

Parallel Imports of Pharmaceutical Products in the European Union

Mattias Ganslandt[⌘]

and

Keith E. Maskus[⌘]

Please do not quote without permission.

Abstract

We study the effects of parallel trade in the pharmaceutical industry. We develop a model in which an original manufacturer competes in its home market with parallel-importing firms. The theoretical analysis results in two key hypotheses. First, if the potential for parallel imports is unlimited, the manufacturer chooses deterrence and international prices converge. Second, with endogenously limited arbitrage the manufacturing firm accommodates and the price in the home market falls as the volume of parallel trade rises. Simple empirical tests favor the accommodation hypothesis with a time lag. Using data from Sweden we find that the prices of drugs subject to competition from parallel imports increased less than other drugs during the period 1995-1998. Approximately 3/4 of this effect can be attributed to lower prices of parallel imports and 1/4 to lower prices charged by the manufacturing firm. Econometric analysis finds that rents to parallel importers (or resource costs in parallel trade) could be more than the gain to consumers from lower prices.

Keywords: Parallel imports, international arbitrage, drug pricing

JEL Codes: F12, I11, L12

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

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

1 Introduction

Parallel imports are legitimately produced goods imported legally into a country without the authorization of a trademark, copyright, or patent holder. The essential purpose of such trade is arbitrage between countries with different prices. For several years, parallel trade of pharmaceutical products has been an important issue for the European pharmaceutical industry and numerous policy institutions, including the European Commission, the European Court of Justice and Member States of the European Union. At the heart of the issue is the evident conflict between diverging price regulations in the different Member States, on the one hand, and the consequences of parallel trade, on the other.

The conflict between market integration and market regulation is widely recognized. The Commission notes in its Communication on The Single Market in Pharmaceuticals (1998) that "Unless parallel trade can operate dynamically on prices, it creates inefficiencies because most, but not all, of the financial benefits accrue to the parallel trader rather the health care system or patient." Even so the Commission concludes that "... parallel trade must equally be seen as an important driving force for market integration and, consequently, for achieving the Single Market."

More recently, both American consumers and policy makers have grown increasingly concerned about the relatively high prices of patented drugs in the United States. Several new measures, including direct regulation of drug prices as well as a policy to admit parallel imports, have been proposed by U.S. policy makers. In an effort to reduce drug costs for American consumers both the House and the Senate approved a measure in July 2000 that would permit pharmacists and wholesalers to import cheaper drugs from other countries (WSJ, 2000). In December 2000, however, the Clinton Administration refused to implement the bill but the high prices of patented drugs in US remains an issue.

The core problem with parallel imports is that the welfare effects are ambiguous. A policy that admits parallel imports of pharmaceuticals involves at least two important trade-offs. In terms of dynamic efficiency, there is a tension between two major public-policy objectives; innovation and development of new drugs on the one hand and welfare of patients, which may require cost-containment strategies, on the other. The research-intensive pharmaceutical industry relies heavily on patents, which provide a limited period of marketing exclusivity within which innovators may recoup their drug development costs. The

value of the patent depends on its scope, including definition of the geographical area over which patent rights are exhausted upon first sale. The narrower the area of exhaustion, the greater the scope for price-differentiation. Consequently, incentives to innovate are stronger at the expense of consumer welfare in high-price countries. In this context, permission of parallel imports could reduce incentives to innovate while consumers in high-price countries gain.¹

The second trade-off involves static efficiency, which is determined by the costs and benefits of parallel trade. The short-run welfare effect is more likely to be positive for a country with high pharmaceutical prices when the real resources used in arbitrage activities are small relative to the effects stemming from price moderation. It is worth noting that a positive static effect of parallel trade is a necessary but not sufficient condition for the total effect to be positive.

The tension between policy objectives is evident in EU case law². The principle of free movement of goods allows individuals or firms within the EU to trade goods across borders without the consent of the producer, while national patents and trademarks provide exclusive rights to the producer to determine where and how to put products on the market. The European Court of Justice has held that free circulation of goods takes precedence over protection of intellectual property rights. In *Merck v Stephar* (C 187/80) the European Court of Justice held that a patent holder marketing its product in two different member states cannot prevent arbitrage between the two local markets, despite differences in intellectual property protection in the two countries. Thus, exhaustion applies upon first sale anywhere in the EU. Moreover, varying degrees of price control across countries do not justify prevention of parallel imports from countries with more rigorous regulations to markets with less rigorous regulations, as found in *Merck v Primecrown* (joined cases C-267/95 and C-268/95). Furthermore, parallel importers have limited rights to use original trademarks in marketing their products (*Dior v Evora*, C-337/95, and *BMS and Others v Paranova*, joined cases C-427/93, C-429/93, and C-436/93). Finally, manufacturers cannot partition the single market by introducing a new variety in member states, which could have the effect of replacing market authorization for the prior variety, where its product is subject to competition from parallel imports (*Rhône-Poulenc Rorer*, case C-94/98).

However, exhaustion in the European Union has important limitations. Most

¹ Danzon (1998) argues that a segmented equilibrium with price-discriminating monopolies can be optimal from a welfare perspective.

² See Ganslandt and Maskus (2000) for further details.

importantly, it does not extend to countries outside the common market (EMI v CBS, case C-51/75 and *Silhouette v Hartlauer*, case C-355/96). Thus, the ECJ has established a principle of "community exhaustion" but rejected the idea of international exhaustion. Moreover, the principle of community exhaustion does not extend to cases where the goods are sold in a member state under a compulsory license, as established in *Pharmon v Hoechst* (C-19/84). To summarize, the EU system essentially mandates free parallel imports within its territory, despite the existence of national intellectual property regimes and price controls, so long as the manufacturer has placed the good voluntarily on the market.

Despite the theoretically ambiguous relationship between welfare and parallel trade in R&D-intensive industries, direct estimates of the costs and benefits of parallel trade are rare. The aim of this paper is, therefore, to study the effects of parallel trade in the pharmaceutical industry. Our attention is limited to the static impacts and we do not consider the dynamic effects on R&D in the long run. Thus, the analysis focuses on explaining the volume and costs of parallel imports, the entry of parallel importers, the strategic responses by pharmaceutical firms, and the effects of these activities on prices.

