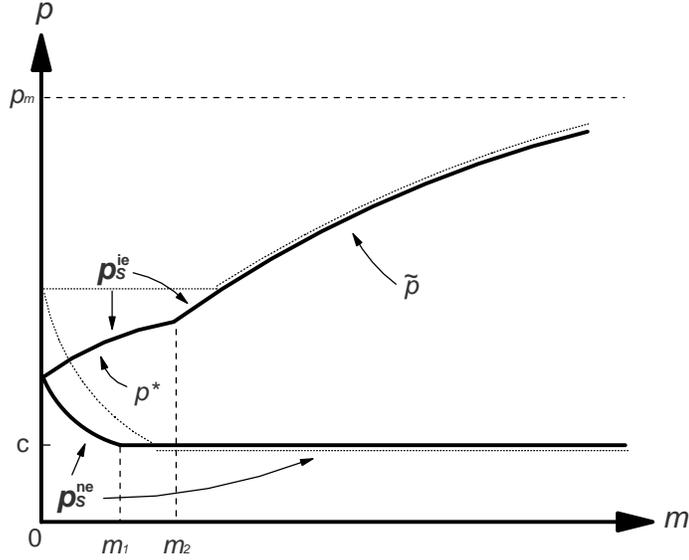


Figure 4: Price Ceilings with Several Southern Countries

world rate of innovation as the typical such country becomes smaller. An increase in n_S shifts the downward-sloping portion of the p_S^{ne} curve downward, and of course results in an intersection with c at a smaller value of m . This is shown in Figure 4, where the bold curves represent a case with $n_S > 1$ and the dotted curves apply for $n_S = 1$ and are reproduced from Figure 1 for reference.

Now suppose that the North instead practices international exhaustion of IPR. We first determine \tilde{p}_i , the critical price ceiling for country i at which the Northern monopolists are indifferent between supplying country i with innovative products and not, considering the arbitrage that results. We are looking for a symmetric equilibrium among the Southern countries; in such an equilibrium, $\tilde{p}_i = \tilde{p}$ for all i . We will show that \tilde{p} is the solution to

$$(M_S + M_N) \pi(\tilde{p}) = M_N \pi(p_m) ;$$

i.e., \tilde{p} is independent of the number of countries in the South and has the same value as before.

To see this, notice first that if all Southern countries set a price ceiling above \tilde{p} , the Northern firms prefer to serve all of them to serve none (or some) of them, even considering the erosion of profits in the North that results from arbitrage. But if all Southern countries set a price below \tilde{p} , the Northern firms prefer to serve none than to serve some or all of them.

They can extinguish supply to the South and the resulting arbitrage activity by charging the monopoly price p_M worldwide. Finally, and most relevantly, suppose that all countries in the South except country i set a price ceiling of \tilde{p} . Will country i be served if it sets a lower ceiling? If it does so, then i will have the lowest price and, by arbitrage, the price p_i will also prevail in the North. The other Southern countries have national exhaustion, so they will continue to pay \tilde{p} . By serving country i at price p_i and the remaining Southern countries at \tilde{p} , a Northern firm earns

$$(m_S + M_N) \pi(p_i) + (n_S - 1) \pi(\tilde{p}) < (M_S + M_N) \pi(\tilde{p}) = M_N \pi(p_m).$$

So the firm does better by serving none of the Southern countries than by serving some or all of them. In other words, if all Southern countries set a price ceiling of \tilde{p} , none has any incentive to lower its ceiling still further, because doing so would cause its market (and all others in the South) to be excluded from sales of innovative products.

Next we derive a new formula for p^* , the optimal price ceiling in a typical Southern country. We are looking for a symmetric equilibrium in which all Southern countries set p^* . But it must be a Nash equilibrium, which means that no Southern country has an incentive to deviate *when the other countries set $p^* \geq \tilde{p}$* . So, we assume that all Southern countries except i set $p^* \geq \tilde{p}$ as their price ceiling and examine a deviation by country i . When evaluated at p^* , such a deviation must not be profitable.

