Center for
European,
Governance and
Economic Development Research

cege

Discussion Papers

Number 180 – December 2013

PHARMACEUTICAL REGULATION AT THE WHOLESALE LEVEL AND PARALLEL TRADE

Laura Birg

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

ISSN: 1439-2305

Pharmaceutical regulation at the wholesale level and parallel trade

Laura Birg*

December 2013

Abstract

This paper studies the effect of pharmaceutical regulation at the wholesale level, if markets are integrated by parallel trade, i.e. trade outside the manufacturer's authorized distribution channel. In particular, maximum wholesale margins, a restriction of pricing by the intermediary, and mandatory rebates, a restriction of the pricing by the manufacturer, are analyzed with respect to their effect on drug prices, quantities, and public pharmaceutical expenditure. Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. In a symmetric equilibrium, maximum wholesale margins of both countries partly offset each other. Mandatory rebates may be a policy alternative, as they exhibit a reinforcing effect with respect to drug prices.

JEL classification: F12, I11, I18

Keywords: parallel trade, regulation, maximum markups, spillovers, mandatory rebates

1 Introduction

This paper studies the effect of pharmaceutical regulation at the wholesale level, if markets are integrated by parallel trade, i.e. trade outside the manufacturer's authorized distribution channel. In particular, maximum markups, a restriction of pricing by intermediaries, are analyzed with respect to its effect on drug prices, quantities, and public health expenditure. As a policy alternative, mandatory rebates, mandatory discounts by the manufacturer, are suggested.

This analysis is motivated by the observation that pharmaceutical regulation and parallel trade are interdependent. Pharmaceutical parallel trade occurs in highly regulated markets. The continuous increase in public health expenditure in many countries over the last decades has

*Center for European, Governance and Economic Development Research, University of Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, laura.birg@wiwi.uni-goettingen.de. I would like to thank Horst Raff, Jürgen Zerth and seminar participants at Lausanne, Kiel, Sevilla, and Warsaw for helpful comments and suggestions.

induced a considerable number of government interventions (Maynard & Bloor, 2003). Consequently, pharmaceuticals markets are characterized by a variety of regulatory instruments that are partly overlapping and impede each other (see Espin & Rovira 2007 for an overview of regulatory interventions in the European Union).

Cross-country differences in regulation result in price differences, which are a precondition for engaging in parallel trade. The profitability of parallel trade depends on substantial price differences. In the European Economic Area, where parallel trade is legal, price differences of up to 300 % percent between countries can be observed (Maskus, 2000b; Glynn, 2009). They may stem from pharmaceutical manufacturer's price discrimination between different countries and/or differences in national pharmaceutical regulations in the individual member states (Kanavos et al., 2004; Enemark et al., 2006; EU Commission, 2003). Consequently, by creating price differences, pharmaceutical regulation may trigger parallel trade or determine the extent of parallel trade.

Parallel trade itself may induce regulation. That is, there is a direct response of regulation to parallel trade: Many destination countries provide incentives for patients to purchase lower-priced parallel imports (via the cost-sharing mechanism) or legal requirements to dispense parallel imported drugs, which ensures the sale of parallel imports for parallel traders (Kanavos et al., 2004). In addition, the regulatory authority in source countries of parallel imports may change its behavior, if parallel trade takes place: Under segmented markets, a price cap only affects the regulating country, but via the channel of parallel imports, a sufficiently low maximum price may reduce prices in other countries as well, amplifying the negative impact of a price cap on the manufacturer's profit. Taking this link into account, regulatory bodies refrain from setting prices too low, if the manufacturer can credibly threaten to refuse to supply the respective market. Königbauer (2004) and Grossman & Lai (2008) suggest this argument. In destination countries of parallel imports, the lower prices of parallel imports may reveal the information which price level is still profitable for manufacturers and regulatory authorities may adjust maximum prices downwards. At last, in countries with a strong pharmaceutical industry, where pharmaceutical regulation also takes industrial policy goals into account, regulation may respond to parallel trade, as it reduces manufacturers' profits.

In this paper, pharmaceutical regulation directly addresses price changes induced by parallel trade: Price interdependencies limit the manufacturer's ability to address the double marginalization problem created by vertical separation in imperfectly competitive markets by a two-part tariff. Pharmaceutical manufacturers do not sell directly, but through independent wholesalers (Taylor, Mrazek & Mossialos, 2004). The manufacturer and the wholesaler both add a markup to their marginal cost, without considering the impact of their pricing decision on the respective other actor (Rey & Verge, 2008). For separated markets, the manufacturer's optimal strategy to avoid the double marginalization problem is to specify a two-part tariff with a low wholesale price and a fixed fee that extracts the intermediary's profit. However, in the presence of parallel trade, the manufacturer raises both wholesale prices in response to parallel trade, creating a double marginalization effect with higher prices than in the absence of parallel trade. The first best solution to the double marginalization effect would be to stimulate competition or to enforce vertical integration. The first is impossible at the manufacturer level due to patent protection, large fixed cost of entry and economies of scale and scope inhibit entry at the retail level. The latter is prohibited by national regulation (Taylor; Mrazek & Mossialos, 2004). Therefore, the regulatory instrument analyzed here, maximum wholesale margins, attempts to address the double marginalization effect by restricting the markup surcharged by the intermediary. Maximum wholesale margins are applied in virtually all European countries. As a policy alternative, mandatory rebates are suggested. Mandatory rebates are a fictitious alternative, based on the Herstellerzwangsrabatt (compulsory manufacturer discounts) according to § 130a Social Security Code V in Germany, which force the manufacturer to grant a discount on the wholesale price.

The effects on drug prices and quantities in destination and source countries of parallel imports are studied in a two-country model following Maskus & Chen (2002) and Chen & Maskus (2005).

Parallel trade provides the manufacturer with the possibility to exploit the strategic effect of exclusive territories in the destination country. It generates a competition effect in the destination country and double marginalization effects in both countries, resulting in a higher price than from direct sales.

Maximum wholesale margins try to mitigate this effect by limiting markups of intermediaries. They also result in an adjustment of wholesale prices by the manufacturer. Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. In the symmetric equilibrium with both countries applying maximum wholesale margins, regulatory instruments exhibit an offsetting effect. This is, a restriction of pricing by intermediaries in the destination country reduces drug prices in the destination country, but raises the price in the source country. Similarly, a restriction of pricing by the intermediary in the source country results in higher prices in the destination country.

Mandatory rebates are a policy alternative that also addresses the double marginalization effect by restricting pricing. It also reduces drug prices in both countries. In the symmetric equilibrium with both countries applying mandatory rebates, regulatory instruments exhibit a reinforcing effect with respect to prices. This is, a restriction of the wholesale price in the destination country reduces drug prices in both countries.

The rest of the paper is organized as follows. Section 2 studies the equilibrium without regulation, section 3 examines the equilibrium with maximum wholesale margins. Section 4 analyzes mandatory rebates as policy alternative, section 5 concludes.

2 The Model

The structure of the model is based on Maskus & Chen (2002), (2005). Consider a (domestic) manufacturer M selling a brand-name drug b in two countries, its home country D and a foreign country S. In both countries, the manufacturer does not sell directly, but through an independent intermediary I_j $(j = D, S)$[1]. With respect to the intermediaries, the manufacturer adopts a two-part pricing strategy, it charges each intermediary a wholesale price w_j per unit and a fixed fee ϕ_j.

In a regime of international exhaustion of intellectual property rights, due to lack of complete vertical control, intermediaries may engage in parallel trade and resell the drug b in the respective other country as parallel import (hereafter noted as β). By assumption the foreign intermediary in country S, I_S takes advantage of this opportunity, but the domestic intermediary in country D, I_D does not (one-way parallel trade). Accordingly, the intermediary I_S exports the drug from country S and sells it in country D as a parallel import. That is, the foreign country is the source country of the parallel import and the home country is the destination country. Therefore, the home country will be denoted as country D and the foreign country as country S.

While consumers in country S buy the drug from the foreign intermediary I_S, consumers in country D have the choice between the locally sourced version b when purchasing from the local intermediary I_D and the parallel import β when buying from the foreign intermediary I_S. Consumers associate a lower quality with the parallel import, which is captured by a discount factor τ in consumer valuation. The perception of parallel imports as qualitatively inferior results from differences in appearing and packaging (Maskus (2000)). In addition, following Schmalensee (1982), uncertainty regarding product characteristics can be translated into quality differentials. If consumers are not sure whether the parallel import is identical with the locally sourced version of the drug, their willingness to pay for the parallel import will be lower and the intermediary must offer a price reduction in order to convince consumers to try and learn about the parallel import. Moreover, there is evidence that the price of a drug may serve as a quality indicator (Waber et al. (2008)). Accordingly, due to a lower price, the parallel import may be associated with lower quality.

Consumers in both countries are heterogeneous with respect to the gross valuation of drug treatment, represented by a parameter θ which is uniformly distributed on the interval $[0, 1]$. Thus, the total mass of consumers is given by 1 in both countries.

Each consumer demands either one or zero units of the most preferred drug. The utility derived from no drug consumption is zero, while a consumer who buys one unit of drug i obtains a net utility

$$U\left(\theta, \tau, \gamma_j, p_i\right) = \begin{cases} \theta - \gamma_j p_{i,j} & \text{if } i = b \\ \theta\left(1-\tau\right) - \gamma_j p_{i,j} & \text{if } i = \beta \end{cases} \tag{1}$$

where $\tau \in (0, 1)$ reflects the perceived quality difference between both versions b and β of the drug, γ_j is the coinsurance rate in country j $(j = D, S)$, and $p_{i,j}$ is the price of drug i in country j. For $\tau = 1$, consumers associate no value at all with the parallel import, for $\tau = 0$, both products are homogenous and are thus considered perfect substitutes.

A consumer with a positive net utility of drug consumption will choose the most preferred

[1] This setup differs from Maskus & Chen (2002), (2005), who assume that the manufacturer sells directly in its home country and through an intermediary in the foreign country.

drug version by trading off perceived drug quality against drug copayment. The higher the gross valuation of drug treatment θ, the more the consumer is willing to pay in order to purchase the (high-quality) locally sourced drug. The consumer heterogeneity with respect to valuation θ can be interpreted as differences in willingness to pay for a locally sourced version, differences in risk aversion regarding the trial of substitutes or differences in the severity of the condition or differences in prescription practices (see e.g. Brekke, Holmas & Straume (2010)).

In each country, health insurance reimburses a fraction $0 < \kappa_j < 1$ of the drug price, the remaining fraction $1 - \kappa_j = \gamma_j$ is paid by the patient. Thus, the effective price of the drug to the patient amounts to the proportion γ_j of the market price set by the respective intermediary I_j (Zweifel et al. (2009)).

I assume that the dispersion of coinsurance rates, i.e. price elasticities, across both markets is sufficiently low:

$$\gamma_S \leq \gamma_D \frac{4}{(1-\tau)(2-3\tau)}. \tag{2}$$

This ensures that the manufacturer finds it profitable to serve both markets in equilibrium.

In country D, consumers in country D have the choice between the locally sourced version (b) from the domestic intermediary I_D or the parallel import (β) from the foreign intermediary I_S. The marginal consumer who is indifferent between buying the locally sourced version b and the parallel import β has a gross valuation $\theta_D^{b,\beta}$, given by

$$\theta_D^{b,\beta} - \gamma_D p_{b,D} = \theta_D^{b,\beta}(1-\tau) - \gamma_D p_{\beta,D} \;\Leftrightarrow \theta_D^{b,\beta} = \frac{\gamma_D (p_{b,D} - p_{\beta,D})}{\tau}, \tag{3}$$

while a consumer who is indifferent between buying the parallel import (β) and not buying at all (0) has a gross valuation $\theta_D^{\beta,0}$, given by

$$\theta_D^{\beta,0}(1-\tau) - \gamma_D p_{\beta,D} = 0 \;\Leftrightarrow \theta_D^{\beta,0} = \frac{\gamma_D p_{\beta,D}}{(1-\tau)}. \tag{4}$$

Consequently, in country D, if the parallel import is available, demand for the authorized product b and for the parallel import β is given by

$$q_{b,D}^{*} = 1 - \frac{\gamma_D (p_{b,D} - p_{\beta,D})}{\tau} \text{ and } q_{\beta,D}^{*} = \frac{\gamma_D (p_{b,D} - p_{\beta,D})}{\tau} - \frac{\gamma_D p_{\beta,D}}{(1-\tau)}. \tag{5}$$

In country S, only a locally sourced version of the brand-name drug, sold by the intermediary I_S, is available. A consumer who is indifferent between buying the drug and not buying has a gross valuation $\theta_S^{b,0}$, given by

$$\theta_S^{b,0} - \gamma_S p_{b,S} = 0 \iff \theta_S^{b,0} = \gamma_S p_{b,S}. \tag{6}$$

Accordingly, in country S demand for the authorized product b is given by

$$q_{b,S} = 1 - \gamma_S p_{b,S}. \tag{7}$$

Production technologies exhibit constant marginal costs, which are normalized to zero for simplicity. It is assumed that parallel trade is costless.

The structure of the model can be summarized by the following three-stage game: In the first stage, the manufacturer specifies for each intermediary a wholesale price w_j and fixed fee ϕ_j. In the second and final stage, the foreign intermediary I_S sets the price in country S (that is, $p_{b,S}$) and the price for the parallel import in country D (namely $p_{\beta,D}$), while the domestic intermediary I_D sets the price for the locally sourced version in country D (that is, $p_{b,D}$).

3 Equilibrium without Regulation

As a benchmark consider the case of unregulated markets, when the manufacturer and both intermediaries can set prices freely. If parallel trade is allowed, the manufacturer's pricing decisions – the wholesale price w^*_D charged the intermediary I_D and the wholesale price w^*_S charged the intermediary I_S – are interdependent.