The motivation for this limitation is to keep the analysis tractable within the confines of available data. Note, however, that unless we find a positive static effect we cannot expect a net positive effect, including dynamic impacts, of parallel trade in pharmaceutical products.

The paper presents a simple two-country model of arbitrage. Income levels differ between the two markets and pharmaceutical companies have an incentive to price-discriminate between markets. It is assumed that price ceilings in the exporting nation are exogenously given and, accordingly, an optimal system of price regulation is not derived. In the first specification we assume that the potential volume of arbitrage is unlimited. In the second specification we derive an endogenous choice of maximum potential parallel imports. While the limit here is a choice variable, one motivation for this idea is that parallel importers could face problems in finding more than a limited quantity in export markets. A related contribution is Klepper (1992), who presents a simple model of a price-discriminating monopoly that is exposed to price controls in one market and faces limited arbitrage between markets.

The theoretical analysis yields a number of hypotheses. First, if the potential volume of arbitrage is unlimited, the manufacturing firm would deter parallel imports by reducing its price in the home market for products that are subject to possible parallel trade. Second, if the potential volume of arbitrage is small,

the manufacturing firm would accommodate parallel imports and the price in the home market would fall in the volume of actual parallel trade.

In the empirical part we informally test our theoretical hypotheses on data from the Swedish market during 1995-1998. The Swedish market provides a natural test for our theoretical hypotheses. Before 1995 Sweden prohibited parallel imports of pharmaceutical products. However, entry into the European Union, on January 1, 1995, required Sweden to allow them.

For this purpose we have compiled a unique set of data consisting of prices for 50 major pharmaceutical products, the sales of patent holders, the identity and time of entry of parallel importers, volumes of parallel imports, and the sources of parallel imports. Because we use data assembled on a detailed product level the price comparisons are not subject to the methodological problems with general price comparisons discussed in Danzon and Chao (1998). We find that prices in Sweden of drugs subject to parallel imports have fallen relative to other pharmaceutical prices, with the effect concentrated at the end of the period. However, in a restricted data set we find that parallel imports have not resulted in convergence of wholesale prices between Sweden and the countries providing the exports.

The rest of this paper is organized as follows. The theoretical model of parallel imports is presented in section 2. The hypotheses from the theoretical analysis are investigated in section 3 and section 4 contains concluding remarks.

2 The Economics of Parallel Imports

Consider a model with two markets - home and foreign - denoted h and f . Throughout the analysis the home market is the high-income country and the foreign market the low-income country. A representative consumer in market $i = h, f$ has income m_i , where $m_h > m_f$. Demand for a specific pharmaceutical product is

$$D_i(p_i) = \alpha_i \gamma_i p_i \quad (1)$$

for market $i = h, f$, where γ_i is proportional to the marginal utility of money and decreasing in m_i .³ It is assumed that no substitution with other pharmaceutical products or therapies is possible.

³This system of demand functions can approximately be obtained with a linear-quadratic utility function, i.e. $u(x, y) = ax + by - \frac{1}{2}x^2 - \frac{1}{2}y^2$; as long as the expenditure on x is a relatively small share of the consumers budget and we make the appropriate variable substitution, that is $\alpha = a/b$ and $\gamma = v'(m) = b$. It is assumed that $v'(m) > 0$ and $v''(m) < 0$.

The product is patented in both countries and produced by a single manufacturing firm at marginal cost c . For simplicity $c = 0$. The manufacturing firm sells the product in the home and foreign market at prices p_h^m and p_f^m , where the subscript denotes the market and the superscript refers to the firm. Individual arbitrage between the two markets is prohibited.

There is an autonomous government in the foreign country with a capacity to set a price cap in its own market without intervention. The price cap in market f is denoted \bar{p}_f . The price cap in the foreign market is always binding.

We assume that there is a small number of symmetric parallel-importing firms, n , with a total arbitrage capacity of k , which can be infinitely large or a binding constraint depending on the situation. The marginal cost of parallel trade is t . The parallel-imported product is a perfect substitute to the product sold by the original manufacturing firm directly to consumers in the local market. Firms compete in prices. The parallel-importing firms set a common price in the import market denoted p_h^a . If the price of a unit sold by the original manufacturer is identical to the price set by the parallel-importing firms, then the consumer buys the good from the original manufacturer. The manufacturing firm's demand in the import market is

$$D_h^m(p_h^m) = D_h(p_h^a) - k \quad (2)$$

The strategic interaction is modelled as a multi-stage game. In the first stage, each parallel-importing firm i orders a quantity k_i from a wholesaler in the foreign market. The order quantity k_i is immediately made known to all other firms. In the second stage, the manufacturing firm sets prices in both markets, p_h^m and p_f^m ; subject to the exogenously given price cap \bar{p}_f . Prices are made public. In the third stage, the parallel-importing firm sets a price in the import market, ships a quantity q_i^a and markets clear.

In the theoretical analysis we consider two situations. The first situation is a benchmark with unlimited arbitrage. More formally, the arbitrage capacity k_i is not binding in the last stage of the game.

In the second situation, arbitrage is limited. It is assumed that the order quantity in the first stage is a binding constraint on the volume of arbitrage. Parallel importing firms can choose to limit arbitrage and the total volume of arbitrage is endogenously determined.

2.1 Parallel imports lead to price convergence

A natural benchmark is an equilibrium in which potential arbitrage is unrestricted and the maximum volume is never a binding constraint on the quantity shipped by the parallel-importing firms. In this case these firms have an incentive to undercut the price set by the manufacturing firm in the home market as long as the price difference is larger than the trade cost. The manufacturing firm achieves no sales in the home market as long as parallel imports occur. The only way to obtain any sales in the home market is to deter parallel imports by setting a price in the home market that makes parallel imports unprofitable. Thus, we define a deterrence strategy as a manufacturer's price that precludes parallel imports. An alternative strategy is accommodation, in which the manufacturer permits the level of trade chosen by the parallel-importing firms and prices accordingly. When potential arbitrage is unlimited, deterrence is more profitable than accommodation. The deterrence strategy results in a per-unit revenue equal to the price in the foreign market plus the trade cost, while under accommodation the manufacturer sells only in the foreign market and its per-unit revenue is the foreign price. In other words, price convergence is a result of the manufacturer's response to potential arbitrage rather than actual parallel trade activities.