The tempting deviation for country i is to cut its price ceiling below p^* . If it does so, its price will be the lowest among all countries in the South and, by arbitrage, its price will also prevail in the North. But its price will not spill over to the other Southern countries, because, by assumption, these countries do not allow parallel imports. The trade-off facing country i is between the gain in consumer surplus that results from a lower price and the slowdown in innovation that occurs when firms earn lower profits in one Southern country *and in the North*. This trade-off is expressed in the marginal net benefit to country i from a small change in p_i , which is

$$\frac{d\phi_N}{dv} \frac{dv^{ie}}{dp_i} [TC(p_i) + (\bar{T} - T) C_c] - \phi_N T x(p_i).$$

The marginal profit change is $(M_N + m_s) \pi'(p_i)$ and so $dv^{ie}/dp_i = T(M_N + m_s) \pi'(p_i)$. Substituting, as before, and evaluating at $p_{-i} = p_i = p^*$ (because the countries besides i choose p^* by assumption, and we are looking for a symmetric Nash equilibrium where country i also

finds it optimal to set $p_i = p^*$) we find that the marginal net benefit vanishes when

$$\frac{\beta}{1-\beta} \left[C(p^*) + \left(\frac{\bar{T} - T}{T} \right) C_C \right] \left[\frac{\pi'(p^*)}{x(p^*)} \right] = \left[\frac{1+m}{\frac{1}{n_S} + m} \right] \pi(p^*) . \quad (18)$$

Notice that this gives the same formula for p^* as (12) when $n_S = 1$ and that p^* is increasing in m when $n_S > 1$. An increase in the relative size of the North increases p^* , because each Southern country affects world innovation to an extent that depends on the share of the North plus one Southern market in a firm's global profits. As n_S rises, p^* falls. This again signifies the declining influence on innovation of the price ceiling in a typical Southern country as the number of countries grows. So the flat portion of p_S^{ie} falls, and of course, $\tilde{p} > p$ for a wider range of m 's. Finally, note that when $m = 0$, the formula for p^* and for p_S^{ne} are the same. So the p_S^{ne} and p_S^{ie} curves continue to emanate from the same point; see Figure 4.

The figure shows that the Southern countries set higher price ceilings and accept lower consumer surplus when the North allows parallel imports than when it prohibits such trade, no matter how many countries are located in the South. There are two parts to the explanation for this finding. First, with international exhaustion of IPR in the North, each Southern country must set a price ceiling of at least \tilde{p} (which is always greater than c) in order to ensure that Northern firms are willing to supply its market despite the arbitrage that occurs. In contrast, when national exhaustion of IPR prevails in the North, the only constraint on the price ceiling in a Southern country is that it must not fall below the marginal cost c at which goods are produced. Second, with national exhaustion of IPR in the North, the policies in a small Southern country have limited impact on innovation, because firms earn a small (perhaps tiny) share of their profits there. With international exhaustion of IPR, the incentives facing a Southern government bite when its price ceiling is the lowest among all Southern countries, in which case its policies determine the price not only in its own small market, *but also in the North*. Therefore, each country's potential influence over world innovation remains vital even as its own market becomes small.

We shall not repeat the remaining analysis of Section 3. Suffice it to say that a variant of Lemma 1 continues to hold under the same Condition 1.²⁷ That is, if Southern countries do not allow parallel imports, a patent holder earns greater aggregate profits when the North practices international exhaustion of IPR and the Southern countries set the price ceiling p^* defined in (18) than when the North practices national exhaustion of IPR and the Southern countries set their price ceilings at $p_S^{ne} > c$. Also, Proposition 2 applies with any number of

²⁷The proof parallels that in Appendix A, except that now $k \equiv -n_S \pi'(p^*) / \Omega'(p^*)$, $\hat{\Omega}(p) \equiv k [\Omega(p) / n_S - \pi(p^*)]$, and $\hat{\pi}(p) \equiv \pi(p^*) - \pi(p)$. A detailed proof is available from the authors upon request.

countries in the South; with national exhaustion of IPR throughout the South, the North fares better by allowing parallel imports than by prohibiting them. Again, the key is the different behavior of the Southern governments under the alternative parallel trade regimes in the North. When the North practices international exhaustion of IPR, the countries in the South set higher price ceilings than under national exhaustion, which in turn leads to aggregate profits that are at least as high for Northern firms and innovation that is at least as fast. Since arbitrage delivers lower prices and greater consumer surplus in the North, $W_N^{ie} > W_N^{ne}$ no matter what is the size of a typical country in the South.