The manufacturer's profit is given as

$$\pi^*_M = \underbrace{w^*_D\left(1 - \frac{\gamma_D(p^*_{b,D} - p^*_{\beta,D})}{\tau}\right)}_{\pi^*_{w_b,D}} + \underbrace{w^*_S\left(1 - \gamma_S p^*_{b,S}\right)}_{\pi^*_{w_b,S}} + \underbrace{w^*_S\left(\frac{\gamma_D(p^*_{b,D} - p^*_{\beta,D})}{\tau} - \frac{\gamma_D p^*_{\beta,D}}{(1-\tau)}\right)}_{\pi^*_{w_\beta}} + \phi^*_D + \phi^*_S, \tag{8}$$

where $\pi^*_{w_b,D}$ denotes the wholesale profit from the intermediary I_D's sales in country D, $\pi^*_{w_b,S}$ the wholesale profit from the intermediary I_S's sales in country S, $\pi^*_{w_\beta}$ the wholesale profit from the intermediary I_S's sales as parallel imports in country D, and ϕ^*_D and ϕ^*_S the fixed fees paid by the intermediaries.

The intermediaries' profits are given as

$$\pi^*_{I_D} = \underbrace{\left(p^*_{b,D} - w^*_D\right)\left(1 - \frac{\gamma_D(p^*_{b,D} - p^*_{\beta,D})}{\tau}\right)}_{\pi^*_{b,D}} - \phi^*_D, \tag{9}$$

$$\text{and } \pi^*_{I_S} = \underbrace{\left(p^*_{b,S} - w^*_S\right)\left(1 - \gamma_S p^*_{b,S}\right)}_{\pi^*_{b,S}} + \underbrace{\left(p^*_{\beta,D} - w^*_S\right)\left(\frac{\gamma_D(p^*_{b,D} - p^*_{\beta,D})}{\tau} - \frac{\gamma_D p^*_{\beta,D}}{(1-\tau)}\right)}_{\pi^*_{\beta,D}} - \phi^*_S, \tag{10}$$

where $\pi^*_{b,D}$ and $\pi^*_{b,S}$ denotes the profit from sales in country D and S, resp. and $\pi^*_{\beta,D}$ the profit from sales as parallel imports in country D.

In country D, the domestic intermediary I_D maximizes (9) with respect to $p^*_{b,D}$ which yields the first order condition

$$\left(1-\frac{\gamma_D(p^*_{b,D}-p^*_{\beta,D})}{\tau}\right)+\left(p^*_{b,D}-w^*_D\right)\underbrace{\left(-\frac{\gamma_D}{\tau}\right)}_{\frac{\partial q^*_{b,D}}{\partial p^*_{b,D}}}=0 \tag{11}$$

and the best response function

$$p^*_{b,D}=\frac{1}{2\gamma_D}\left(\tau+p^*_{\beta,D}\gamma_D+\gamma_D w^*_D\right). \tag{12}$$

The foreign intermediary I_S maximizes (10) with respect to $p^*_{\beta,D}$ which yields the first order condition

$$\left(\frac{\gamma_D(p^*_{b,D}-p^*_{\beta,D})}{\tau}-\frac{\gamma_D p^*_{\beta,D}}{(1-\tau)}\right)+\left(p^*_{\beta,D}-w^*_S\right)\underbrace{\left(-\frac{\gamma_D}{\tau}-\frac{\gamma_D}{(1-\tau)}\right)}_{\frac{\partial q^*_{\beta,D}}{\partial p^*_{\beta,D}}}=0 \tag{13}$$

and the best response function

$$p^*_{\beta,D}=\frac{1}{2}\left(w^*_S+p^*_{b,D}\left(1-\tau\right)\right). \tag{14}$$

Equilibrium prices are $p^*_{b,D}=\frac{2\tau+\gamma_D(w^*_S+2w^*_D)}{\gamma_D(\tau+3)}$ and $p^*_{\beta,D}=\frac{(1-\tau)\tau+\gamma_D(2w^*_S+w^*_D(1-\tau))}{\gamma_D(\tau+3)}$. Note that both drug prices in country D, $p^*_{b,D}$ and $p^*_{\beta,D}$ increase in both intermediaries' marginal costs, i.e. both wholesale prices w^*_D and w^*_S, with the effect of the intermediary's own marginal cost being stronger than the effect of the competitor's marginal cost ($\frac{\partial p^*_{b,D}}{\partial w^*_D}>\frac{\partial p^*_{b,D}}{\partial w^*_S}$, $\frac{\partial p^*_{\beta,D}}{\partial w^*_S}>\frac{\partial p^*_{\beta,D}}{\partial w^*_D}$).

In country S, the intermediary maximizes (10) with respect to $p^*_{b,S}$. The first order condition to this maximization problem is

$$\left(1-\gamma_S p^*_{b,S}\right)+\left(p^*_{b,S}-w^*_S\right)\left(-\gamma_S\right)=0, \tag{15}$$

resulting in the price $p^*_{b,S}=\frac{1+w^*_S\gamma_S}{2\gamma_S}$. Note that the drug price $p^*_{b,S}$ increases in the wholesale price w^*_S.

With fixed fees of

$$\phi^*_D=\underbrace{\frac{\left(2\tau+\gamma_D w^*_S-\gamma_D w^*_D\left(\tau+1\right)\right)^2}{\tau\gamma_D\left(\tau+3\right)^2}}_{\pi^*_{b,D}} \tag{16}$$

$$\text{and } \phi^*_S=\underbrace{\frac{\left(1-w^*_S\gamma_S\right)^2}{4\gamma_S}}_{\pi^*_{b,S}}+\underbrace{\frac{\left(\gamma_D w^*_D\left(1-\tau\right)+\left(1-\tau\right)\tau-\gamma_D w^*_S\left(1+\tau\right)\right)^2}{\tau\gamma_D\left(1-\tau\right)\left(\tau+3\right)^2}}_{\pi^*_{\beta,D}} \tag{17}$$

the manufacturer extracts the intermediaries' total profits.

Substituting (16), (17), and equilibrium prices into (8) and maximizing with respect to w_D^* and w_S^* gives the wholesale prices $w_D^* = \frac{(1-\tau)(2\tau+\gamma_D w_S^*)}{\gamma_D(3\tau+1)}$ and $w_S^* = \frac{2(1-\tau)\left(5\tau-\tau^2+2\gamma_D w_D^*(1-\tau)\right)}{4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2}$. Wholesale prices mutually reinforce one another; a higher wholesale price in the destination country, w_D^* induces a higher wholesale price in the source country, w_S and vice versa.

Equilibrium wholesale prices are given as:

$$w_D^* = \frac{2(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \text{ and } w_S^* = \frac{2(1-\tau)}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)}. \tag{18}$$

For segmented markets, the manufacturer sets the wholesale prices equal to marginal cost, i.e. $w_D = w_S = 0$. This avoids the double marginalization problem resulting from vertical separation in imperfectly competitive markets. However, if parallel trade is allowed and results in market integration, the manufacturer raises both wholesale prices. This allows him to induce higher retail prices and reduce competition from parallel trade in the destination country.

As a result of price competition between the two intermediaries, both drug prices in country D, $p_{b,D}^*$ and $p_{\beta,D}^*$ increase in both wholesale prices w_D^* and w_S^*. The choice of the wholesale price w_D^* therefore includes a strategic effect: An increase of w_D^* raises not only the price for the locally sourced version but also the price for the parallel import. The same effect holds for the wholesale price w_S^*, an increase of w_S^* raises both the price for the parallel import and the locally sourced version. This allows the manufacturer to exploit a strategic effect: By raising both w_D^* and w_S^*, he can enforce a coordinated price increase in the destination country, i.e. induce higher retail prices for both versions of the drug.

Consider Figure **??** for a visualization of this effect. Dashed lines are best response functions for $w_D^* = w_S^* = 0$, yielding retail price equilibrium A. Solid lines are best response functions for $w_D^* > 0$, $w_S^* > 0$, yielding retail price equilibrium B. The increase of wholesale prices (from $w_D^* = w_S^* = 0$ to $w_D^* > 0$, $w_S^* > 0$) shifts the retail price equilibrium from A to B, inducing higher retail prices. Note that also the increase of only one wholesale price would result in higher retail prices.

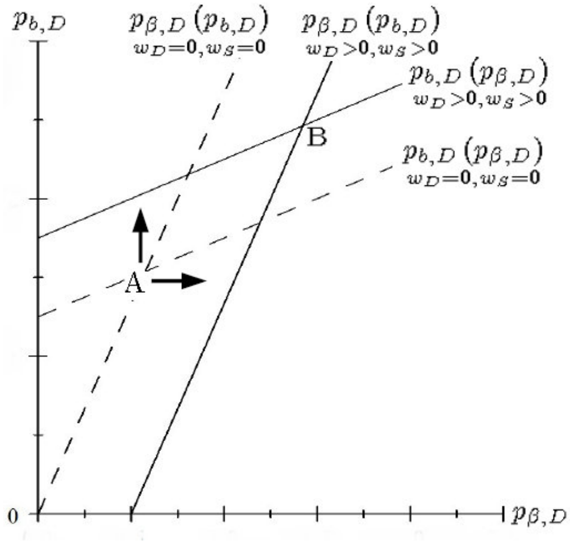

Figure 1: Best response functions for $w_D = w_S = 0$ and $w_D > 0, w_S > 0$.

This strategic incentive for the manufacturer to raise wholesale price is described by Rey & Stiglitz (1995), who show that two competing manufacturers can use exclusive territories also to reduce interbrand competition.

In my model, parallel trade results in competition between two intermediaries with exclusive territories in the destination country, but they are supplied by the *same* manufacturer. The manufacturer cannot suppress this form of intrabrand competition due to lack of vertical control and international, respectively regional exhaustion of intellectual property rights. But parallel trade provides the manufacturer with the option to exploit the strategic effect of exclusive territories, namely inducing higher retail prices and reducing competition by increasing wholesale prices. This effect is stronger, when products are close substitutes and prices increase more in response to wholesale price increases, i.e. the degree of product differentiation is small.

At the same time, an increase of w_S^* also increases the drug price and decreases the quantity sold in the source country. If price elasticity in the source country is high, a given price increase results in a higher reduction of quantity. A wholesale price of zero would be profit-maximizing for the manufacturer with respect to the source country. Thus, the impact of an increase of w_S^* on the profit from the source country restricts the manufacturer in exploiting this strategic effect.

The first order conditions illustrate the effects of the choice of the wholesale price on the manufacturer's profit, see Appendix A for details.

Equilibrium drug prices are

$$p_{b,D}^* = \frac{2\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{\gamma_D\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)},\ p_{\beta,D}^* = \frac{\left(1-\tau\right)\left(2\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{\gamma_D\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} \text{ and} \qquad (19)$$

$$p_{b,S}^{*} = \frac{4\gamma_D + 3\gamma_S\left(1-\tau^2\right)}{2\gamma_S\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)}. \tag{20}$$

As compared to segmented markets, parallel trade reduces the price for the locally sourced version in country D, $p_{b,D}^{*}$, but raises the drug price in country S, $p_{b,S}^{*}$.

Equilibrium quantities are

$$q_{b,D}^{*} = \frac{2\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)},\ q_{\beta,D}^{*} = \frac{\left(1-\tau\right)\gamma_S}{4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)} \text{ and} \tag{21}$$

$$q_{b,S}^{*} = \frac{4\gamma_D - \gamma_S\left(1-3\tau\right)\left(1-\tau\right)}{2\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)}. \tag{22}$$

In country D, the quantity of the locally sourced version of the drug $q_{b,D}^{*}$ is lower under parallel trade, but the total quantity of both versions of the drug, $q_{b,D}^{*}+\ q_{\beta,D}^{*}$ is higher under parallel trade than under segmented markets. In country S, the quantity sold, $q_{b,S}^{*}$ is lower under parallel trade than under segmented markets.

Under parallel trade, the manufacturer may increase the profit allocated to the destination country. Competition from parallel trade has a profit-decreasing effect, but the strategic effect of reducing competition by increasing the wholesale prices and inducing higher retail prices works in the opposite direction. If the effect of a higher wholesale profit from sales as parallel imports and a higher fixed fee extracted from intermediary I_S exceeds the effect of competition in the destination country[2], the profit allocated to the destination country is higher under parallel trade. The strategic effect of reducing competition by increasing wholesale prices is crucial; the manufacturer's profit is always lower under direct sales in the destination country. The profit earned in the source country is always lower due to the double marginalization effect with a higher drug price and a lower quantity sold. The total effect of parallel trade on the manufacturer's profit depends on the relative size of these two effects and with it on the price elasticities in both countries (i.e. coinsurance rates) and the substitutability of both products (i.e. the degree of vertical product differentiation).

4 Equilibrium with Maximum Wholesale Margins

In this model, both the manufacturer and each intermediary have a monopoly position, allowing them to set prices freely. The vertical separation in an imperfectly competitive market results in inefficient successive markups by both the manufacturer and each intermediary. Under segmented markets, the manufacturer solves this problem by setting wholesale prices equal to marginal cost (and extracting the profits of the intermediaries via the fixed fee). Under parallel trade, the manufacturer exploits the strategic effect of exclusive territories in the destination country and

[2] That is a lower wholesale profit from sales as locally sourced and a higher fixed fee extracted from intermediary I_D.

raises wholesale prices to induce higher retail prices and reduce competition. This results in a double marginalization effect in the source country. Consequently, in both countries, drug prices are higher than they would be if the manufacturer sold directly[3]. In addition, in the source country, the drug price is higher than in the absence of parallel trade.