Proposition 1 Let the potential volume of parallel imports be unlimited. The manufacturing firm deters parallel imports by setting a price in the home market which makes arbitrage unprofitable. More precisely, $p_h^m = p_f + t$, in the unique equilibrium. The price in the foreign market remains unchanged.

Proof. If $p_h^m > p_f + t$ then the parallel importer sets a price $p^a = p_h^m - t > p_f + t$ and the profit of the manufacturing firm would be $D_h(p^a) p_f$ in the home market. But the deterrence profit is $D_h(p_f + t) (p_f + t)$ which is higher. Any price $p_h^m < p_f + t$ deters parallel imports but results in less profit than $p_h^m = p_f + t$. Hence, by contradiction $p_f + t$ is the unique equilibrium.

This result is worth commentary. The threat of parallel imports results in price convergence relative to the segmented equilibrium. More specifically, prices converge (up to variable trade cost) due to a price reduction in the home market while the price in the foreign market remains unchanged. Interestingly, potential arbitrage is sufficient to induce price convergence and no parallel imports occur in equilibrium. From a welfare point of view this is desirable as no real resources are used in arbitrage activities.

A related point is that both the price effect in the home market and price convergence are unrelated to the volume or share of parallel imports, and the change in average price in the home market is identical to the change of the manufacturer's price.

2.2 Costly trade activities

The effects of parallel trade are sensitive to the existence of a constraint on the maximum volume of parallel imports. In particular, a quantity constraint on parallel trade gives the manufacturing firm an incentive to accommodate rather than to deter parallel imports. The manufacturing firm prefers to sell a somewhat smaller quantity, specifically its residual demand, in the home market at a higher price instead of reducing the price sufficiently to preclude parallel trade.

In the accommodation equilibrium the price difference between the foreign and home markets admits profitable arbitrage in the last stage and the parallel-importing firms undercut the manufacturing firm's price in the home market sufficiently to sell its whole capacity. Working backwards we may focus on the home market in the first stage. The result from the last stage is a residual demand for the manufacturing firm's product in the home market and the manufacturing firm solves

$$\max_{p_h^m} (\alpha_i - \beta_h p_h^m - k) p_h^m + k p_f \quad (3)$$

The optimum price is

$$p_h^m = \frac{\alpha_i - k}{2\beta_h} \quad (4)$$

which we call the accommodation price. Interestingly, the accommodation price is falling in the volume of parallel trade, so long as parallel imports are positive. The equilibrium profit may be obtained by inserting the accommodation price in the profit function. Correspondingly, we obtain the deterrence profit by inserting the deterrence price, $p_h^m = p_f + t$, in the profit function. The two profit levels give a condition for deterrence

$$\frac{(\alpha_i - k)^2}{4\beta_h} + k p_f = (\alpha_i - \beta_h (p_f + t)) (p_f + t); \quad (5)$$

where the left hand side is the profit in the accommodation equilibrium and the right hand side is the profit from deterrence. For sufficiently small k , the profit

from accommodation falls in the volume of arbitrage.⁴ The profit from deterrence increases in the trade cost. In other words, accommodation is relatively more attractive when the potential volume of parallel trade is small and the trade cost is relatively high. More formally, we have the following result:

Proposition 2 For a sufficiently small k and t , the manufacturing firm chooses not to deter parallel imports. Moreover, the equilibrium price set by the manufacturing firm in the home market is falling in the actual volume of parallel imports.

Proof. Define the profit from accommodation minus the profit of deterrence to be

$$f(k; t) = \frac{(\bar{p}_h - k)^2}{4 - \bar{p}_h} + k\bar{p}_f - (\bar{p}_h - t)(\bar{p}_f + t);$$

which is (strictly) positive for $k = 0; t = 0$. For any k the function is negative for $t > \bar{p}_h - \bar{p}_f$. For any given t , it is negative for $k > \bar{p}_h$. From the theorem of intermediate values it follows that there exist a function $k(t)$ such that $f(k(t); t) = 0$ and for $k < k(t)$ accommodation is more profitable than deterrence. Moreover, in this equilibrium the manufacturer's price in the home market is falling in the volume of parallel imports as

$$\frac{dp_h^m}{dk} = -\frac{1}{2 - \bar{p}_h} < 0$$

which concludes the proof.

The main intuition for the result is straightforward. First, a small potential volume of parallel imports hurts the manufacturing firm's revenue less in the accommodation equilibrium than does a large volume. A high trade cost, on the other hand, makes deterrence relatively easy and, thus, less costly for the manufacturing firm. A combination of these two effects gives the main result. Second, the accommodation price falls in the volume of arbitrage for a simple reason. The direct effect of a price cut on revenues is smaller than in the segmented equilibrium since a share of the quantity is sold by the parallel-importing firms. The manufacturing firm does not internalize the effect on the parallel-importing firms' profit in reaching its pricing decision in the accommodation equilibrium.

In the accommodation equilibrium the manufacturing firm sets a price that is a function of the actual volume of arbitrage. A larger quantity of parallel imports results in a lower price set by the manufacturing firm, though it remains

⁴We derive the maximum level of k for which this is true in the next sub-section.

above the deterrence price. The proposition, therefore, emphasizes that actual parallel imports, rather than potential arbitrage, may be essential for the degree of international price convergence. In the accommodation equilibrium the real resources used in arbitrage activities are increasing in the volume of parallel imports.

Note that the cost of obtaining price convergence is, however, not a continuous function in the potential volume of arbitrage and trade cost. A large potential volume and high trade cost makes deterrence more attractive to the manufacturing firm. More formally,

Proposition 3 Let the volume of parallel import be limited to a maximum quantity of k and the cost of parallel imports be t . The manufacturing firm is more likely to deter parallel imports when the trade cost, t , is high or the potential volume of arbitrage, k , is large.