Note finally that the same conclusion applies in the limit, as the number of countries in the South grows large. With a continuum of small Southern countries, national exhaustion of IPR in the North leads the Southern countries to set a price ceiling of $p_S^{ne} = c$. International exhaustion of IPR in the North leads the Southern countries to set a price ceiling of $p_S^{ie} = \tilde{p}$.²⁸ By the definition of \tilde{p} , aggregate profits per firm equal $M_N\pi(p_M)$ in either case, which means that the value of patent and the rate of innovation are the same in either case. But the North enjoys greater consumer surplus at the lower price of \tilde{p} that prevails with parallel imports than at the higher price of p_M that prevails without them.

5.2 International Exhaustion of IPR in the South

The argument is simpler and more straightforward when the various countries in the South all permit parallel imports. Consider the choice of price ceilings by the Southern governments when the South practices international exhaustion of IPR and the North does not. Again we look for a symmetric equilibrium in which all Southern governments choose the price ceiling p_S^{ne} . Starting from a situation in which all but one of the Southern countries sets their ceilings at some p_S , the policy in the remaining country i will affect prices and allocations if and only if $p_i \leq p_S$; otherwise, parallel imports from the other Southern countries will make the choice of a high price ceiling by country i irrelevant. But if $p_i \leq p_S$, a marginal change in policy by country i will affect prices *in all countries in the South*. Therefore, the influence of country i on world innovation is exactly the same in these circumstances as would be the case if the South were one large country. It follows that the optimal choice of price ceiling for country i is the same as with one Southern country; the implicit formula for p_S^{ne} in a symmetric Nash equilibrium is given by equation (9).

Now suppose that the North permits parallel imports, so that international exhaustion of IPR prevails worldwide. It should be apparent that \tilde{p} —the minimum price ceiling in a

²⁸To make this statement, we need to show that $p^* < \tilde{p}$ for all $m > 0$ as n_S tends to infinity. First we note that $p^* = \tilde{p} = c$ when $m = 0$. By totally differentiating the implicit definitions for \tilde{p} and p^* that apply as $n_S \rightarrow \infty$, we find that $d\tilde{p}/dm > dp^*/dm$ for all m . It follows that $\tilde{p} > p^*$ for $m > 0$.

Southern country that allows for positive consumption—is the same as before. If all Southern countries but country i set a ceiling of \tilde{p} as given in (10), and i sets its ceiling at $p_i < \tilde{p}$, a producer of innovative products earns higher profits by selling only in the North at the monopoly price p_M than by selling worldwide at the common (due to arbitrage) price p_i . Moreover, the price ceiling p^* that prevails in a symmetric equilibrium when m is not too large also is the same as when there is only one country in the South. If all Southern countries other than country i were to set the p^* given by (11) as their maximum price, a lower ceiling of $p_i < p^*$ would generate a world price of p_i in all markets via parallel trade, and thus would lead to lower welfare for country i than if it set a ceiling of p^* . And if all countries other than country i were to set a common price ceiling above the p^* given by (11), the best response for country i would be to set a ceiling of p^* .

In short, with tolerance of parallel imports throughout the South, the outcomes based on what the North chooses as its parallel trade regime are the same in any symmetric equilibrium as those that would arise were the South to be a single, unified country. Each Southern country’s choice of price ceiling matters only when it is the lowest one, and when that is the case the policy determines prices and profit opportunities throughout the region. It follows from the analysis in Section 3 that the North fares better by allowing parallel imports than by forbidding such imports when the South uniformly practices international exhaustion of IPR, just as when it practices national exhaustion. Evidently, our results do not require that the number of countries in the South be small.

6 Conclusions

Many countries impose price controls on pharmaceutical products to hold down the cost of medicine to national health plans and local consumers. Such controls have generated sizable differences in the prices of similar drugs in different countries. The price gaps in turn create incentives for arbitrage in the form of “re-importation” or parallel trade.