This sections studies maximum wholesale margins as a possible regulatory intervention addressing the problem of double marginalization by restricting the markups charged by intermediaries. This instrument is applied in virtually all European countries. Commonly, intermediaries are granted a certain (percentage) markup on the wholesale price[4]. In a strict specification, this cannot be described in this model[5]. Therefore, I model maximum wholesale margins instead as a restriction of the markup charged under free pricing. In country j, intermediaries may charge a fraction $\left(1-\mu_j\right)$ of the markup under free pricing:

$$\begin{aligned} p_{b,D}^{*\mu} &= w_D^{*\mu} + (1-\mu_D)\, m_{b,D}^{*}, \text{ with } m_{b,D}^{*} = p_{b,D}^{*} - w_D^{*}, \\ p_{\beta,D}^{*\mu} &= w_S^{*\mu} + (1-\mu_D)\, m_{\beta,D}^{*}, \text{ with } m_{\beta,D}^{*} = p_{\beta,D}^{*} - w_S^{*}, \\ \text{and } p_{b,S}^{*\mu} &= w_S^{*\mu} + (1-\mu_S)\, m_{b,S}^{*}, \text{ with } m_{b,S}^{*} = p_{b,S}^{*} - w_S^{*}. \end{aligned} \tag{23}$$

The case of $\mu_j = 0$ corresponds to no restrictions on pricing, i.e. no regulation (the intermediaries may set prices freely), while the case of $\mu_j = 1$ corresponds to the strictest regulation possible (the intermediaries are forced to price at marginal cost). This construction allows me to analyze different degrees of regulation explicitly.

To illustrate the way markup restrictions affect the equilibrium, I start with the extreme case of one country enforcing marginal cost pricing. Then I describe the general effects for any degree of regulation in the symmetric equilibrium of both countries applying maximum wholesale margins.

4.1 Maximum Wholesale Margin Regulation in the Destination Country

Consider first the case of the destination country prohibiting markups by intermediaries, i.e. $\mu_D = 1$, whereas in the source country, pricing is free.

In the destination country, both intermediaries are forced to price at marginal cost, $p_{b,D}^{*\mu} = w_D^{*\mu}$ and $p_{\beta,D}^{*\mu} = w_S^{*\mu}$. This implies that intermediaries make zero profits in the destination

[3] Note that although the price of the locally sourced version of the drug in country D is lower under parallel trade, direct sales would bring about an even lower price, avoiding the impact of double marginalization. The price change associated with the switch from segmented markets to parallel trade is the net of a competition and a double marginalization effect.

[4] In all European countries except for Italy maximum wholesale markups are defined in terms of wholesale prices.

[5] Note that in the symmetric equilibrium regulatory instruments of both countries mutually offset one another. This implies that a binding restriction on markups depends on the restriction in the respective other countries. Therefore, a symmetric equilibrium with both countries effectively restricting markups can only exist, if restriction on markups are scaled in terms of the markup under free pricing.

country, the corresponding fixed fees ($\phi_D^{*\mu}$ and the part of $\phi_S^{*\mu}$ associated with parallel trade) are also zero. In the source country, intermediary I_S may set the drug price freely.

The enforcement of marginal cost pricing results in the manufacturer not being able to exploit the strategic effect of exclusive territories anymore. With free pricing of intermediaries, the manufacturer raises both wholesale prices to induce higher retail prices and to reduce competition from parallel trade. Under maximum wholesale margins, the regulation of pricing cuts the link between $w_D^{*\mu}$ and $p_{\beta,D}^{*\mu}$ and the link between $w_S^{*\mu}$ and $p_{b,D}^{*\mu}$. Accordingly, an increase of $w_D^{*\mu}$ does not raise the price for the parallel import, an increase of $w_S^{*\mu}$ does not raise the price for the locally sourced version.

Consequently, not being able to restrict competition via (inducing higher) retail prices, the manufacturer raises the wholesale prices to reduce competition by parallel trade:

$$w_D^{*\mu} - w_D^* > 0,\; w_S^{*\mu} - w_S^* > 0. \tag{24}$$

An increase of the wholesale price $w_S^{*\mu}$ aggravates the double marginalization effect in the source country by increasing the drug price and reducing the quantity sold. But the effect of an increase of $w_S^{*\mu}$ on increasing the price of the parallel import (and thus reducing competition) is stronger than on increasing the price of the drug in the source country (and thus aggravating the double marginalization effect). Consequently, the manufacturer can reduce competition to larger extent by increasing wholesale prices under maximum wholesale margins than by increasing wholesale prices to induce higher retail prices under free pricing. The increase of the wholesale price under maximum wholesale margins translates directly to an increase in the retail prices by the same amount, while under free pricing, intermediaries increase their prices less than proportionally. Thus, under maximum wholesale margins, the price difference between the two versions of the drug is lower.

However, the double marginalization effect prevents the manufacturer from raising $w_D^{*\mu}$ sufficiently to induce the same retail price for locally sourced version as under free pricing:

$$p_{b,D}^{*\mu} - p_{b,D}^* < 0. \tag{25}$$

Note that an increase of $w_D^{*\mu}$ requires also equivalent raise of $w_S^{*\mu}$ to restrict competition.

Depending on the price elasticity in the source country, the relative importance of the double marginalization effect may be low enough to result in an increase of price of parallel import:

$$p_{\beta,D}^{*\mu} - p_{\beta,D}^* > 0, \text{ if } \gamma_S < \frac{2}{(1-\tau)}\gamma_D. \tag{26}$$

That is, under maximum wholesale margins, the manufacturer is able to reduce competition to a larger extent, but he fails to maintain the same price level for the locally sourced version. He increases the price for the parallel import in relative (or even absolute terms). This results in a shift of demand from the parallel import to the locally sourced version, the quantity of the locally sourced version increases, the quantity of parallel import decreases:

$$q_{b,D}^{*\mu} - q_{b,D}^{*} > 0,\ q_{\beta,D}^{*\mu} - q_{\beta,D}^{*} < 0. \tag{27}$$

In the source country, the increase of the wholesale price $w_S^{*\mu}$ increases the drug price and decreases quantity sold:

$$p_{b,S}^{*\mu} - p_{b,S}^{*} > 0,\ q_{b,S}^{*\mu} - q_{b,S}^{*} < 0. \tag{28}$$

Thus, if the destination country implements marginal cost pricing, the manufacturer cannot exploit the strategic effect of exclusive territories. Instead he increases wholesale prices to reduce competition directly. Under maximum wholesale margins, competition from parallel trade is weaker, but takes place on a lower price level. In the source country, the increase of the wholesale price aggravates the double marginalization effect.

4.2 Maximum Wholesale Margin Regulation in the Source Country

Consider now the case of the source country prohibiting markups by intermediaries, i.e. $\mu_S = 1$, whereas the destination country is not regulated.

In the source country, the intermediary I_S is forced to price at marginal cost, $p_{b,S}^{*\mu} = w_S^{*\mu}$. This implies that the intermediary makes only profits from parallel importing, the profit from sales in the source country is zero. In the destination country, this intermediary is not constrained in setting a price for the parallel import. Intermediary I_D may also set the price for the locally sourced version freely.

In the source country, the enforcement of marginal cost pricing results resolves the double marginalization effect. The drug price includes only the manufacturer's markup.

This allows the manufacturer to increases both wholesale prices to exploit the strategic effect of exclusive territories to a greater degree and to further reduce competition from parallel trade in the destination country:

$$w_D^{*\mu} - w_D^{*} > 0,\ w_S^{*\mu} - w_S^{*} > 0. \tag{29}$$

The profit from selling to the sourced country is maximized for a retail price of $p_{b,S}^{*} = \frac{1}{2\gamma_S}$. Under segmented markets, the manufacturer induces this retail price by setting the wholesale price w_S^{*} to zero, since the intermediary surcharges a monopoly markup. Thus, if the intermediary's monopoly markup disappears and the intermediary sets the price to marginal cost, a wholesale price of $w_S^{*\mu} = \frac{1}{2\gamma_S}$ would maximize profits from selling in the source country. However, when setting wholesale prices, the manufacturer also considers the strategic effect of higher wholesale prices for the destination country. Consequently, it may be profitable to raise $w_S^{*\mu}$ above this profit-maximizing level of $w_S^{*\mu} = \frac{1}{2\gamma_S}$ and reduce competition from parallel trade, increasing profits from selling in the destination country,while reducing profits from selling in the source country by decreasing the quantity sold[6]. Thus, the manufacturer increases wholesale

[6] Note that an increase of the wholesale price above the profit maximizing $w_S^{*\mu} = \frac{1}{2\gamma_S}$ induces a higher reduction

prices balancing the increase in profits from limiting competition destination country and the decrease in profits from reducing the quantity in the source country.

In the source country, the reduction in quantity prevents the manufacturer from raising $w_S^{*\mu}$ sufficiently to induce the same retail price for locally sourced version as under free pricing:

$$p_{b,S}^{*\mu} - p_{b,S}^{*} < 0. \tag{30}$$

Due to the lower drug price, the quantity is higher under maximum wholesale margins:

$$q_{b,S}^{*\mu} - q_{b,S}^{*} > 0. \tag{31}$$

By increasing the wholesale prices the manufacturer induces even higher retail prices in the destination country:

$$p_{b,D}^{*\mu} - p_{b,D}^{*} > 0,\ p_{\beta,D}^{*\mu} - p_{\beta,D}^{*} > 0. \tag{32}$$

The effect of higher prices dominates the effect of reducing competition from parallel trade and the quantity of the locally sourced version is lower:

$$q_{b,D}^{*\mu} - q_{b,D}^{*} < 0. \tag{33}$$

Depending on the price elasticity, i.e. the coinsurance rate the manufacturer may limit competition from parallel trade or even block parallel entirely:

$$q_{\beta,D}^{*\mu} - q_{\beta,D}^{*} < 0,\ q_{\beta,D}^{*\mu} > 0 \text{ if } \gamma_S > \frac{1}{(1-\tau)}\gamma_D. \tag{34}$$

Thus, if the source country implements marginal cost pricing and thereby avoids the double marginalization effect, the manufacturer is able to exploit the strategic effect of wholesale price increases in the destination country to a greater extent. This increases drug prices in the destination country and reduces competition from parallel trade. In the source country, the double marginalization effect is mitigated, the drug price is lower.

4.3 Symmetric Equilibrium under Maximum Wholesale Margins

Consider now the symmetric equilibrium with both countries restricting markups of intermediaries, which is the case for basically all member states in the European Union.

If the destination country restricts markups, the manufacturer cannot exploit the strategic effect of exclusive territories and increases wholesale prices to reduce competition directly. This effect is limited by the double marginalization effect in the source country. If the source country also restricts markups, the double marginalization effect is mitigated. This increases the incentive to raise wholesale prices to reduce competition. The total effect on drug prices in the destination

of quantity and thus a higher reduction of profits than an increase of the wholesale price above zero under free pricing.

country is ambiguous, as the further increase of wholesale prices due to the restriction of markups is not offset by stricter markup limits in the destination country. Also in the source country, the total effect on the drug price is ambiguous. The increase of the wholesale price emerging from the restriction of markups in the destination country is not compensated by a tougher restriction of markups in the source country.

Similarly, if the source country restricts markups and mitigates the double marginalization effect, the manufacturer raises wholesale prices to exploit the strategic effect of wholesale price increases in the destination country to a greater extent. If the destination country restricts markups as well, the impact of wholesale price increases on reducing competition is even higher. Intermediaries do not absorb wholesale price increases partly, but pass them on to retail prices completely. Accordingly, the manufacturer increases wholesale prices more, with the total effect on drug prices in both countries being ambiguous.

Thus, in a symmetric equilibrium, regulatory instruments of both countries mutually offset one another, with the total effect being ambiguous.

Proposition 1 summarizes the effect of maximum wholesale margins on drug prices and quantities:

Proposition 1 *A restriction of markups in the destination country i) reduces competition from parallel trade by reducing the relative price difference between both versions in the destination country, but reduces the price for the locally sourced version and ii) aggravates the double marginalization effect by increasing the drug price in the source country. A restriction of markups in the source country i) mitigates the double marginalization effect by decreasing the drug price in the source country and ii) increases the drug prices and reduces competition from parallel trade in the destination country. Regulatory instruments of both countries partly offset each other.*

5 Policy Alternative: Mandatory Rebates

As an alternative to the restriction of pricing for intermediaries, pricing of the manufacturer can also be restricted to mitigate the double marginalization effect and reduce drug prices.

Under mandatory rebates, the manufacturer is forced to grant a discount on the wholesale price. This reduces or even avoids the monopolistic markups on the first market stage.

When being forced to grant discounts and subject to free pricing at the same time, manufacturers may simply avoid discounts by increasing prices. Therefore, mandatory rebates are applied in combination with price freezes, which prevent strategic price increases in response to discounts. Here, price freezes apply to wholesale prices only, as the regulatory intervention is intended to be limited with respect to one level only[7].

[7] In Germany, the increase of mandatory rebates from 6 to 16% in the SHI-Amending Law (GKV-ÄndG) of 2010 was combined with a price freeze at the retail level. As a price freeze at the retail level would leave drug prices and quantities sold unchanged and only affect marginal cost, this analysis restricts the price freeze to the wholesale level only.

Under mandatory rebates, wholesale prices are discounted by the factor ψ_j in country j. In country D, the wholesale price amounts to:

$$w_D^{*\psi} = (1 - \psi_D)\, w_D^* \tag{35}$$

and in country S to:

$$w_S^{*\psi} = (1 - \psi_S)\, w_S^*. \tag{36}$$

The case of $\psi_j = 0$ corresponds to no restrictions on pricing, i.e. no regulation (the manufacturer may set wholesale prices freely), while the case of $\mu_j = 1$ corresponds to the strictest regulation possible (the manufacturer is forced to price at marginal cost). Similarly as for maximum wholesale margins, I first consider the extreme case of one country enforcing marginal cost pricing.

To illustrate the way mandatory rebates affect the equilibrium, I start with the extreme case of one country enforcing marginal cost pricing. Then I describe the general effects for any degree of regulation in the symmetric equilibrium of both countries applying maximum wholesale margins.

5.1 Mandatory Rebates in the Destination Country

Consider first the case of the destination country implementing marginal cost pricing and forcing the manufacturer to set the wholesale price to zero, i.e. $\psi_D = 1$ and $w_D^{*\psi} = 0$. In the source country, pricing is free. This implies that the wholesale profit from intermediary I_D's sales in the destination country is zero.