Proof. First, the accommodation profit is falling in k

$$\frac{d}{dk} \left(\frac{(\bar{p}_i - k)^2}{4 - \bar{p}_h} + k p_f \right) = \bar{p}_i (\bar{p}_h - \bar{p}_i) < 0$$

and for $\bar{p}_f + t < \bar{p}_h$, the deterrence profit is increasing in t

$$\frac{d}{dt} \left((\bar{p}_i - \bar{p}_h (\bar{p}_f + t)) (\bar{p}_f + t) \right) = \bar{p}_i - 2 \bar{p}_h (\bar{p}_f + t) > 0$$

and a combination of the two gives the proof.

The intuition for this result is similar to that for the previous proposition. A high trade cost makes deterrence relatively easy and a large volume of parallel imports makes accommodation costly. Hence, the manufacturing firm is more likely to deter if the potential volume of parallel imports is large and/or the trade cost is high.

The last proposition establishes a link between the potential volume of parallel imports and the price effect in the home market, but no parallel trade is actually observed in a deterrence equilibrium. The real cost to obtain full price convergence is, therefore, kept at a minimum level. It is worth noting that high trade costs make deterrence more likely but the price effect of parallel imports in the home market is smaller. The reason is that the manufacturing firm need not reduce its price as much to deter parallel imports in the home market, since the trade cost helps to block arbitrage activities.

2.3 Strategic behavior by parallel-importing firms

In the previous section we showed that the manufacturing firm deters parallel imports when the potential volume of parallel imports is sufficiently large. This is clearly not a desirable outcome for the parallel-importing firms, which prefer an equilibrium in which the potential volume is limited and the manufacturing firm accommodates. We now develop a formal justification for the existence of endogenously limited maximum parallel trade

First, a commitment to an upper level of parallel imports in the first stage would result in accommodation. Let \bar{k} denote the maximum level of parallel imports at which the manufacturing firm accommodates. The deterrence condition (5) gives the maximum potential volume of arbitrage:

$$\bar{k} = \frac{\sum_i 2^{-h} p_f i^{-h} t}{\sum_i 2 p_f i^{-h} t} \quad (6)$$

Second, a commitment to an upper level of parallel imports removes some of the parallel-importing firms' incentive to undercut each other in the last stage. Competition between these firms in the last stage is, therefore, less aggressive. Working backward, each parallel-importing firm non-cooperatively chooses to order a quantity k_i^a in the first stage. If we restrict our attention to outcomes in which the manufacturing firm accommodates, the first order condition is

$$\sum_i 2k_i^a - \sum_i k_i^a i^{-h} (p_f + t) = 0 \quad (7)$$

where $\sum_i k_i^a$ is the sum of the ordered quantities by all other parallel-importing firms than i . Considering only symmetric equilibria gives a total optimal ordered quantity k_n^a , which is

$$k_n^a = \frac{n}{n+1} t \left[\sum_i 2^{-h} (p_f + t) \right] \quad (8)$$

The profit-maximizing quantity, however, must be lower than the maximum potential volume of arbitrage or else parallel imports are deterred by the manufacturing firm. The maximum volume of arbitrage is not binding for low variable trade costs. The profit-maximizing volume of arbitrage is less than the maximum volume of arbitrage, if $t < \bar{t}$ where

$$\bar{t} = \frac{\sum_i 2^{-h} p_f i^{-h}}{2^{-h} (2n^2 + 2n + 1)} \quad (9)$$

The critical level \bar{t} is falling in the number of parallel-importing firms. The reason is that the profit-maximizing volume increases with the number of parallel-importing firms while the maximum volume remains unchanged.

We summarize the formal results in the following proposition:

Proposition 4 Let the ordered volume in stage one be a binding constraint on parallel imports in stage three. In a symmetric equilibrium with n parallel-importing firms the total volume of parallel imports is $k = \min \{\bar{k}; k_n^a\}$ and the manufacturing firm accommodates. In this equilibrium, the price in the home market is falling in the variable trade cost t and falling in the number of parallel-importing firms.

Proof. The parallel importing firm prefers the accommodation equilibrium i^a

$$p_n^m(k) - p_f - t > 0 \quad (10)$$

which is equivalent to $k < \frac{2n}{n+1} (p_f + t)$. The maximum total volume of parallel imports in the accommodation equilibrium is \bar{k} ; given by equation (6); and the optimal non-cooperative volume is k_n^a ; given by equation (8); subject to the condition $k_n^a \leq \bar{k}$. Now, for all p_f and t we have $k > \min \{\bar{k}; k_n^a\}$, which guarantees that the parallel-importing firms prefer the accommodation equilibrium over the deterrence equilibrium.

To show the last part of the proposition we can use proposition 2 in which the relationship between the home market price and the volume of arbitrage was established. Proposition 2 shows that the price set by the manufacturing firm in the home market falls in the volume of parallel imports. Now, the quantity of parallel imports depends on the variable trade cost and the number of firms. More precisely,

$$\frac{dk}{dt} = \begin{cases} < 0 & \text{if } t < \bar{t} \\ > 0 & \text{if } t > \bar{t} \end{cases} \quad (11)$$

and

$$\frac{dk}{dn} = \begin{cases} > 0 & \text{if } t < \bar{t} \\ 0 & \text{if } t > \bar{t} \end{cases} \quad (12)$$

which concludes the proof.

The result is interesting for several reasons. First, a relatively low variable trade cost makes deterrence costly and the maximum volume of arbitrage in the accommodation equilibrium is high. The profit-maximizing parallel-importing firms, however, prefers to restrict parallel imports to maintain a higher margin on the volume of parallel imports that actually take place. The volume of

parallel imports is increasing in the number of firms due to the non-cooperative interaction between the firms. Each such firm will choose its quantity without regarding the effect on the profits of other parallel-importing firms.