Current trade rules leave the legality of such trade in the hands of the importing country. In many of these countries, the question as to whether intellectual property rights should be exhausted nationally or internationally has been highly controversial. Consumer advocates argue for parallel importation as means to temper monopoly pricing, while the pharmaceutical lobby points to the adverse effects of price controls on the returns to research. By allowing parallel imports, they claim, a country would import its trade partner’s price controls and thereby undermine intellectual property and dampen incentives for innovation.

In this paper, we have examined the welfare economics of parallel trade in a setting in which such trade is motivated by differential price controls. Our analytical framework is

one in which the protection of IPR creates a static distortion due to monopoly pricing, but provides incentives for ongoing innovation. Our analysis takes off from the observation that governments face different incentives for regulating prices when parallel imports are allowed by their trade partners than when they are not. Therefore, an evaluation of the costs and benefits of allowing parallel imports should incorporate the indirect effects of the trade regime on the level of foreign price controls.

We considered a two-country world in which product development is highly (or entirely) concentrated in one country (“North”) and compared outcomes when this country practices international versus national exhaustion of its patent rights. We found that the less innovative country (“South”) always sets a more lenient price ceiling in the former regime than the latter. In other words, the prospect of arbitrage induces the low-price country to accept higher prices than otherwise. This is because arbitrage forces the country to export its price controls and so impedes its ability to free-ride on the research incentives provided by its trade partner. This prediction of the model is consistent with the observation by Kanavos and Costa-Font (2005) that the growth of parallel trade in Europe has been accompanied by a relaxation of price controls in low-price countries such as Portugal, Italy and France, which now tolerate prices closer to European averages than they did before.

We first considered a setting in which South does no R&D and North imposes no price controls. We provided a sufficient condition under which global profit flows per patent, the value of a patent, and the pace of innovation are higher when North allows parallel imports than when it does not. Since reimportation also reduces prices in North, this country enjoys both higher surplus per product and greater product variety when it tolerates parallel trade. In contrast, the legalization of parallel imports by North spells a welfare loss for South, where the pain of higher prices outweighs the gain of faster innovation. We showed that our sufficient condition is satisfied by all linear and constant-elasticity demand functions.

Relaxing our assumptions of no Southern innovation and no Northern price regulation introduced some caveats but did not alter the gist of our conclusions. With positive but limited innovation in South, North still fares better by allowing parallel imports to cross its borders. When both countries impose price ceilings non-cooperatively in a world without parallel trade, a switch in trade regime again raises the return to R&D. However, in this case, faster innovation may come at the expense of higher drug prices in both countries.

Our results also extend to situations with several or many countries in the South. We examined symmetric equilibria in a model with multiple, identical Southern countries. We considered both the possibility that these countries tolerate parallel imports and the possibility that they do not. In either case, the North fares better by allowing parallel imports than by prohibiting them. Even a small Southern country has incentive to set a reasonable

price ceiling when the North is open to parallel trade, because its policy influence extends via arbitrage to the large market of the North and because doing otherwise might lead Northern firms to set prices that preclude sales to its (small) market.

Lobbies for pharmaceutical firms have vigorously encouraged their trade authorities to negotiate the relaxation or elimination of foreign price controls. Our analysis suggests that legalizing the reimportation of prescription medicines, perhaps paradoxically, might further their cause.