In the destination country, the manufacturer can no longer exploit the strategic effect of inducing higher retail prices by increasing both wholesale prices. However, this is no longer necessary, the manufacturer can control drug prices with the wholesale $w_S^{*\psi}$ effectively: As the effect of a change in w_S^{ψ} is stronger for the price for the parallel import than for the price of the locally sourced version, setting $w_S^{*\psi}$ slightly higher than $w_D^{*\psi}$ is sufficient to induce a shift in demand from the parallel import to the locally sourced version. With w_D^{ψ} being set to zero, the manufacturer can decrease $w_S^{*\psi}$ and reduce competition from parallel trade more effectively:

$$w_S^{*\psi} - w_S^* > 0. \tag{37}$$

Both wholesale prices are lower than under free pricing and accordingly, also drug prices are lower in the destination country:

$$p_{b,D}^{*\psi} - p_{b,D}^* < 0,\ p_{\beta,D}^{*\psi} - p_{\beta,D}^* < 0. \tag{38}$$

The price difference decreases and induces a shift of demand from the parallel import to the locally sourced version. The quantity of locally sourced version is higher:

$$q_{b,D}^{*\psi} - q_{b,D}^{*} > 0. \tag{39}$$

Depending on the price elasticity, i.e. the coinsurance rate the manufacturer may limit competition from parallel trade or even block parallel entirely:

$$q_{\beta,D}^{*\psi} - q_{\beta,D}^{*} < 0,\ q_{\beta,D}^{*\psi} > 0 \text{ if } \gamma_S > \frac{2}{(3+\tau)}\gamma_D. \tag{40}$$

In the source country, the decrease of the wholesale price $w_S^{*\psi}$ mitigates the double marginalization problem. The drug price is lower, the quantity sold is higher:

$$p_{b,S}^{*\psi} - p_{b,S}^{*} < 0,\ q_{b,S}^{*\psi} - q_{b,S}^{*} > 0. \tag{41}$$

Thus, if the destination country implements marginal cost pricing, the manufacturer can reduce competition from parallel trade more effectively. This decreases drug prices in the destination country, but also reduces competition from parallel trade. In the source country, the double marginalization effect is mitigated.

5.2 Mandatory Rebates in the Source Country

Consider now the case of the source country implementing marginal cost pricing and forcing the manufacturer to set the wholesale price to zero, i.e. $\psi_S = 1$ and $w_S^{*\psi} = 0$. In the destination country, pricing is free.

In the source country, a wholesale price of zero avoids the double marginalization effect.

The drug price is lower than under free pricing and corresponds to the drug price under segmented markets. Similarly, the quantity is higher than under free pricing and corresponds to the quantity under segmented markets:

$$p_{b,S}^{*\psi} - p_{b,S}^{*} < 0,\ q_{b,S}^{*\psi} - q_{b,S}^{*} > 0. \tag{42}$$

In the destination country, a wholesale price $w_S^{*\psi}$ of zero induces the maximum competition from parallel trade. To cope with competition and to also sell the locally sourced version, the manufacturer decreases the wholesale price $w_D^{*\psi}$ as well:

$$w_D^{*\psi} - w_D^{*} < 0. \tag{43}$$

This reduces both drug prices:

$$p_{b,D}^{*\psi} - p_{b,D}^{*} < 0,\ p_{\beta,D}^{*\psi} - p_{\beta,D}^{*} < 0. \tag{44}$$

A change in the wholesale price $w_D^{*\psi}$ still exhibits the strategic effect, both drug prices increase in $w_D^{*\psi}$. Consequently, the manufacturer does not set $w_D^{*\psi}$ to zero. This implies that the relative prices decreases. There is a shift in demand from the locally sourced version to the parallel import. The quantity of the locally sourced version is lower than under free pricing, the quantity of the parallel import is higher:

$$q_{b,D}^{*\psi} - q_{b,D}^{*} < 0, \; q_{\beta,D}^{*\psi} - q_{\beta,D}^{*} > 0. \tag{45}$$

If the source country implements marginal cost pricing, the double marginalization effect is mitigated, the drug price is lower. In the destination country, competition from parallel trade is intensified, resulting in lower drug prices.

5.3 Symmetric Equilibrium under Mandatory Rebates

Consider now the symmetric equilibrium with both countries adopting mandatory rebates and restricting the wholesale prices.

If the destination country restricts the wholesale price, reducing competition from parallel trade requires a lower wholesale price for the intermediary in the source country. Reducing the wholesale price in the source country mitigates the double marginalization effect and reduces the drug price in the source country. If the source country restricts the wholesale price, competition from parallel trade is intensified and the manufacturer reduces the wholesale price in the destination country. This reduces drug prices in the destination country.

Thus, in a symmetric equilibrium, regulatory instruments of both countries mutually reinforce one another with respect to drug prices in both countries. The total effect on competition from parallel trade is ambiguous, mandatory rebates in the destination country tend to limit competition from parallel trade, whereas mandatory rebates in the source country work towards intensified competition from parallel trade.

Proposition 2 summarizes the effect of mandatory rebates:

Proposition 2 *A mandatory rebate on the wholesale price in the destination country i) reduces competition from parallel trade by reducing the relative price difference between both versions in the destination country, but reduces drug prices and ii) mitigates the double marginalization effect by decreasing the drug price in the source country. A mandatory rebate on the wholesale price in the source country i) mitigates the double marginalization effect by decreasing the drug price in the source country and ii) decreases the drug prices and intensifies competition from parallel trade in the destination country. Regulatory instruments of both countries mutually reinforce each other with respect to drug prices.*

6 Conclusion

In this paper, I have studied the effect of pharmaceutical regulation at the wholesale level, if markets are integrated by parallel trade. In particular, I have analyzed maximum wholesale margins and mandatory rebates as a policy alternative with respect to effects on drug prices, quantities, and public pharmaceutical expenditure.

In the model used in this paper, parallel trade provides the manufacturer with the possibility to exploit the strategic effect of exclusive territories in the destination country. He raises wholesale prices in both countries to induce higher retail prices in the destination country and to reduce competition from parallel trade. This results in a double marginalization effect in both the destination country and the source country of parallel imports, increasing drug prices and reducing quantities sold. In the absence of the possibility to stimulate downstream competition or to enforce vertical integration, regulatory authorities may implement regulatory instruments such as maximum wholesale margins or mandatory rebates to limit pricing at one of the two market stages. At the same time, the respective other market stage is also affected by these instruments. Under maximum wholesale margins, pricing by intermediaries is restricted, but the manufacturer incorporates this effect in his price setting and adjusts wholesale prices in response. Under segmented markets, the manufacturer may neutralize the effect of maximum wholesale margins by increasing wholesale prices, but parallel trade prevents the manufacturer from offsetting this effect completely. Under mandatory rebates, wholesale prices set by the manufacturer are restricted; intermediaries, however, do not pass through discounts completely, but keep a part of it.

Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. Restrictions of pricing by intermediaries in the destination country shift the competition-reducing effect from retail prices to wholesale prices, which improves the effect of price increases. Under maximum wholesale margins, intermediaries do not absorb wholesale price increases partly, but pass them on to retail prices completely. Restrictions of pricing by intermediaries in the source country mitigate the double marginalization effect, allowing the manufacturer to focus on exploiting the strategic effect in the destination country. Since the manufacturer cannot offset the effect of markups completely, restrictions of pricing also reduce drug prices in the respective country. But via the increase of wholesale prices, they increase drug prices in the respective other country. In the symmetric equilibrium with both countries applying maximum wholesale margins, regulatory instruments exhibit an offsetting effect. Therefore, this regulatory instrument may be inappropriate in a setting where markets are integrated by parallel trade.

Mandatory rebates may be an alternative, they restrict wholesale prices set by the manufacturer. A restriction of the wholesale price in the destination country allows the manufacturer to reduce competition from parallel trade more easily, since a shift in demand from the parallel import to the locally sourced version requires a lower wholesale price for the foreign manufacturer than under free pricing. At the same time, drug prices in both countries are reduced.

A restriction of the wholesale price in the source country intensifies competition from parallel trade. It also reduces drug prices in both countries. In the symmetric equilibrium with both countries applying mandatory rebates, regulatory instruments exhibit a reinforcing effect. This is, the regulation of wholesale prices has a positive externality.

Generally, the results of the model are conditional on the contract choice of the manufacturer. The model assumes the manufacturer to have full contract freedom and thus, being able to write a two-part tariff in order to avoid the double marginalization problem. Limited contract choice may also have an impact on the consequences of parallel trade, which needs to be investigated further.

References

[1] Brekke, K. R.; Holmas, T. & Straume, O. R. (2010): Reference pricing, competition, and pharmaceutical expenditures: theory and evidence from a natural experiment, CESifo Working Paper No. 3258.

[2] Chen, Y. & Maskus, K. E. (2005): Vertical pricing and parallel imports. The journal of international trade & economic development 14, 1-18.

[3] Enemark, U.; Møller Pedersen, K.; Sørensen, J. (2006): The economic impact of parallel imports of pharmaceuticals. University of Odense.

[4] EU Commission (2003): Communication COM(2003) 839 Commission Communication on parallel imports.

[5] Ganslandt, M. & Maskus, K. E. (2007): Vertical distribution, parallel trade, and price divergence in integrated markets. European Economic Review 51, 943-970.

[6] Glynn, D. (2009): The effects of parallel trade on affordable access to medicines. Eurohealth 15, 1-5.

[7] Grossman, G. M. & Lai, E. L.-C. (2008): Parallel imports and price controls. The Rand journal of economics 39, 378-402.

[8] Hurley, J. (2001): An Overview of the Normative Economics of the Health Sector, in: Culyer, A. & Newhouse, J. (eds.): Handbook of Health Economics, Amsterdam, 55-118.

[9] Kanavos, P.; Costa-Font, J.; Merkur, S; Gemmill, M. (2004): The Economic Impact of Pharmaceutical Parallel Trade in European Union Member States: A Stakeholder Analysis. London: LSE Health and Social Care Special Research Paper.

[10] Koenigbauer, I. (2004): Die Auswirkung von Parallelimporten auf die optimale Regulierung von Arzneimittelpreisen. Vierteljahreshefte zur Wirtschaftsforschung 73, 592 - 604.

[11] Maskus, K. E. (2000): Parallel imports. The world economy, 23, 1269-1284.

[12] Maskus, K. E. & Chen, Y. (2002): Parallel imports in a model of vertical distribution: theory, evidence, and policy. Pacific economic review, 7, 319-334.

[13] Maynard, A.; Bloor, K. (2003): Dilemmas In Regulation Of The Market For Pharmaceuticals, Health Affairs 22, 31-41.

[14] Mossialos, E. & Le Grand, J. (1999): Cost containment in the EU: an overview. in: Mossialos, E. & Le Grand, J. (eds.): Health care and cost containment in the European Union. Ashgate.

[15] OECD Health Data (2010): Selected Data, http://stats.oecd.org/Index.aspx?DataSetCode=HEALTH.

[16] Rey, P. & Verge, T. (2008): Economics of Vertical Restraints. in: Buccirossi, P. (ed.): Handbook of Antitrust Economics, 353-390.

[17] Schmalensee, R. (1982): Product differentiation advantages of pioneering brands. American Economic Review, 72, 349-365.

[18] Taylor, D.; Mrazek, M. & Mossialos, E. (2004): Regulating pharmaceutical distribution and retail pharmacy in Europe. in: Mossialos, E.; Mrazek, M. & Walley, T.(eds.): Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality, Maidenhead, 55-79.

[19] Waber, R. L.; Shiv B.; Carmon Z. & Ariely D. (2008): Commercial Features of Placebo and Therapeutic Efficacy. Journal of the American Medical Association 299, 1016-1017.

[20] Zweifel, P.; Breyer, F. & Kifmann, M. (2009): Health Economics. Springer.

Appendix A: The Effect of Parallel Trade

Demand

If parallel trade is not allowed (regime of national exhaustion of intellectual property rights), only the locally sourced version is available in country D. The marginal consumer who is indifferent between buying the locally sourced version from the domestic intermediary (b) or not purchasing at all (0), has a gross valuation $\theta_D^{b,0}$, given by

$$\theta_D^{b,0} - \gamma_D p_{b,D}^* = 0 \Leftrightarrow \theta_D^{b,0} = \gamma_D p_{b,D}^*. \tag{46}$$

Hence, in country D, if the parallel import is not available, demand for the locally sourced version b is given by

$$q_{b,D}^* = 1 - \gamma_D p_{b,D}^*. \tag{47}$$

An asterisk is used to denote variables associated with segmented markets.

If parallel trade is legal (international exhaustion of intellectual property rights), consumers in country D have the choice between the locally sourced version (b) from the domestic intermediary I_D or the parallel import (β) from the foreign intermediary I_S. The marginal consumer who is indifferent between buying the locally sourced version b and the parallel import β has a gross valuation $\theta_D^{b,\beta}$, given by

$$\theta_D^{b,\beta} - \gamma_D p_{b,D} = \theta_D^{b,\beta}(1-\tau) - \gamma_D p_{\beta,D} \;\Leftrightarrow \theta_D^{b,\beta} = \frac{\gamma_D\left(p_{b,D} - p_{\beta,D}\right)}{\tau}, \tag{48}$$

while a consumer who is indifferent between buying the parallel import (β) and not buying at all (0) has a gross valuation $\theta_D^{\beta,0}$, given by

$$\theta_D^{\beta,0}(1-\tau) - \gamma_D p_{\beta,D} = 0 \;\Leftrightarrow \theta_D^{\beta,0} = \frac{\gamma_D p_{\beta,D}}{(1-\tau)}. \tag{49}$$

Consequently, in country D, if the parallel import is available, demand for the authorized product b and for the parallel import β is given by

$$q_{b,D} = 1 - \frac{\gamma_D\left(p_{b,D} - p_{\beta,D}\right)}{\tau} \text{ and } q_{\beta,D} = \frac{\gamma_D\left(p_{b,D} - p_{\beta,D}\right)}{\tau} - \frac{\gamma_D p_{\beta,D}}{(1-\tau)}. \tag{50}$$

Demand in country S is not affected by the availability of parallel imports. Here, only a locally sourced version of the brand-name drug, sold by the intermediary I_S, is available. A consumer who is indifferent between buying the drug and not buying has a gross valuation $\theta_S^{b,0}$, given by

$$\theta_S^{b,0} - \gamma_S p_{b,S} = 0 \iff \theta_S^{b,0} = \gamma_S p_{b,S}. \tag{51}$$

Accordingly, in country S demand for the authorized product b is given by

$$q_{b,S} = 1 - \gamma_S p_{b,S}. \tag{52}$$

Equilibrium without Parallel Trade

When parallel trade is not allowed and markets are segmented, pricing decisions by the manufacturer with respect to both countries – wholesale prices w_D and w_S, which determine drug prices in both countries – are independent.