Second, as the variable trade cost increases the effect of parallel imports on the price in the home market is weaker and the equilibrium price is higher. For a relatively low variable trade cost, the reason is that the firms choose a lower volume of arbitrage. The intuition is that the marginal revenue remains unchanged while the marginal cost increases, which results in a lower aggregate volume of parallel imports. For a relatively high variable trade cost, on the other hand, the reason is that deterrence is relatively less costly. The maximum accommodation volume is, therefore, falling in the variable trade cost and the equilibrium price in the home market is higher.

Our two cases are illustrated with Figures 1a and 1b. In both diagrams, the manufacturer sets the monopoly price p_h^m in the segmented equilibrium. In so doing, the manufacturing firm generates quasi-rents equal to the area of the box below the line extending from p_h^m to point A. These quasi-rents may be used to cover costs of R&D. In the deterrence equilibrium in Figure 1a, the manufacturer is forced to reduce price to the foreign price plus tariff. The result is a transfer from the manufacturer to consumers of area $(\alpha + \beta)$ and a consumer efficiency gain of β . Furthermore, area $(\gamma + \delta)$ reflects additional rents to the manufacturer from the higher sales volume. Here, area γ ordinarily might be resources wasted in trade but there is no actual trade in equilibrium. Although static welfare is higher in the home market, this simply reflects the gains from wider dissemination of a monopolized good. Because quasi-rents in the deterrence equilibrium are lower than in the monopoly case, the issue for policymakers is whether the loss in quasi-rents, and the implied reduction in future R&D, outweighs the static gains to consumers.

In the accommodation equilibrium in Figure 1b, the parallel-importing firms sell the quantity k in the home market. The manufacturer acts as monopolist on the residual demand, generating an equilibrium at point C, with the manufacturer selling volume q_m^h . The consumer efficiency gain is area β . The manufacturer loses quasi-rents equal to the box to the left of point A plus area δ_1 (transferred to parallel-importing firms) and area γ_1 (lost to trade costs). However, that firm gains quasi-rents of area \hat{A} on additional foreign sales to support the parallel import volume. In this case profits of area δ_2 are generated for the parallel-importing firms on additional consumption in the home market (with their total rents equaling area $\delta_1 + \delta_2$). Resources used in transport are the

area $(\lambda_1 + \lambda_2)$. The manufacturer would be indifferent between deterrence and accommodation where (for a given t), the parallel imports k are large enough to set residual marginal revenue equal to zero at price $p_f^{\wedge} + t$.

For lower import volumes the manufacturer prefers accommodation and for higher import volumes the manufacturer would deter. Because the parallel-importing firms understand this trade-off, they would choose a limited trade volume. While there are necessarily static overall gains from deterrence (if not necessarily net dynamic gains), the home economy may be better off or worse off strictly in static terms from accommodation. Note that it matters for home welfare whether the parallel-importing firms are located abroad or at home. If they are home firms, the net static welfare effect at home is $(\frac{1}{2} + \lambda_2 - \lambda_1 - \lambda_2 + \tilde{A})$, whereas if they are foreign firms, the net static welfare effect at home is $(\frac{1}{2} - \lambda_1 - \lambda_1 + \tilde{A})$, assuming the parallel-importing firms absorb the transport costs.

3 Empirical Analysis

We now proceed to an empirical analysis of the price effects of parallel imports. For this purpose we have collected detailed data about the pharmaceutical market in Sweden. The Swedish market provides a natural test for our theoretical hypotheses. As Sweden joined the European Union on the 1st of January 1995 the policy on parallel imports was drastically changed. Before 1995 parallel imports of pharmaceutical products were prohibited but they were allowed January 1 of that year.⁵

The main data sample was provided by LIF, the Swedish Association of the Pharmaceutical Industry. It includes panel data for the 50 molecules with the highest sales values in the Swedish market in 1994, 1995, 1996, 1997 and 1998. Approximately 35-38 percent of the Swedish pharmaceutical market in value terms is included in the data set.

An observation is a "product", defined as a patented molecule with a specific anatomical therapeutic classification (ATC) code, form and concentration. More specifically, the data include 164 different forms and concentrations distributed over 50 patented molecules. For each product there are quantities and prices for both original manufacturing firms and all parallel importing firms on a yearly basis.

⁵Parallel imports are subject to a regulatory approval period, which effectively delayed the regime shift.

Moreover, we have collected data on approvals to parallel import the products in our sample for 1995-1998. An approval is a formal decision by the Swedish Medical Products Agency and it allows a specific parallel importing firm to import a unique patented molecule in a specific form and dose from a specified export country.

In the supplemental sample we have detailed prices for 26 of the 50 molecules in two main parallel-exporting countries, Italy and Spain, for 1995 and 1998. These prices were obtained from IMS Health, a private consulting firm. The narrow sample was restricted to drugs that are precisely comparable on an international basis. Thus, they have the same ATC code, form, and concentration and originate from the same manufacturers across all markets.

The remaining parts of this empirical section are an overview of the pharmaceutical market in Sweden, an empirical analysis of the price effects in the Swedish market and, finally, a test of the price-convergence hypothesis between the export and import market.

3.1 The Pharmaceutical Market and Parallel Imports

The Swedish pharmaceutical market was approximately 0.8-0.9 percent of GDP valued at wholesale prices during the period 1995-1998. It ranged from 13.393 million Swedish Kronor (SEK) in 1995 to 16.567 million SEK in 1998, as summarized in table 1. Sales were relatively concentrated in a number of patented molecules. The 50 highest-sold molecules accounted for a third of the total market. In 1998 these 50 top drugs accounted for 37 percent of total sales in the pharmaceutical market valued at wholesale prices.

Parallel imports have increased substantially since Sweden joined the European Union, both in terms of actual sales and approvals to engage in the activity. In 1995 no parallel imports occurred and no applications to import were filed. By 1998 parallel imports had grown to 1.007 million SEK, which corresponded to 6 percent of the total market, and 226 approvals to import pharmaceutical products had been granted by the Medical Products Agency. In 1995 no firms entered the market and in 1996 still only one firm was granted permission to start parallel trade. However, by 1997 the number had grown to four firms and in 1998 the number was ten. The market was nevertheless dominated by a small number of parallel importers. The largest parallel-importing firm accounted for 100 percent of total parallel trade in 1996, 85 percent in 1997 and 59 percent in 1998. The four largest companies accounted for 96 percent of all parallel trade

in 1998.