References

- [1] Arfwedson, Jacob (2004), "Re-importation (Parallel Trade) in Pharmaceuticals," Institute for Policy Innovation Policy Report 182, IPI Center for Technology Freedom.
- [2] Barfield, Claude E. and Groombridge, Mark A. (1998), "The Economic Case for Copyright and Owner Control over Parallel Imports," *The Journal of World Intellectual Property* 1, 903-939.
- [3] Barfield, Claude E. and Groombridge, Mark A. (1999), "Parallel Trade in the Pharmaceutical Industry: Implications for Innovation, Consumer Welfare, and Health Policy," *Fordham Intellectual Property, Media & Entertainment Law Journal* 10, 185-265.
- [4] Cambridge Pharma Consultancy (2004), *Pricing and Reimbursement Review 2003*, Cambridge UK: IMS Health - Management Consulting..
- [5] Chard, J.S. and Mellor, C.J. (1989), "Intellectual Property Rights and Parallel Imports," *The World Economy* 12, 69-83.
- [6] Danzon, Patricia M. (1997), *Pharmaceutical Price Regulation*, Washington D.C.: The American Enterprise Institute Press.
- [7] Danzon, Patricia M. (1998), "The Economics of Parallel Trade," *Pharmacoeconomics* 13, 293-304.
- [8] Danzon, Patricia M. And Towse, Adrian (2003), "Differential Pricing for Pharmaceuticals: Reconciling Access, R&D and Patents," *International Journal of Health Care Finance and Economics* 3, 183-205.
- [9] Ganslandt, Mattias and Maskus, Keith E. (2004), "The Price Impact of Parallel Imports in Pharmaceuticals: Evidence from the European Union," *Journal of Health Economics* 23, 1035-1057.
- [10] Grossman, Gene M. and Lai, Edwin L.-C. (2004), "International Protection of Intellectual Property," *American Economic Review* 94, 1635-1653.
- [11] Kanavos, Panos and Costa-Font, Joan, "Pharmaceutical Parallel Trade in Europe: Stakeholder and Competition Effects," *Economic Policy* 44, 751-798.
- [12] Li, Changying and Maskus, Keith E. (2006), "The Impact of Parallel Imports on Investment in Cost-Reducing Research and Development," *Journal of International Economics*, forthcoming.

- [13] Malueg, David A. and Schwartz, Marius (1994), "Parallel Imports, Demand Dispersion, and International Price Discrimination," *Journal of International Economics* 37, 167-196.
- [14] Maskus, Keith E. (2000a), *Intellectual Property Rights in the Global Economy*, Washington D.C.: The Institute for International Economics.
- [15] Maskus, Keith E. (2000b), "Parallel Imports," *The World Economy* 23, 1269-1284.
- [16] Maskus, Keith E. (2001), "Parallel Imports in Pharmaceuticals: Implications for Competition and Prices in Developing Countries," Final Report to the World Intellectual Property Organization, Geneva.
- [17] Pecorino, Paul (2002), "Should the US Allow Prescription Drug Reimports from Canada," *Journal of Health Economics* 21, 699-708.
- [18] Rey, Patrick (2003), "The Impact of Parallel Imports on Prescription Medicines," University of Toulouse, manuscript.
- [19] Szymanski, Stefan and Valletti, Tommaso (2005), "Parallel Trade, Price Discrimination, Investment, and Price Caps," *Economic Policy* 44, 705-749.
- [20] Valletti, Tommaso M. (2006), "Differential Pricing, Parallel Trade, and the Incentive to Invest," *Journal of International Economics*, forthcoming.

7 Appendix

Proof of Lemma 1

Define $k \equiv -\pi'(p^*)/\Omega'(p^*)$, $\hat{\Omega}(p) \equiv k [\Omega(p) - \pi(p^*)]$, and $\hat{\pi}(p) \equiv \pi(p^*) - \pi(p)$. Note that $\hat{\Omega}(p^*) = \hat{\pi}(p^*) = 0$ and $\hat{\Omega}'(p^*) = \hat{\pi}'(p^*) = -\pi'(p^*)$. Thus, $\hat{\Omega}(p)$ and $\hat{\pi}(p)$ are tangent at $p = p^*$. Note too that $\hat{\pi}''(p) = -\pi''(p) > 0$, while $\hat{\Omega}''(p) = \Omega''(p) \geq 0$. If Condition 1 is satisfied, then $\hat{\Omega}(p) \leq \hat{\pi}(p)$ for all $p \in [c, p_m]$, with equality only at $p = p^*$.

Suppose Condition 1 is satisfied. Then

$$k [\Omega(p) - \pi(p^*)] \leq \pi(p^*) - \pi(p),$$

with equality only at $p = p^*$. Therefore

$$k [\Omega(p_S^{ne}) - \pi(p^*)] < \pi(p^*) - \pi(p_S^{ne}) \quad (\text{A1})$$

and

$$k [\Omega(p_M) - \pi(p^*)] < \pi(p^*) - \pi(p_M) . \quad (\text{A2})$$

But $\Omega(p_M) = 0$, which, with (A1) and (A2), implies

$$\frac{\pi(p^*) - \pi(p_S^{ne})}{\pi(p_M) - \pi(p^*)} > \frac{\Omega(p_S^{ne}) - \pi(p^*)}{\pi(p^*)} .$$

This in turn implies

$$\frac{\Omega(p_S^{ne}) - \pi(p_S^{ne})}{\pi(p_M)} > \frac{\Omega(p_S^{ne}) - \pi(p^*)}{\pi(p^*)} .$$

Recognizing that the left-hand side of this inequality is equal to m , we have

$$(m + 1)\pi(p^*) > \Omega(p_S^{ne}) = m\pi(p_M) + \pi(p_S^{ne}) .$$

Proof of Proposition 5

We distinguish five cases.