The manufacturers profit is given as

$$\pi_M^* = \underbrace{w_D^* \left(1 - \gamma_D p_{b,D}^*\right)}_{\pi_{w_b,D}^*} + \underbrace{w_S^* \left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{w_b,S}^*} + \phi_D^* + \phi_S^*, \tag{53}$$

where $\pi_{w_b,D}^*$ and $\pi_{w_b,S}^*$ denote the wholesale profit from the intermediaries' sales in country D and S resp. and ϕ_D^* and ϕ_S^* the fixed fees, which are used to extract the intermediaries' profits.

For the intermediary I_D, profit is given as:

$$\pi_{I_D}^* = \underbrace{\left(p_{b,D}^* - w_D^*\right)\left(1 - \gamma_D p_{b,D}^*\right)}_{\pi_{b,D}^*} - \phi_D^* \tag{54}$$

and for the intermediary I_S as:

$$\pi_{I_S}^* = \underbrace{\left(p_{b,S}^* - w_S^*\right)\left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{b,S}^*} - \phi_S^*, \tag{55}$$

where $\pi_{b,D}^*$ and $\pi_{b,S}^*$ denote the profits from sales in country D and S, respectively.

In country D, the intermediary I_D maximizes (54) with respect to $p_{b,S.}^*$ The first order condition to this problem is

$$\underbrace{\left(1 - \gamma_D p_{b,D}^*\right)}_{I} + \underbrace{\left(p_{b,D}^* - w_D^*\right)\underbrace{\left(-\gamma_S\right)}_{\frac{\partial q_{b,D}^*}{\partial p_{b,D}^*}}}_{II} = 0, \tag{56}$$

resulting in the monopoly drug price $p_{b,D}^* = \frac{(1+w_D^*\gamma_D)}{2\gamma_D}$. The drug price $p_{b,D}^*$ increases in the wholesale price w_D^*.

In country S, the intermediary I_S maximizes (55) with respect to $p_{b,S.}^*$ The first order condition to this problem is

$$\left(1-\gamma_S p_{b,S}^*\right)+\left(p_{b,S}^*-w_S^*\right)\underbrace{\left(-\gamma_S\right)}_{\frac{\partial q_{b,S}^*}{\partial p_{b,S}^*}}=0, \tag{57}$$

resulting in the monopoly drug price $p_{b,S}^* = \frac{(1+w_S^*\gamma_S)}{2\gamma_S}$. The drug price $p_{b,S}^*$ increases in the wholesale price w_S^*.

Turning to the second stage of the game, the manufacturer M sets the fixed fees to

$$\phi_D^* = \pi_{b,D}^* = \frac{\left(1-w_D^*\gamma_D\right)^2}{4\gamma_D} \text{ and } \phi_S^* = \pi_{b,S}^* = \frac{\left(1-w_S^*\gamma_S\right)^2}{4\gamma_S}. \tag{58}$$

in order to extract the intermediaries' profits. In the absence of parallel trade and for segmented markets, the manufacturer's optimal strategy is to set the wholesale price equal to the marginal cost of production, i.e. $w_D^* = w_S^* = 0$[8]. This pricing decision avoids the double marginalization problem and results in the same drug price and sales volume as if the manufacturer sold directly to the consumers.

Equilibrium drug prices are

$$p_{b,D}^* = \frac{1}{2\gamma_D} \text{ and } p_{b,S}^* = \frac{1}{2\gamma_S}. \tag{59}$$

Equilibrium quantities are

$$q_{b,D}^* = \frac{1}{2},\ q_{b,S}^* = \frac{1}{2}. \tag{60}$$

The manufacturer's profit is

$$\pi_M^* = \frac{\left(1-\gamma_D\right)^2}{4\gamma_D} + \frac{\left(1-\gamma_S\right)^2}{4\gamma_S}.$$

Equilibrium with Parallel Trade

If parallel trade is allowed, the manufacturer's pricing decisions –the wholesale price w_D charged the intermediary I_D and the wholesale price w_S charged the intermediary I_S – are no longer independent.

The manufacturer's profit is given as

$$\pi_M = \underbrace{w_D\left(1-\frac{\gamma_D(p_{b,D}-p_{\beta,D})}{\tau}\right)}_{\pi_{w_b,D}} + \underbrace{w_S\left(1-\gamma_S p_{b,S}\right)}_{\pi_{w_b,S}} + \underbrace{w_S\left(\frac{\gamma_D(p_{b,D}-p_{\beta,D})}{\tau}-\frac{\gamma_D p_{\beta,D}}{(1-\tau)}\right)}_{\pi_{w_\beta}} + \phi_D + \phi_S, \tag{61}$$

where $\pi_{w_b,D}$ denotes the wholesale profit from the intermediary I_D's sales in country D, $\pi_{w_b,S}$

[8] This result can also be obtained by substituting (58) and equilibrium prices into (53) and maximizing with respect to w.

the wholesale profit from the intermediary I_S's sales in country S, π_{w_β} the wholesale profit from the intermediary I_S's sales as parallel imports in country D, and ϕ_D and ϕ_S the fixed fees.

The manufacturer's profit differs from the profit under segmented markets in two aspects: First, as the domestic intermediary I_D faces competition by the foreign intermediary I_S in country D, the wholesale profit from I_D's sales in country D and the fixed fee ϕ_D extracted from I_D are lower. Second, the intermediary I_S's sales as reimports result in additional wholesale profit for the manufacturer and for a given wholesale price, the fixed fee extracted from the intermediary I_S, ϕ_S is higher

The intermediaries' profits are given as

$$\pi_{I_D} = \underbrace{(p_{b,D} - w_D)\left(1 - \frac{\gamma_D(p_{b,D} - p_{\beta,D})}{\tau}\right)}_{\pi_{b,D}} - \phi_D, \tag{62}$$

and $$\pi_{I_S} = \underbrace{(p_{b,S} - w_S)(1 - \gamma_S p_{b,S})}_{\pi_{b,S}} + \underbrace{(p_{\beta,D} - w_S)\left(\frac{\gamma_D(p_{b,D} - p_{\beta,D})}{\tau} - \frac{\gamma_D p_{\beta,D}}{(1-\tau)}\right)}_{\pi_{\beta,D}} - \phi_S, \tag{63}$$

where $\pi_{b,D}$ and $\pi_{b,S}$ denotes the profit from sales in country D and S, resp. and $\pi_{\beta,D}$ the profit from sales as parallel imports in country D.

In country D, the domestic intermediary I_D maximizes (62) with respect to $p_{b,D}$ which yields the first order condition

$$\underbrace{\left(1 - \frac{\gamma_D(p_{b,D} - p_{\beta,D})}{\tau}\right)}_{I} + \underbrace{(p_{b,D} - w_D)\left(-\frac{\gamma_D}{\tau}\right)}_{II} = 0$$

and the best response function

$$p_{b,D} = \frac{1}{2\gamma_D}\left(\tau + p_{\beta,D}\gamma_D + \gamma_D w_D\right)$$

Compared to the first order condition for segmented markets, part I and consequently $p_{b,D}$ are lower under parallel trade, if $p_{\beta,D} < p_{b,D}(1-\tau)$, i.e. if the parallel import is priced below the discounted price of the locally sourced drug, which is specified by the vertical product differentiation. Part II of the first order condition differs by the factor $\frac{1}{\tau}$ from the first order condition without parallel trade. For $0 < \tau < 1$, part II and consequently $p_{b,D}$ are lower under parallel trade.

The foreign intermediary I_S maximizes (63) with respect to $p_{\beta,D}$ which yields the first order condition

$$\left(\frac{\gamma_D(p_{b,D} - p_{\beta,D})}{\tau} - \frac{\gamma_D p_{\beta,D}}{(1-\tau)}\right) + (p_{\beta,D} - w_S)\left(-\frac{\gamma_D}{\tau} - \frac{\gamma_D}{(1-\tau)}\right) = 0$$

and the best response function

$$p_{\beta,D} = \frac{1}{2}\left(w_S + p_{b,D}\left(1-\tau\right)\right). \tag{64}$$

Solving for equilibrium prices results in $p_{b,D} = \frac{2\tau+\gamma_D(w_S+2w_D)}{\gamma_D(\tau+3)}$ and $p_{\beta,D} = \frac{(1-\tau)\tau+\gamma_D(2w_S+w_D(1-\tau))}{\gamma_D(\tau+3)}$.

In country S, the intermediary maximizes (63) with respect to $p_{b,S}$. The first order condition to this maximization problem is

$$\left(1-\gamma_S p_{b,S}\right) + \left(p_{b,S} - w_S\right)\left(-\gamma_S\right) = 0, \tag{65}$$

resulting in the price $p_{b,S} = \frac{1+w_S\gamma_S}{2\gamma_S}$. The first order condition is identical to the first order condition under segmented markets. Note that as $p_{b,S}$ increases in the wholesale price w_S, $p_{b,S}$ will be higher under parallel trade, if $w_S > 0$.

With fixed fees of

$$\phi_D = \underbrace{\frac{\left(2\tau + \gamma_D w_S - \gamma_D w_D\left(\tau+1\right)\right)^2}{\tau\gamma_D\left(\tau+3\right)^2}}_{\pi_{b,D}} \tag{66}$$

$$\text{and } \phi_S = \underbrace{\frac{\left(1-w_S\gamma_S\right)^2}{4\gamma_S}}_{\pi_{b,S}} + \underbrace{\frac{\left(\gamma_D w_D\left(1-\tau\right) + \left(1-\tau\right)\tau - \gamma_D w_S\left(1+\tau\right)\right)^2}{\tau\gamma_D\left(1-\tau\right)\left(\tau+3\right)^2}}_{\pi_{\beta,D}} \tag{67}$$

the manufacturer extracts the intermediaries' total profits.

Substituting (**??**), (**??**), and equilibrium prices into (**??**) and maximizing with respect to w_D and w_S gives the wholesale prices $w_D = \frac{(1-\tau)(2\tau+\gamma_D w_S)}{\gamma_D(3\tau+1)}$ and $w_S = \frac{2(1-\tau)\left(5\tau-\tau^2+2\gamma_D w_D(1-\tau)\right)}{4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2}$. Wholesale prices mutually reinforce one another; a higher wholesale price in the destination country, w_D induces a higher wholesale price in the source country, w_S and vice versa.

Equilibrium wholesale prices are given as:

$$w_D = \frac{2\left(1-\tau\right)\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{\gamma_D\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} \text{ and } w_S = \frac{2\left(1-\tau\right)}{4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)}. \tag{68}$$

For segmented markets, the manufacturer sets the wholesale prices equal to marginal cost, i.e. $w_D = w_S = 0$. This avoids the double marginalization problem resulting from vertical separation in imperfectly competitive markets. However, if parallel trade is allowed and results in market integration, the manufacturer raises both wholesale prices. This allows him to induce higher retail prices and reduce competition from parallel trade in the destination country.

As a result of price competition between the two intermediaries, both drug prices in country D, $p_{b,D}$ and $p_{\beta,D}$ increase in both wholesale prices w_D and w_S. The choice of the wholesale price w_D therefore includes a strategic effect: An increase of w_D raises not only the price for the locally sourced version but also the price for the parallel import. The same effect holds for the wholesale price w_S, an increase of w_S raises both the price for the parallel import and the locally sourced version. This allows the manufacturer to exploit a strategic effect: By raising both w_D and w_S, he can enforce a coordinated price increase in the destination country, i.e. induce higher retail prices for both versions of the drug.

In this model, parallel trade results in competition between two intermediaries with exclusive territories in the destination country, but they are supplied by the same manufacturer. The manufacturer cannot suppress this form of interbrand competition due to lack of vertical control and international, resp. regional exhaustion of intellectual property rights. But at the same time, parallel trade provides the manufacturer with a situation, where he can exploit the strategic effect of exclusive territories, namely inducing higher retail prices and reducing competition by increasing wholesale prices. This effect is stronger, when products are close substitutes and prices increase more in response to wholesale price increases, i.e. the degree of product differentiation is small.

At the same time, an increase of w_S also increases the drug price and decreases the quantity sold in the source country. If price elasticity in the source country is high, a given price increase results in a higher reduction of quantity. A wholesale price of zero would maximize profits with respect to the source country. Thus, the impact of an increase of w_S on the profit from the source country restricts the manufacturer exploiting this strategic effect.

The first order conditions illustrate the effects of the choice of the wholesale price on the manufacturer's profit. Maximizing the manufacturer's profit with respect to w_D gives the following first order condition:

$$\begin{aligned}\frac{\partial \pi_M}{\partial w_D} = & \underbrace{\left(-\frac{2\gamma_D w_D (\tau+1) - 2\tau - \gamma_D w_S}{\tau(\tau+3)}\right)}_{\frac{\partial \pi_{w_b,D}}{\partial w_D}} + \underbrace{w_S\left(\frac{\gamma_D}{\tau(\tau+3)}\right)}_{\frac{\partial \pi_{w_\beta}}{\partial w_D}} \\ & + \underbrace{\left(-\frac{2(\tau+1)(2\tau+\gamma_D w_S - \gamma_D w_D(\tau+1))}{\tau(\tau+3)^2}\right)}_{\frac{\partial \phi_D}{\partial w_D}} + \underbrace{\left(\frac{2(\tau(1-\tau) - \gamma_D w_S(\tau+1) + \gamma_D w_D(1-\tau))}{\tau(\tau+3)^2}\right)}_{\frac{\partial \phi_S}{\partial w_D}} \qquad (69)\end{aligned}$$

An increase of the wholesale price w_D shifts demand from the locally sourced version to the parallel import by increasing the price for the locally sourced version by more than the price of the parallel import. This affects the wholesale profit from sales of the locally sourced version (first term) through a price effect and a quantity effect. By decreasing demand for the locally sourced version, an increase of w_D decreases the fixed fee extracted from intermediary I_D (third term). By increasing demand for the parallel import, an increase of w_D increases the wholesale profit from sales of the parallel import (second term) and the corresponding part of the fixed fee extracted from intermediary I_S (fourth term).