[SEE TABLE 1]

Parallel imports from 13 different countries had been approved by 1998. The source of parallel imports was, however, heavily concentrated in a few low-price countries in Southern Europe. Two countries, Spain and Italy, were the exporters in approximately two thirds (63 percent) of all cases. The three most important export countries, Spain, Italy and Greece, accounted for 74 percent of all approvals.

In 1998 parallel trade accounted for 6 percent of the total pharmaceutical market in Sweden, which was mainly concentrated in the largest products. Sixteen percent of the sales of the top 50 molecules was parallel imports, but the distribution of parallel trade as a share of total sales for these 50 molecules was not equal. Measured on a unweighted product basis the median share of parallel trade was 0.0 percent and the maximum 72.1 percent. Parallel imports existed in approximately 15 percent of all products. Ranking products from lowest to highest parallel-import shares, at the 95th percentile such imports accounted for 53.6 percent of total sales.

A similar picture appears for approvals on a product basis. Approvals to parallel import drugs into Sweden in 1998 were concentrated in a few high-value products. For 68.3 percent of all products on the top 50 list no entry of a parallel-importing firm had occurred. More than one approval had been granted for 21.9 percent of products.⁶

To summarize this overview we conclude that our sample covers approximately 38 percent of the pharmaceutical market in Sweden. The growth of parallel imports from 1995 to 1998 was considerable and such imports accounted for 16 percent of sales in 1998. A large and rapidly growing number of parallel-importing firms entered the market, in some cases to sell the same products as other firms already present. Overall, sales are largely accounted for by four major parallel importers. Parallel trade is concentrated in a minority of the products in the sample but the share of parallel imports is considerable for approximately 15 percent of the top-value products. Italy and Spain are the source countries for 63 percent of parallel imports.

⁶Eighteen approvals had been granted for one specific product. That this number could be so large has two explanations. First, for some products a specific parallel-importing firm was approved to import the goods from several exporting countries. Second, more than one parallel-importing firm had an approval to import certain drugs, possibly from the same exporting country as other such firms.

3.2 Price effects in the Import Market

Next, consider the effect of parallel imports in the Swedish market. We start with a comparison between products which are subject to parallel imports and products which are not. For this purpose we calculate the relative price change for a specific product between a base year (1994 or 1997) and 1998. The relative price change is defined as the price in SEK in 1998 divided by the corresponding price in the base year minus one. We calculate the change for the average price including parallel imports, referred to as the "Mean incl. PI", as well as the change for the manufacturing firms' prices.

Table 2 reports the unweighted and weighted average price changes for all products. The unweighted average is a simple arithmetic mean. In the weighted average, however, product weights are computed as the product's sales in 1998 divided by the sum of sales for all products included in the average.

[SEE TABLE 2]

Our discussion will primarily focus on the unweighted average. Over the whole period 1994-1998 prices increased on average 6.64 percent for all products and manufacturing firms' prices increased somewhat more at 7.34 percent. On average prices for products subject to parallel imports increased 2.88 percent while manufacturers raised their prices 6.38 percent for these products. As a contrast, prices for products not subject to parallel trade rose 7.57 percent.

The difference was even more pronounced over the shorter period 1997-1998. Average prices increased 0.25 percent. Average manufacturing firms' prices declined 0.34 percent for products subject to parallel trade but rose 0.95 for products not subject to parallel imports. Prices of parallel-imported products fell on average by 3.12 percent.

This first overview seems to confirm that prices of parallel-traded products, and products facing such actual competition, fell in the import market relative to the prices of products not subject to parallel trade. The main effect, approximately three quarters of the fall, results from parallel trade itself while the remaining effect is the change in manufacturing firms' prices.

We undertake simple tests of the hypotheses generated by our two models. First, to test whether the differences between the change in the manufacturing firms' prices for products subject and not subject to parallel imports are significant, we perform t-tests, assuming unequal variances, of the hypothesis that the mean change is the same. The hypothesis that the manufacturing firms' price

changes for goods facing parallel imports and those not facing such imports is the same cannot be rejected at the ten-percent level of significance for the period 1994-1998 ($t = 0.4324$). The hypothesis is, however, rejected at the five-percent level for 1997-1998 ($t = 1.7679$), which confirms that the manufacturing firms' prices increased significantly less for products subject to parallel imports than did prices of other products in the end of the period.

An additional t-test confirms that the mean price change for parallel-imported products was significantly lower than the mean for non-parallel-imported products for 1994-1998 ($t = 1.6854$) and 1997-1997 ($t = 4.8160$).

A third hypothesis, that the average change of parallel-imported goods and the manufacturer's price change for non-parallel-imported products is the same during 1994-1998 cannot be rejected ($t = 1.0255$). However, this difference becomes significant at the one-percent level for 1997-1998 ($t = 2.6297$).

Our preliminary conclusion is, therefore, that the data support the model of accommodation rather than deterrence since the change of the manufacturing firms' prices is not significantly different between goods facing parallel imports and other goods. However, the mean price, including both parallel-imported goods and manufacturers' prices in those goods, increased significantly less than the manufacturing firms' prices not facing such competition. Over the short period 1997-1998, the change of the manufacturing firms' prices was significantly lower for products facing import competition than for other drugs. This result suggests that manufacturing firms react to the volume of arbitrage with a lag, rather than trying to deter parallel imports before they actually enter the market.

To investigate this finding further we examine statistically how changes in manufacturers' prices are affected by the share of parallel trade and the approval of parallel imports. We define PI SHARE, which takes values between 0 and 1, as the share of parallel trade in total sales for a specific product. The variable APPROVAL is a dummy equal to one when there is at least one approval in 1998 to parallel import the product and zero otherwise. The dependent variable, defined at the individual product level, is the relative price change of the manufacturing firms' price over the periods 1995-1998, 1996-1998 and 1997-1998.