(i) $p^* > p_N^{ne}$

Since $p_S^{ie} \geq p^* > p_S^{ne}$, $(m + 1)\pi(p_S^{ie}) \geq (m + 1)\pi(p^*) > m(\pi p_N^{ne}) + \pi(p_S^{ne})$.

(ii) $p_N^{ne} > p_S^{ie} = p^* \geq \tilde{p}$ and $p_S^{ne} > c$

Define k , $\hat{\Omega}(p)$ and $\hat{\pi}(p)$ as in the proof of Lemma 1. We have seen there that if Condition

1 is satisfied,

$$k [\Omega(p) - \pi(p^*)] \leq \pi(p^*) - \pi(p)$$

for all $p \in [c, p_M]$, with equality only at $p = p^*$. Therefore

$$k [\Omega(p_N^{ne}) - \pi(p^*)] < \pi(p^*) - \pi(p_N^{ne}) . \quad (\text{A3})$$

Since $\Omega(p_N^{ne}) > 0$, (A1) and (A3) imply

$$\frac{\pi(p^*) - \pi(p_S^{ne})}{\pi(p_N^{ne}) - \pi(p^*)} > \frac{\Omega(p_S^{ne}) - \pi(p^*)}{\pi(p^*) - \Omega(p_N^{ne})} > \frac{\Omega(p_S^{ne}) - \pi(p^*)}{\pi(p^*)} .$$

This in turn implies

$$\frac{\Omega(p_S^{ne}) - \pi(p_S)}{\pi(p_N^{ne})} > \frac{\Omega(p_S^{ne}) - \pi(p^*)}{\pi(p^*)} .$$

Recognizing that the left-hand side of this inequality is equal to m , we have

$$(m + 1)\pi(p_S^{ie}) = (m + 1)\pi(p^*) > \Omega(p_S^{ne}) = m\pi(p_N^{ne}) + \pi(p_S^{ne}) .$$

(iii) $p_S^{ie} = \min[\tilde{p}, p_N^*]$ and $p_S^{ne} > c$

In part (ii) we established that Condition 1 ensures $(m + 1)\pi(p^*) > m\pi(p_N^{ne}) + \pi(p_S^{ne})$. Here, $\tilde{p} > p^*$ and $p_N^* > p^*$, so $p_S^{ie} > p^*$. Therefore, $(m + 1)\pi(p_S^{ie}) > m\pi(p_N^{ne}) + \pi(p_S^{ne})$.

(iv) $p_S^{ie} = p^* < p_N^{ne}$ and $p_S^{ne} = c$

The fact that $p_S^{ie} \geq \tilde{p}$ implies $(m + 1)\pi(p_S^{ie}) \geq m\pi(p_M) > m\pi(p_N^{ne}) = m\pi(p_N^{ne}) + \pi(p_S^{ne})$.

(v) $p_S^{ie} = \max[\tilde{p}, p_N^*]$ and $p_S^{ne} = c$

(a) Suppose that $\tilde{p} \geq p_S^{ie} = p_N^*$. Note that $p_N^* > p_N^{ne}$. Therefore $(m + 1)\pi(p_S^{ie}) \geq m\pi(p_S^{ie}) \geq m\pi(p_N^*) > m\pi(p_N^{ne}) = m\pi(p_N^{ne}) + \pi(p_S^{ne})$.

(b) Suppose that $p_N^* \geq p_S^{ie} = \tilde{p}$. Then $(m + 1)\pi(p_S^{ie}) = (m + 1)\pi(\tilde{p}) = m\pi(p_M) > m\pi(p_N^{ne}) = m\pi(p_N^{ne}) + \pi(p_S^{ne})$.