By reference to impact of the choice of w_D on the wholesale profit from sales of the locally sourced version ($\frac{\partial \pi_{w_b,D}}{\partial w_D}$), three effects of the choice of w_D on the manufacturer's profit can be illustrated. Consider the following decomposition:

$$\frac{\partial \pi_{w_b,D}}{\partial w_D} = \underbrace{\left(\frac{2\tau + \gamma_D w_S - \gamma_D w_D (\tau+1)}{\tau(\tau+3)}\right)}_{\frac{\partial w_D}{\partial w_D} q_{b,H}} + \underbrace{w_D \left(-\frac{2\gamma_D}{(\tau+3)\tau}\right)}_{w_D \frac{\partial q_{b,H}}{\partial p_{b,H}} \frac{\partial p_{b,H}}{\partial w_D}} + \underbrace{w_D \left(\frac{\gamma_D (1-\tau)}{(\tau+3)\tau}\right)}_{w_D \frac{\partial q_{b,H}}{\partial p_{\beta,H}} \frac{\partial p_{\beta,H}}{\partial w_D}}. \tag{70}$$

The first part illustrates the wholesale profit-increasing effect of a higher wholesale price per unit sold, while ignoring changes in quantity. The second part gives the standard direct effect of an increase of w_D. Via the increase of the price for the locally sourced version, an increase of w_D decreases demand for the locally sourced version and the wholesale profit. The third part indicates the strategic effect: An increase in w_D raises the price for the parallel import as well, thus it increases indirectly the demand for the locally sourced version. Similarly, the impact of the choice of w_D on the other components of the manufacturer's profit can be decomposed.

Maximizing the manufacturer's profit with respect to w_S gives the following first order condition:

$$\frac{\partial \pi_M}{\partial w_S} = \underbrace{\left(\frac{1}{\tau}\gamma_D \frac{w_D}{\tau+3}\right)}_{\frac{\partial \pi_{w_b,D}}{\partial w_S}} + \underbrace{\left(-\frac{(2\gamma_S w_S - 1)}{2}\right)}_{\frac{\partial \pi_{w_b,S}}{\partial w_S}} + \underbrace{\left(-\frac{2\gamma_D w_S (\tau+1) - \tau(1-\tau) - \gamma_D w_D (1-\tau)}{\tau(\tau+3)(1-\tau)}\right)}_{\frac{\partial \pi_{w_\beta}}{\partial w_S}}$$
$$+ \underbrace{\left(\frac{2(2\tau + \gamma_D w_S - \gamma_D w_D (1+\tau))}{\tau(\tau+3)^2}\right)}_{\frac{\partial \phi_D}{\partial w_S}} + \underbrace{\left(-\frac{(1-\gamma_S w_S)}{2} + -\frac{2(\tau+1)(\tau(1-\tau) - \gamma_D w_S (\tau+1) + \gamma_D w_D (1-\tau}{\tau(1-\tau)(\tau+3)^2}\right.}_{\frac{\partial \phi_S}{\partial w_S}}$$

With respect to the destination country, an increase of w_S shifts demand from the parallel import to the locally sourced version by increasing the price for the parallel import by more than the price of the locally sourced version. This affects the wholesale profit from sales of the parallel import (third term) through a price effect and a quantity effect and decreases the fixed fee extracted from intermediary I_S (fifth term) by reducing the quantity of the parallel import. By increasing demand for the locally sourced version, an increase of w_S increases the wholesale profit from sales of the locally sourced version (first term) and the the fixed fee extracted from intermediary I_D (fourth term). With respect to the source country, an increase of w_S affects the wholesale profit from sales of the locally sourced version (second term) through a price effect and a quantity effect and decreases the fixed fee extracted from intermediary I_S (fifth term).

Equilibrium drug prices are

$$p_{b,D} = \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}, \; p_{\beta,D} = \frac{(1-\tau)(2\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \text{ and} \tag{72}$$

$$p_{b,S} = \frac{4\gamma_D + 3\gamma_S(1-\tau^2)}{2\gamma_S(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \tag{73}$$

As compared to segmented markets, the price for the locally sourced version of the drug in country D is lower under parallel trade:

$$\frac{p_{b,D}}{p^*_{b,D}} = \frac{4\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} < 1. \tag{74}$$

The total change of the drug price is the net of a competition effect and a double marginalization effect[9]. In country S, the drug price is higher under parallel trade, as the wholesale price w_S is higher under parallel trade:

$$\frac{p_{b,S}}{p^*_{b,S}} = \frac{\left(4\gamma_D + 3\gamma_S\left(1-\tau^2\right)\right)}{\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} > 1. \tag{75}$$

Equilibrium quantities are

$$q_{b,D} = \frac{2\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)},\ q_{\beta,D} = \frac{\left(1-\tau\right)\gamma_S}{4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)} \text{ and} \tag{76}$$

$$q_{b,S} = \frac{4\gamma_D - \gamma_S\left(1-3\tau\right)\left(1-\tau\right)}{2\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)}. \tag{77}$$

n country D, the quantity of the locally sourced version of the drug $q_{b,D}$ is lower under parallel trade if $\gamma_D < \frac{\left(1-6\tau+5\tau^2\right)}{4}\gamma_S$, but the total quantity of both versions of the drug, $q_{b,D}+ q_{\beta,D}$ is higher under parallel trade than under segmented markets.

$$\begin{aligned}
\frac{q_{b,D}}{q^*_{b,D}} &= \frac{8\left(\gamma_D + \tau\gamma_S\left(1-\tau\right)\right)}{\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} < 1, \text{ if } \gamma_D < \frac{\left(1-6\tau+5\tau^2\right)}{4}\gamma_S, \\
\frac{q_{b,D}+q_{b,D}}{q^*_{b,D}} &= \frac{8\gamma_D + 2\gamma_S\left(1-\tau\right)\left(4\tau+1\right)}{\left(4\gamma_D + \gamma_S\left(1-\tau\right)\left(3\tau+1\right)\right)} > 1.
\end{aligned} \tag{78}$$

In country S, the quantity sold, $q_{b,S}$ is lower under parallel trade than under segmented markets.

$$\frac{q_{b,S}}{q^*_{b,S}} = \frac{\left(4\gamma_D - \gamma_S\left(3\tau-1\right)\left(\tau-1\right)\right)}{\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} < 1. \tag{79}$$

The manufacturer may increase the profit allocable to the destination country. Competition from parallel trade has a profit-decreasing effect, but the strategic effect of reducing competition by increasing the wholesale prices and inducing higher retail prices works in opposite direction. If the effect of a higher wholesale profit from sales as parallel imports and a higher fixed fee extracted from intermediary I_S exceeds the effect of competition in the destination country, i.e. a lower wholesale profit from sales as locally sourced and a higher fixed fee extracted from intermediary I_D, the profit allocable to the destination country is higher under parallel trade. The strategic

[9] For $w_H = 0$, the drug price would be $p_{b,H}\left(w_H = 0\right) = \frac{2\tau(\tau+3)(\gamma_H+\tau\gamma_F(1-\tau))}{\gamma_H\left(4\gamma_H(3\tau+1)+\tau\gamma_F(1-\tau)(\tau+3)^2\right)} < \frac{2(\gamma_H+\tau\gamma_F(1-\tau))}{\gamma_H(4\gamma_H+\gamma_F(3\tau+1)(1-\tau))} = p_{b,H}$.

effect of reducing competition by increasing wholesale prices is crucial; the profit is always lower under direct sales in the destination country. The profit allocable to the source country is always lower due a double marginalization effect with a higher drug price and a lower quantity sold. The total effect of parallel trade on the manufacturer's profit depends on the relative size of these two effects and with it on price elasticity in both countries (i.e. coinsurance rates) and the substitutability of both products (i.e. the degree of vertical product differentiation).

The total profit is higher under parallel trade if γ_S is sufficiently low:

$$
\begin{aligned}
\frac{\pi_M^*}{\pi_M} &= \frac{\left(\gamma_S^2\gamma_D + \gamma_S\gamma_D^2 - 4\gamma_S\gamma_D + \gamma_S + \gamma_D\right)\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)}{4\gamma_D^2 + \gamma_S\gamma_D\left(\tau+1\right)\left(5-3\tau\right) + 4\tau\gamma_S^2\left(1-\tau\right)} < 1, \\
\text{if } \gamma_S &< \gamma_S^* = \frac{4\gamma_D\left(3\tau+1\right)\left(1-\tau\right) + \gamma_D^2\left(\tau+1\right)\left(3\tau-5\right) - \left(1-\tau\right)^2}{2\gamma_D\left(3\tau+1\right)\left(1-\tau\right)} \\
&\quad + \frac{\sqrt{\left(1-\tau\right)^4 - 8\gamma_D\left(3\tau+1\right)\left(1-\tau\right)^3 + 2\gamma_D^2\left(50\tau+69\tau^2+13\right)\left(1-\tau\right)^2 + \gamma_D^3\left(3-2\tau+3\tau^2\right)\left(8\left(3\tau+1\right)\left(1-\right.\right.}}{2\gamma_D\left(3\tau+1\right)\left(1-\tau\right)}
\end{aligned}
$$

Appendix B: The Effect of Contract Choice on the Effectiveness of Regulation under Segmented Markets

Whether the two regulatory instruments, maximum markups and mandatory rebates also have an effect, in other words lower prices, under segmented markets, depends on the form of contract between manufacturer and intermediary: Under two-part tariffs, the increase of wholesale prices neutralizes the effect of maximum markups completely and mandatory rebates cannot applied, as wholesale prices are equal to zero. Under linear pricing, however, both regulatory instruments have an effect on drug prices.

Two-part tariff

No regulation

When parallel trade is not allowed and markets are segmented, pricing decisions by the manufacturer with respect to both countries – wholesale prices w_D and w_S, which determine drug prices in both countries – are independent.

The manufacturers profit is given as

$$
\pi_M^* = \underbrace{w_D^*\left(1-\gamma_D p_{b,D}^*\right)}_{\pi_{w_b,D}^*} + \underbrace{w_S^*\left(1-\gamma_S p_{b,S}^*\right)}_{\pi_{w_b,S}^*} + \phi_D^* + \phi_S^*, \tag{80}
$$

where $\pi_{w_b,D}^*$ and $\pi_{w_b,S}^*$ denote the wholesale profit from the intermediaries' sales in country D and S resp. and ϕ_D^* and ϕ_S^* the fixed fees, which are used to extract the intermediaries' profits.

For the intermediary I_D, profit is given as:

$$\pi^*_{I_D} = \underbrace{\left(p^*_{b,D} - w^*_D\right)\left(1 - \gamma_D p^*_{b,D}\right)}_{\pi^*_{b,D}} - \phi^*_D \tag{81}$$

and for the intermediary I_S as:

$$\pi^*_{I_S} = \underbrace{\left(p^*_{b,S} - w^*_S\right)\left(1 - \gamma_S p^*_{b,S}\right)}_{\pi^*_{b,S}} - \phi^*_S, \tag{82}$$

where $\pi^*_{b,D}$ and $\pi^*_{b,S}$ denote the profits from sales in country D and S, respectively.

In country D, the intermediary I_D maximizes (81) with respect to $p^*_{b,S.}$ The first order condition to this problem is

$$\left(1 - \gamma_D p^*_{b,D}\right) + \left(p^*_{b,D} - w^*_D\right)(-\gamma_S) = 0, \tag{83}$$

resulting in the monopoly drug price $p^*_{b,D} = \frac{(1+w^*_D\gamma_D)}{2\gamma_D}$. The drug price $p^*_{b,D}$ increases in the wholesale price w^*_D.

In country S, the intermediary I_S maximizes (82) with respect to $p^*_{b,S.}$ The first order condition to this problem is

$$\left(1 - \gamma_S p^*_{b,S}\right) + \left(p^*_{b,S} - w^*_S\right)(-\gamma_S) = 0, \tag{84}$$

resulting in the monopoly drug price $p^*_{b,S} = \frac{(1+w^*_S\gamma_S)}{2\gamma_S}$. The drug price $p^*_{b,S}$ increases in the wholesale price w^*_S.

Turning to the second stage of the game, the manufacturer M sets the fixed fees to

$$\phi^*_D = \pi^*_{b,D} = \frac{\left(1 - w^*_D\gamma_D\right)^2}{4\gamma_D} \text{ and } \phi^*_S = \pi^*_{b,S} = \frac{\left(1 - w^*_S\gamma_S\right)^2}{4\gamma_S}. \tag{85}$$

in order to extract the intermediaries' profits. In the absence of parallel trade and for segmented markets, the manufacturer's optimal strategy is to set the wholesale price equal to the marginal cost of production, i.e. $w^*_D = w^*_S = 0$[10]. This pricing decision avoids the double marginalization problem and results in the same drug price and sales volume as if the manufacturer sold directly to the consumers.

Equilibrium drug prices are

$$p^*_{b,D} = \frac{1}{2\gamma_D} \text{ and } p^*_{b,S} = \frac{1}{2\gamma_S}. \tag{86}$$

[10] This result can also be obtained by substituting (85) and equilibrium prices into (80) and maximizing with respect to w.