An ordinary least squares regression finds that the coefficients of PI SHARE and APPROVAL have the expected negative sign in every case, as reported in table 3. The coefficients are, however, insignificant for the longer period 1994-1998. For the shorter period 1997-1998 the coefficient on PI SHARE is

-0.0393 and statistically significant on the 10-percent level. The coefficient on APPROVAL is -0.0125 and significant at the 10-percent level. Thus, the regressions further support the model of accommodation with a lagged reaction to parallel imports. Although the effect of parallel imports on manufacturers' prices is significantly negative it is worth noting that the size of the effect is relatively small. Further, these variables do not explain much of the variation in prices, as the adjusted R^2 for the OLS-regression with PI SHARE and APPROVAL are 0.04 and 0.03, respectively.

[SEE TABLE 3]

3.3 Price Convergence and Rents

So far the empirical analysis has focused on the effects of parallel imports in the home market. More generally, however, we are interested in the effects of parallel trade on the price differential between the export and import markets. In both the deterrence and accommodation cases we expect prices between the two markets to converge.

To test this hypothesis we use bilateral price comparisons between the Swedish market and the two main export markets, Italy and Spain. Prices are wholesale prices, that is, prices quoted to pharmacies, in U.S. dollars in 1994 and 1998. These data include at least one pair of prices for 28 of the top 50 molecules. Nine of the products were subject to parallel trade from one or both of the export markets. Regional varieties in the Swedish market that lack comparable products in the export market were excluded from the sample.

Relative prices for Italy and Spain were calculated on a product-by-product basis for 1994 and 1998. The relative price for a specific product is defined as the dollar price in the export country divided by the dollar price in Sweden. The price change is defined as the relative price in 1998 minus the relative price in 1994. Price convergence with this definition occurs when the price change is positive.

The dependent variable is the change of the relative price in Italy or Spain for all bilateral comparisons. The exogenous variable is a dummy, PI TRADE, which is one when parallel trade between the export and import country occurs and zero otherwise.

The estimated coefficient for PI TRADE was 0.0180 for Italy, 0.0206 for Spain and 0.0176 for the pooled data. All three estimated parameters are insignificantly different from zero. Hence, the data do not support the hypothesis

that prices for products subject to parallel trade converge between the export and import countries. More specifically, we conclude that manufacturing firms' prices in Italy and Spain relative to Sweden did not converge during the period 1994-1998. Substantial price differences for parallel-traded products remained. Indeed, the average price across Italy and Spain of these goods was 68 percent of the average price in Sweden.

Moreover, we can approximately estimate the difference between the price in the export market and the price set by the parallel-importing firm in the import market. Table 4 summarizes the parallel-importing firms' prices for products subject to such imports. The relative price in 1998 is defined as the price set by the parallel-importing firm divided by the price charged by the manufacturing firm for the same product. The average relative price set by the parallel-importing firms in 1998 was 89 percent of the price set by the manufacturing firms in Sweden. The minimum relative price was 85 percent and the maximum 92 percent.

[SEE TABLE 4]

Given these figures, we estimate the parallel-importing firms' margin to be approximately 21 percent of the original manufacturer's price in the Swedish market. The margin for parallel imports from Italy ranged from nine to 39 percent while the margin for parallel imports from Spain ranged from nine to 31 percent. Rents to the parallel-importing firms, or alternatively real resources used in the arbitrage activities, are, therefore, considerable compared to the price reduction in the home market.

Finally, we use the margin and price effects to estimate the impact of parallel trade on consumer surplus and the rents that are shifted from manufacturing firms to parallel-importing firms. The effect of parallel imports on the manufacturing firms' prices is -1.2 percent for the unweighted average and -3.3 percent for the weighted average, compared to products not subject to parallel trade over the period 1994-1998. Assuming that pharmaceuticals are normal goods, we obtain an upper bound of the positive effect on consumer surplus in 1998 by using the quantity consumed at the lower, parallel-import-induced prices, evaluated at fictitious prices had there been no parallel imports. The difference between the fictitious expenditure and the actual expenditure is the effect on consumer surplus. In this regard, the static effect of parallel imports on consumer surplus in 1998 is estimated to be a gain of 150 million SEK with unweighted price changes and 199 million SEK with weighted price changes.

The rents to parallel-importing firms (which include costs of the activity) are calculated using the actual margin between these firms' prices in Sweden and wholesale prices in the export market (Italy or Spain), multiplied by the quantity of parallel-imported drugs. In cases where the export price is missing it is replaced with the average export price, which is 68 percent of the manufacturing firm's wholesale price in Sweden. Using these margins, we calculate the rents to parallel importers to be approximately 188 million SEK in 1998. Note that these rents are of the same magnitude as the consumer surplus gain.

4 Conclusion

We developed a model of manufacturer behavior in the presence of potential parallel imports. In the deterrence outcome we expect the manufacturing firm's price of a product that is subject to such competition to drop in the home market, though no actual imports would occur. More specifically, we expect the price of products subject to potential arbitrage to drop relative to prices of products that are not subject to potential parallel imports in the home market. In the accommodation equilibrium we expect the manufacturing firm's price to fall as the volume of parallel imports rises.

The Swedish market provides a natural laboratory in which to investigate our theoretical findings. The growth of parallel imports from 1995 to 1998 was considerable and accounted for 16 percent of the sales in our sample in 1998. A large number of parallel-importing firms entered the market, in some cases to sell the same products as other such firms already present. However, parallel-trade sales were dominated by four major firms. Parallel trade was concentrated on a minority of the products in the sample but its share was considerable for up to 15 percent of major drugs. The source countries were primarily Italy and Spain.

The empirical analysis confirmed that prices of goods subject to import competition, including parallel-traded products themselves, fell approximately 4 percent in the import market relative to the prices of products not subject to parallel trade. Approximately three quarters of this reduction resulted from parallel trade itself while the remaining effect was the change in the manufacturing firms' prices. The simple tests for differences in price changes suggested that the accommodation reaction came with a lag, as the price reductions were significantly different in the shorter period 1997-1998 than in the full period.

The data fail to support the hypothesis that prices for products subject to parallel trade converge between the exporting and importing countries. Our results suggest that parallel-importing firms exploit a price difference between these markets of approximately 21 percent of the original manufacturer's price in Sweden. Rents to the parallel-importing firms, or alternatively real resources used in arbitrage, are, therefore, considerable compared to the price effect in the home market.