Equilibrium quantities are

$$q^*_{b,D} = \frac{1}{2},\ q^*_{b,S} = \frac{1}{2}. \tag{87}$$

Maximum markups

Under maximum markups, the regulatory body of country j restricts the markup surcharged to μ_j. Drug prices are

$$\begin{aligned} p^{*\mu}_{b,D} &= w^{*\mu}_D + (1-\mu_D)\, m^*_{b,D},\ m^*_{b,D} = p^*_{b,D} - w^*_D = \frac{1}{2\gamma_D} && (88)\\ p^{*\mu}_{b,S} &= w^{*\mu}_S + (1-\mu_S)\, m^*_{b,S},\ m^*_{b,S} = p^*_{b,S} - w^*_S = \frac{1}{2\gamma_S}. && (89)\end{aligned}$$

The manufacturers profit is given as

$$\pi^{*\mu}_M = \underbrace{w^{*\mu}_D\left(1-\gamma_D p^{*\mu}_{b,D}\right)}_{\pi^{*\mu}_{w_b,D}} + \underbrace{w^{*\mu}_S\left(1-\gamma_S p^{*\mu}_{b,S}\right)}_{\pi^{*\mu}_{w_b,S}} + \phi^{*\mu}_D + \phi^{*\mu}_S, \tag{90}$$

where $\pi^{*\mu}_{w_b,D}$ and $\pi^{*\mu}_{w_b,S}$ denote the wholesale profit from the intermediaries' sales in country D and S resp. and $\phi^{*\mu}_D$ and $\phi^{*\mu}_S$ the fixed fees, which are used to extract the intermediaries' profits.

For the intermediary I_D, profit is given as:

$$\pi^{*\mu}_{I_D} = \underbrace{(1-\mu_D)\, m^*_{b,D}\left(1-\gamma_D p^{*\mu}_{b,D}\right)}_{\pi^{*\mu}_{b,D}} - \phi^{*\mu}_D \tag{91}$$

and for the intermediary I_S as:

$$\pi^{*\mu}_{I_S} = \underbrace{(1-\mu_S)\, m^*_{b,S}\left(1-\gamma_S p^{*\mu}_{b,S}\right)}_{\pi^{*\mu}_{b,S}} - \phi^{*\mu}_S, \tag{92}$$

where $\pi^{*\mu}_{b,D}$ and $\pi^{*\mu}_{b,S}$ denote the profits from sales in country D and S, respectively.

Turning to the second stage of the game, the manufacturer M sets the fixed fees to

$$\begin{aligned} \phi^{*\mu}_D &= \pi^{*\mu}_{b,D} = (1-\mu_D)\, m^*_{b,D}\left(1-\gamma_D\left(w^{*\mu}_D + (1-\mu_D)\, m^*_{b,D}\right)\right)\\ \text{and } \phi^{*\mu}_S &= \pi^{*\mu}_{b,S} = (1-\mu_S)\, m^*_{b,S}\left(1-\gamma_S\left(w^{*\mu}_S + (1-\mu_S)\, m^*_{b,S}\right)\right) && (93)\end{aligned}$$

in order to extract the intermediaries' profits. Substituting (93) and equilibrium prices (89) into (90) and maximizing with respect to $w^{*\mu}_D$ and $w^{*\mu}_S$ gives the wholesale prices:

$$w^{*\mu}_D = \frac{\left(1-2\gamma_D(1-\mu_D)\, m^*_{b,D}\right)}{2\gamma_D} \text{ and } w^{*\mu}_S = \frac{\left(1-2\gamma_S(1-\mu_S)\, m^*_{b,S}\right)}{2\gamma_S}.$$

The manufacturer increases wholesale prices in response to maximum markups. The stricter regulation is, i.e. the higher μ_j, the higher the wholesale price is. An increase of the wholesale price decreases demand for the drug and accordingly the profit of the intermediary and the fixed fee. At the same time, there is a positive impact on the wholesale profit, if the effect from a higher price per unit offsets the effect from a lower quantity. The positive effect on the wholesale profit dominates and consequently, the manufacturer raises the wholesale price to the point that the effect from maximum markups on drug prices is neutralized: Equilibrium drug prices are

$$p_{b,D}^{*\mu} = \frac{1}{2\gamma_D} \text{ and } p_{b,S}^{*\mu} = \frac{1}{2\gamma_S}. \tag{94}$$

Higher wholesale prices offset the effect of lower markups allowed completely; drug prices under maximum markups are identical to drug prices under no regulation.

Equilibrium quantities are

$$q_{b,D}^{*\mu} = \frac{1}{2},\ q_{b,S}^{*\mu} = \frac{1}{2}. \tag{95}$$

Mandatory Rebates

As wholesale prices w_D^* and w_S^* are set to zero under segmented markets, this instrument cannot be applied.

Linear Pricing

No Regulation

Under linear pricing, the manufacturer charges a wholesale price per unit, but abstains from charging a fixed fee.

The manufacturers profit is given as

$$\pi_M^* = \underbrace{w_D^* \left(1 - \gamma_D p_{b,D}^*\right)}_{\pi_{w_b,D}^*} + \underbrace{w_S^* \left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{w_b,S}^*}, \tag{96}$$

where $\pi_{w_b,D}^*$ and $\pi_{w_b,S}^*$ denote the wholesale profit from the intermediaries' sales in country D and S resp..

For the intermediary I_D, profit is given as:

$$\pi_{I_D}^* = \underbrace{\left(p_{b,D}^* - w_D^*\right)\left(1 - \gamma_D p_{b,D}^*\right)}_{\pi_{b,D}^*} \tag{97}$$

and for the intermediary I_S as:

$$\pi_{I_S}^* = \underbrace{\left(p_{b,S}^* - w_S^*\right)\left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{b,S}^*}, \tag{98}$$

where $\pi^*_{b,D}$ and $\pi^*_{b,S}$ denote the profits from sales in country D and S, respectively.

In country D, the intermediary I_D maximizes (97) with respect to $p^*_{b,S.}$ The first order condition to this problem is

$$\left(1-\gamma_D p^*_{b,D}\right)+\left(p^*_{b,D}-w^*_D\right)(-\gamma_S)=0, \tag{99}$$

resulting in the monopoly drug price $p^*_{b,D}=\frac{(1+w^*_D\gamma_D)}{2\gamma_D}$. The drug price $p^*_{b,D}$ increases in the wholesale price w^*_D.

In country S, the intermediary I_S maximizes (98) with respect to $p^*_{b,S.}$ The first order condition to this problem is

$$\left(1-\gamma_S p^*_{b,S}\right)+\left(p^*_{b,S}-w^*_S\right)(-\gamma_S)=0, \tag{100}$$

resulting in the monopoly drug price $p^*_{b,S}=\frac{(1+w^*_S\gamma_S)}{2\gamma_S}$. The drug price $p^*_{b,S}$ increases in the wholesale price w^*_S.

Substituting equilibrium prices into (96) and maximizing with respect to w^*_D and w^*_S gives the wholesale prices:

$$w^*_D=\frac{1}{2\gamma_D} \text{ and } w^*_S=\frac{1}{2\gamma_S}.$$

Under linear pricing, the manufacturer extracts profits through the wholesale price instead of through the fixed fee, as under two part-tariffs. Accordingly, the manufacturer sets profit-maximizing wholesale prices (= monopolistic wholesale prices) and ignores the impact on higher wholesale prices on the intermediary's profit, as it cannot be appropriated.

Equilibrium drug prices are

$$p^*_{b,D}=\frac{3}{4\gamma_D} \text{ and } p^*_{b,S}=\frac{3}{4\gamma_S}. \tag{101}$$

The intermediary surcharges a monopolistic markup on the wholesale price. Thus, the final drug price is comprised of two monopolistic markups (double marginalization effect) and is higher than if the intermediary sold directly.

Equilibrium quantities are

$$q^*_{b,D}=\frac{1}{4},\ q^*_{b,S}=\frac{1}{4}. \tag{102}$$

Maximum Markups

Under maximum markups, the regulatory body of country j restricts the markup surcharged to μ_j. Drug prices are

$$p^{*\mu}_{b,D} = w^{*\mu}_D+(1-\mu_D)\,m^*_{b,D},\ m^*_{b,D}=p^*_{b,D}-w^*_D=\frac{1}{4\gamma_D} \tag{103}$$

$$p^{*\mu}_{b,S} = w^{*\mu}_S+(1-\mu_S)\,m^*_{b,S},\ m^*_{b,S}=p^*_{b,S}-w^*_S=\frac{1}{4\gamma_S}. \tag{104}$$

The manufacturers profit is given as

$$\pi_M^{*\mu} = \underbrace{w_D^{*\mu}\left(1-\gamma_D p_{b,D}^{*\mu}\right)}_{\pi_{w_b,D}^{*\mu}} + \underbrace{w_S^{*\mu}\left(1-\gamma_S p_{b,S}^{*\mu}\right)}_{\pi_{w_b,S}^{*\mu}}, \tag{105}$$

where $\pi_{w_b,D}^{*\mu}$ and $\pi_{w_b,S}^{*\mu}$ denote the wholesale profit from the intermediaries' sales in country D and S resp..

For the intermediary I_D, profit is given as:

$$\pi_{I_D}^{*\mu} = \underbrace{(1-\mu_D)\, m_{b,D}^{*}\left(1-\gamma_D p_{b,D}^{*\mu}\right)}_{\pi_{b,D}^{*\mu}} \tag{106}$$

and for the intermediary I_S as:

$$\pi_{I_S}^{*\mu} = \underbrace{(1-\mu_S)\, m_{b,S}^{*}\left(1-\gamma_S p_{b,S}^{*\mu}\right)}_{\pi_{b,S}^{*\mu}}, \tag{107}$$

where $\pi_{b,D}^{*\mu}$ and $\pi_{b,S}^{*\mu}$ denote the profits from sales in country D and S, respectively.

Substituting equilibrium prices (104) into (105) and maximizing with respect to $w_D^{*\mu}$ and $w_S^{*\mu}$ gives the wholesale prices:

$$w_D^{*\mu} = \frac{1-\gamma_D\,(1-\mu_D)\, m_{b,D}}{2\gamma_D} \text{ and } w_S^{*\mu} = \frac{1-\gamma_S\,(1-\mu_S)\, m_{b,S}}{2\gamma_S}.$$

Wholesale prices are lower than under no regulation. The stricter regulation is, i.e. the higher μ_j, the higher the wholesale price is.

Equilibrium drug prices are

$$p_{b,D}^{*\mu} = \frac{(1+\gamma_D\,(1-\mu_D)\, m_{b,D})}{2\gamma_D} = \frac{(5-\mu_D)}{8\gamma_D} \tag{108}$$

$$\text{and } p_{b,S}^{*\mu} = \frac{(1+\gamma_S\,(1-\mu_S)\, m_{b,S})}{2\gamma_S} = \frac{(5-\mu_S)}{8\gamma_S}. \tag{109}$$

Drug prices are lower than under no regulation. That is, under linear pricing, the manufacturer cannot offset the effect of maximum markups by raising wholesale prices. The stricter regulation is, i.e. the higher μ_j, the higher the drug price is.

Equilibrium quantities are

$$q_{b,D}^{*\mu} = \frac{(3+\mu_D)}{8},\; q_{b,S}^{*\mu} = \frac{(3+\mu_S)}{8}. \tag{110}$$

Quantities are higher than under no regulation. The stricter regulation is, i.e. the higher μ_j, the lower the quantity is.

Mandatory Rebates

Under mandatory rebates, wholesale prices are discounted by the factor ψ_j in country j. In country D, the wholesale price amounts to:

$$w_D^{*\psi} = (1 - \psi_D)\, w_D^* = (1 - \psi_D)\, \frac{1}{2\gamma_D} \tag{111}$$

and in country S to:

$$w_S^{*\psi} = (1 - \psi_S)\, w_S^* = (1 - \psi_S)\, \frac{1}{2\gamma_S}. \tag{112}$$

For the intermediary I_D, profit is given as:

$$\pi_{I_D}^{*\psi} = \underbrace{\left(p_{b,D}^{*\psi} - w_D^{*\psi}\right)\left(1 - \gamma_D p_{b,D}^{*\psi}\right)}_{\pi_{b,D}^{*\psi}} \tag{113}$$

and for the intermediary I_S as:

$$\pi_{I_S}^{*} = \underbrace{\left(p_{b,S}^{*\psi} - w_S^{*\psi}\right)\left(1 - \gamma_S p_{b,S}^{*\psi}\right)}_{\pi_{b,S}^{*\psi}}, \tag{114}$$

where $\pi_{b,D}^{*\psi}$ and $\pi_{b,S}^{*\psi}$ denote the profits from sales in country D and S, respectively.

In country D, the intermediary I_D maximizes (97) with respect to $p_{b,S.}^{*\psi}$ The first order condition to this problem is

$$\left(1 - \gamma_D p_{b,D}^{*\psi}\right) + \left(p_{b,D}^{*\psi} - w_D^{*\psi}\right)(-\gamma_S) = 0, \tag{115}$$

resulting in the monopoly drug price $p_{b,D}^{*\psi} = \frac{\left(1 + w_D^{*\psi}\gamma_D\right)}{2\gamma_D}$.

In country S, the intermediary I_S maximizes (98) with respect to $p_{b,S.}^{*\psi}$ The first order condition to this problem is

$$\left(1 - \gamma_S p_{b,S}^{*\psi}\right) + \left(p_{b,S}^{*\psi} - w_S^{*\psi}\right)(-\gamma_S) = 0, \tag{116}$$

resulting in the monopoly drug price $p_{b,S}^{*\psi} = \frac{\left(1 + w_S^{*\psi}\gamma_S\right)}{2\gamma_S}$.

Equilibrium drug prices are

$$p_{b,D}^{*\mu} = \frac{(3 - \psi_D)}{4\gamma_D} \text{ and } p_{b,S}^{*\mu} = \frac{(3 - \psi_S)}{4\gamma_S}. \tag{117}$$

Drug prices are lower than under no regulation. That is, an obligatory discount on wholesale prices is passed on to drug prices. However, intermediaries do not pass discounts on completely.

Equilibrium quantities are

$$q^*_{b,D} = \frac{(1+\psi_D)}{4},\ q^*_{b,S} = \frac{(1+\psi_S)}{4}. \tag{118}$$

Quantities are higher than under no regulation.