References

- [1] CEC (1998), "Commission Communication on the Single Market in Pharmaceuticals", COM(98)588 Final, Brussels, November 1998
- [2] Danzon, Patricia M. (1998), "The Economics of Parallel Trade", *PharmacoEconomics* v13, n3 (March): 293-304
- [3] Danzon, P. and L. Chao (1995), "The Uses and Abuses of International Price Comparisons", in Helms "Competitive Strategies in the Pharmaceutical Industry", AEI Press, Washington
- [4] Ganslandt, Mattias and Keith E. Maskus (2000), "Parallel Imports of Pharmaceutical Products in the European Union", mimeo, IUI, Stockholm, Sweden
- [5] Klepper, Gernot (1992), "Pharmaceuticals: who's afraid of 1992", CEPR discussion paper no. 675, March 1992
- [6] WSJ (2000), "Senate Approves Measure On Import of Cheap Drugs", *The Wall Street Journal*, July 20, 2000

Table 1: The Pharmaceutical Market in Sweden 1995-1998.

Sweden	1995	1996	1997	1998
Gross Domestic Product (MSEK)	1.649.922	1.688.200	1.738.859	1.816.042
Total pharmaceutical sales (MSEK)	13.393	15.808	14.263	16.567
Sales of top 50 molecules (MSEK)	4.576	5.977	5.201	6.203
Parallel imports (MSEK)	0	>0	269	1.007
Parallel imports of top 50 (MSEK)	0	>0	269	920
C50 by molecule	0.34	0.38	0.36	0.37
Parallel imports/Total sales	0.00	0.00	0.02	0.06
Parallel imports/Top 50 sales	0.00	0.00	0.05	0.16
Concentration ratio (C1)	n/a	1.00	0.85	0.59
Concentration ratio (C4)	n/a	1.00	1.00	0.96
Total number of P.I. approvals	0	1	45	226
P.I. approvals for top 50 molecules	0	1	31	131
Total number of P.I. forms	0	1	4	10

Source: Statistics on approvals compiled from data obtained from the Medical Product Agency (<http://www.mpa.se>), "Approvals" and sales statistics compiled from LIF data (<http://www.lif.se>).

Note: C50 is defined as the fraction of the sales of the top 50 molecules of total sales of pharmaceuticals. Sales are in nominal wholesale prices, i.e. price to pharmacy (AIP), incl vet and excl extempore. An observation of an approval is defined as a formal approval to a specific form to parallel import a unique patented molecule in a specific form and dose from a specified export country.

Table 2: Price changes of pharmaceuticals in Sweden 1995-1998.

		Unweighted average		Weighted average	
		1994-1998	1997-1998	1994-1998	1997-1998
Sweden					
All products	Mean incl. PI	0.06636 (0.1344)	0.00253 (0.0352)	0.00800 (0.0276)	-0.01370 (0.0061)
	Manuf.'s price	0.07336 (0.1330)	0.00731 (0.0308)	0.02791 (0.0217)	-0.00156 (0.0032)
PI products	Mean incl. PI	0.02881 (0.1213)	-0.03117 (0.0409)	-0.04384 (0.0448)	-0.03846 (0.0086)
	Manuf.'s price	0.06381 (0.1199)	-0.00343 (0.0349)	0.00308 (0.0360)	-0.00668 (0.0052)
Non-PI prod	Mean	0.07574 (0.1365)	0.00955 (0.0296)	0.03646 (0.0171)	0.00150 (0.0035)
No. of obs.		125	151	125	151

Source: Author's calculations based on data from LIF (<http://www.lif.se>).

Note: Std dev in parentheses. Data includes 164 different forms and dosages distributed over 50 patented molecules. The relative price changes are calculated for each unique package. The price change for all firms selling a unique product is a weighted average of the manufacturing firm's price and the parallel importing firm's price.

Table 3: The effects of P.I. on the manufacturing firm's prices.

	Manufactureur's price change in the import market from base-year to 1998					
	1994	1994	1994	1997	1997	1997
CONST.	0.0808 ^{***} (0.0152)	0.0747 ^{***} (0.0127)	0.0808 ^{***} (0.0151)	0.0115 ^{***} (0.0030)	0.0094 ^{***} (0.0026)	0.0115 ^{***} (0.0030)
PI SHARE	0.0073 (0.0830)	-0.0217 (0.0729)	-	-0.0266 (0.0186)	-0.0393 ^{**} (0.0163)	-
APPROVAL	-0.0206 (0.0279)	-	-0.0194 (0.0244)	-0.0084 (0.0059)	-	-0.0125 ^{***} (0.0052)
No. of obs.	125	125	125	151	151	151
Adj. R2	0.00	0.00	0.00	0.04	0.04	0.03

Source: Author's calculations based on data from LIF (<http://www.lif.se>).

Note: Std err in parentheses. Data includes 164 different forms and dosages distributed over 50 patented molecules. The relative price changes are calculated for each unique package.

Table 4: Prices of parallel imports and parallel importing firms' markups in 1998

Relative price in 1998	Price relative to manufacturing firm's price in Sweden					
	PI to Sweden from Italy			PI to Sweden from Spain		
	Mean (std. dev.)	Max Min	Obs.	Mean (std. dev.)	Max Min	Obs
PI price in Sweden	0.8917 (0.0125)	0.9155 0.8506	28	0.8917 (0.0125)	0.9155 0.8506	28
PI price in export market	0.6819 (0.1145)	0.8258 0.5095	7	0.6786 (0.0683)	0.7874 0.5919	8
PI markup	0.2140 (0.1136)	0.3890 0.0897	7	0.2116 (0.0749)	0.3071 0.0913	8

Source: Author's calculations based on data from LIF (<http://www.lif.se>).

Note: Std. dev. in parentheses. The relative PI price in Sweden is defined as the parallel importing firm's price relative to the price set by the manufacturing firm. The narrow sample includes molecules with price observations in the export market (Italy or Spain).