7 Appendix C: Equilibrium with Maximum Markups

Under maximum markups, the regulatory body of country j restricts the markup surcharged to a fraction μ_j of the markup surcharged under unregulated markets[11]:

$$\begin{aligned} p^{\mu}_{b,D} &= w^{\mu}_D + (1-\mu_D)\, m_{b,D}, \\ \text{with } m_{b,D} &= p_{b,D} - w_D = \frac{2\tau\left(\gamma_D + \tau\gamma_S(1-\tau)\right)}{\gamma_D\left(4\gamma_D + \gamma_S(3\tau+1)(1-\tau)\right)}, \end{aligned} \tag{119}$$

$$\begin{aligned} p^{\mu}_{\beta,D} &= w^{\mu}_S + (1-\mu_D)\, m_{\beta,D}, \\ \text{with } m_{\beta,D} &= p_{\beta,D} - w_S = \frac{\tau\gamma_S(\tau-1)^2}{\gamma_D\left(4\gamma_D + \gamma_S(3\tau+1)(1-\tau)\right)}, \end{aligned} \tag{120}$$

$$\begin{aligned} \text{and } p^{\mu}_{b,S} &= w^{\mu}_S + (1-\mu_S)\, m_{b,S}, \\ \text{with } m_{b,S} &= p_{b,S} - w_S = \frac{4\gamma_D - \gamma_S(1-3\tau)(1-\tau)}{2\gamma_S\left(4\gamma_D + \gamma_S(3\tau+1)(1-\tau)\right)}. \end{aligned} \tag{121}$$

The case of $\mu_j = 0$ corresponds to no restrictions on pricing, i.e. no regulation (the intermediary may charge the profit-maximizing markup as under no regulation), while the case of $\mu_j = 1$ corresponds to the strictest regulation possible (the intermediary is forced to price at marginal cost).

7.1 Maximum Markup Regulation in the Destination Country

Consider first the case of the destination country restricting pricing by intermediaries by restricting markups. Assume that markups are completely restricted, i.e. $\mu_D = 1$. In the source country, pricing by intermediary I_S is free.

[11] In all European countries except for Italy maximum wholesale markups are defined in terms of wholesale prices.

The manufacturer's profit is given as:

$$\pi_M^\mu = \underbrace{w_D^\mu \left(1 - \frac{\gamma_D (p_{b,D}^\mu - p_{\beta,D}^\mu)}{\tau}\right)}_{\pi^\mu_{w_b,D}} + \underbrace{w_S^\mu \left(1 - \gamma_S p_{b,S}^\mu\right)}_{\pi^\mu_{w_b,S}} + \underbrace{w_S^\mu \left(\frac{\gamma_D (p_{b,D}^\mu - p_{\beta,D}^\mu)}{\tau} - \frac{\gamma_D p_{\beta,D}^\mu}{(1-\tau)}\right)}_{\pi^\mu_{w_\beta}} + \phi_D^\mu + \phi_S^\mu, \tag{122}$$

with fixed fees given as

$$\phi_D^\mu = 0 \text{ and } \phi_S^\mu = \frac{(1 - w_S^\mu \gamma_S)^2}{4\gamma_S}. \tag{123}$$

Substituting (123), and equilibrium prices into (122) and maximizing with respect to w_D and w_S gives the wholesale prices $w_D^\mu = \frac{(\tau + 2\gamma_D w_S^\mu)}{2\gamma_D}$ and $w_S^\mu = \frac{4\gamma_D w_D^\mu (1-\tau)}{4\gamma_D + \tau\gamma_S(1-\tau)}$. Equilibrium wholesale prices are given as

$$w_D^\mu = \frac{(4\gamma_D + \tau\gamma_S (1-\tau))}{2\gamma_D (4\gamma_D + \gamma_S (1-\tau))} \text{ and } w_S^\mu = \frac{2(1-\tau)}{(4\gamma_D + \gamma_S (1-\tau))}. \tag{124}$$

Both wholesale prices are higher than under free pricing:

$$\begin{aligned} w_D^\mu - w_D &= \frac{\tau \left(16\gamma_D^2 + 4\gamma_S\gamma_D (4\tau+1)(1-\tau) + \gamma_S^2 (7\tau - 3)(1-\tau)^2\right)}{2\gamma_D (4\gamma_D + \gamma_S (1-\tau))(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} > 0, \\ w_S^\mu - w_S &= \frac{6\tau\gamma_S (1-\tau)^2}{(4\gamma_D + \gamma_S (1-\tau))(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} > 0. \end{aligned} \tag{125}$$

Equilibrium drug prices are:

$$\begin{aligned} p_{b,D}^\mu &= \frac{(4\gamma_D + \tau\gamma_S (1-\tau))}{2\gamma_D (4\gamma_D + \gamma_S (1-\tau))}, p_{\beta,D}^\mu = \frac{2(1-\tau)}{(4\gamma_D + \gamma_S (1-\tau))} \\ \text{and } p_{b,S}^\mu &= \frac{4\gamma_D + 3\gamma_S (1-\tau)}{2\gamma_S (4\gamma_D + \gamma_S (1-\tau))}. \end{aligned} \tag{126}$$

In the destination country, the price of the locally sourced version is lower than under free pricing, the price of the parallel import may be higher or lower than under free pricing. In the source country, the drug price is higher than under free pricing:

$$\begin{aligned} p_{b,D}^\mu - p_{b,D} &= -\frac{3\tau\gamma_S^2 (1-\tau)^3}{2\gamma_D (4\gamma_D + \gamma_S (1-\tau))(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} < 0 \\ p_{\beta,D}^\mu - p_{\beta,D} &= \frac{(2\gamma_D - \gamma_S (1-\tau))\tau\gamma_S (1-\tau)^2}{\gamma_D (4\gamma_D + \gamma_S (1-\tau))(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} \lesseqgtr 0, \\ p_{b,S}^\mu - p_{b,S} &= \frac{3\tau\gamma_S (1-\tau)^2}{(4\gamma_D + \gamma_S (1-\tau))(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} > 0 \end{aligned} \tag{127}$$

Equilibrium quantities are:

$$\begin{aligned}
q^{\mu}_{b,D} &= \frac{1}{2}, \\
q^{\mu}_{\beta,D} &= \frac{\gamma_S(1-\tau)}{2(4\gamma_D+\gamma_S(1-\tau))} \text{ and} \\
q^{\mu}_{b,S} &= \frac{(4\gamma_D-\gamma_S(1-\tau))}{2(4\gamma_D+\gamma_S(1-\tau))}.
\end{aligned} \tag{128}$$

In the destination country, maximum markups shift demand from the parallel import to the locally sourced version. In the source country, the quantity is lower than under free pricing:

$$\begin{aligned}
q^{\mu}_{b,D}-q_{b,D} &= \frac{\gamma_S(\tau-1)^2}{2(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))} > 0 \\
q^{\mu}_{\beta,D}-q_{\beta,D} &= -\frac{\gamma_S(1-\tau)(4\gamma_D+\gamma_S(1-3\tau)(1-\tau))}{2(4\gamma_D+\gamma_S(1-\tau))(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))} < 0 \\
q^{\mu}_{b,S}-q_{b,S} &= -\frac{3\tau\gamma_S^2(1-\tau)^2}{(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))(4\gamma_D+\gamma_S(1-\tau))} < 0.
\end{aligned} \tag{129}$$

7.2 Maximum Markup Regulation in the Source Country

Consider now the case of the source country restricting pricing by intermediary I_S by restricting his markup. Assume that the markup is completely restricted, i.e. $\mu_S = 1$. In the destination country, pricing by intermediaries is free.

The manufacturer's profit is given as:

$$\pi^{\mu}_M = \underbrace{w^{\mu}_D\left(1-\frac{\gamma_D(p^{\mu}_{b,D}-p^{\mu}_{\beta,D})}{\tau}\right)}_{\pi^{\mu}_{w_b,D}} + \underbrace{w^{\mu}_S\left(1-\gamma_S p^{\mu}_{b,S}\right)}_{\pi^{\mu}_{w_b,S}} + \underbrace{w^{\mu}_S\left(\frac{\gamma_D(p^{\mu}_{b,D}-p^{\mu}_{\beta,D})}{\tau}-\frac{\gamma_D p^{\mu}_{\beta,D}}{(1-\tau)}\right)}_{\pi^{\mu}_{w_\beta}} + \phi^{\mu}_D + \phi^{\mu}_S, \tag{130}$$

with fixed fees given as

$$\phi^{\mu}_D = \frac{(2\tau+\gamma_D w^{\mu}_S-\gamma_D w^{\mu}_D(\tau+1))^2}{\tau\gamma_D(\tau+3)^2} \text{ and } \phi^{\mu}_S = \frac{(\gamma_D w^{\mu}_D(1-\tau)+(1-\tau)\tau-\gamma_D w^{\mu}_S(1+\tau))^2}{\tau\gamma_D(1-\tau)(\tau+3)^2}. \tag{131}$$

Substituting (131), and equilibrium prices into (123) and maximizing with respect to w_D and w_S gives the wholesale prices $w^{\mu}_D = \frac{(1-\tau)\left(2\tau+\gamma_D w^{\mu}_S\right)}{\gamma_D(3\tau+1)}$ and $w^{\mu}_S = \frac{(1-\tau)\left(\tau\left(14+5\tau+\tau^2\right)+2\gamma_D w^{\mu}_D(1-\tau)\right)}{2\left(\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}$. Equilibrium wholesale prices are given as

$$w^{\mu}_D = \frac{(1-\tau)(\gamma_D(2-\tau)+4\tau\gamma_S(1-\tau))}{2\gamma_D(\gamma_D+\gamma_S(3\tau+1)(1-\tau))} \text{ and } w^{\mu}_S = \frac{(3\tau+2)(1-\tau)}{2(\gamma_D+\gamma_S(3\tau+1)(1-\tau))}. \tag{132}$$

Both wholesale prices are higher than under no regulation:

$$\begin{aligned} w_D^\mu - w_D &= \frac{(1-\tau)^2 \left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{2\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} > 0, \\ w_S^\mu - w_S &= \frac{(3\tau+1)(1-\tau)\left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{2\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} > 0. \end{aligned} \tag{133}$$

Equilibrium drug prices are:

$$\begin{aligned} p_{b,D}^\mu &= \frac{(2-\tau)\gamma_D + 4\tau\gamma_S (1-\tau)}{2\gamma_D \left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)}, \; p_{\beta,D}^\mu = \frac{(1-\tau)\left(\gamma_D (2+\tau) + 2\tau\gamma_S (1-\tau)\right)}{2\gamma_D \left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} \\ \text{and } p_{b,S}^\mu &= \frac{(3\tau+2)(1-\tau)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)}. \end{aligned} \tag{134}$$

In the destination country, prices are higher than under free pricing. In the source country, the price is lower than under free pricing:

$$\begin{aligned} p_{b,D}^\mu - p_{b,D} &= \frac{(1-\tau)\left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} > 0 \\ p_{\beta,D}^\mu - p_{\beta,D} &= \frac{(\tau+1)(1-\tau)\left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} > 0, \\ p_{b,S}^\mu - p_{b,S} &= -\frac{4\gamma_D^2 - \gamma_S\gamma_D (1-3\tau)(1-\tau) + \gamma_S^2 (3\tau+1)(1-\tau)^2}{2\gamma_S \left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} < 0 \end{aligned} \tag{135}$$

Equilibrium quantities are:

$$\begin{aligned} q_{b,D}^\mu &= \frac{\gamma_D (2-\tau) + 4\tau\gamma_S (1-\tau 1)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)}, q_{\beta,D}^\mu = \frac{\gamma_S (1-\tau) - \gamma_D}{\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} \\ \text{and } q_{b,S}^\mu &= \frac{2\gamma_D + 3\tau\gamma_S (1-\tau)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)}. \end{aligned} \tag{136}$$

In the destination country, maximum markups shift demand from the parallel import to the locally sourced version. In the source country, the quantity is higher than under free pricing:

$$\begin{aligned} q_{b,D}^\mu - q_{b,D} &= \frac{\gamma_D (1-\tau)\left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} \\ q_{\beta,D}^\mu - q_{\beta,D} &= -\frac{\gamma_D \left(4\gamma_D - \gamma_S (2-3\tau)(1-\tau)\right)}{\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)} \\ q_{b,S}^\mu - q_{b,S} &= \frac{4\gamma_D^2 - \gamma_S\gamma_D (1-3\tau)(1-\tau) + \gamma_S^2 (3\tau+1)(1-\tau)^2}{2\left(\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)\left(4\gamma_D + \gamma_S (3\tau+1)(1-\tau)\right)}. \end{aligned} \tag{137}$$

8 Appendix D: Mandatory Rebates

Under mandatory rebates, wholesale prices are discounted by the factor ψ_j in country j. In country D, the wholesale price amounts to:

$$w_D^{\psi} = (1 - \psi_D)\, w_D \tag{138}$$

and in country S to:

$$w_S^{\psi} = (1 - \psi_S)\, w_S. \tag{139}$$

8.1 Mandatory Rebates in the Destination Country

Consider first the case of the destination country implementing marginal cost pricing and forcing the manufacturer to set the wholesale price to zero, i.e. $\psi_D = 1$ and $w_D^{\psi} = 0$. In the source country, pricing is free.

The manufacturer's profit is given as:

$$\pi_M^{\psi} = \underbrace{w_S^{\psi}\left(1 - \gamma_S p_{b,S}^{\psi}\right)}_{\pi_{w_b,S}^{\psi}} + \underbrace{w_S^{\psi}\left(\frac{\gamma_D(p_{b,D}^{\psi} - p_{\beta,D}^{\psi})}{\tau} - \frac{\gamma_D p_{\beta,D}^{\psi}}{(1-\tau)}\right)}_{\pi_{w_\beta}^{\psi}} + \phi_D^{\psi} + \phi_S^{\psi}. \tag{140}$$

with the fixed fees given as

$$\phi_D^{\psi} = \frac{\left(2\tau + \gamma_D w_S^{\psi}\right)^2}{\tau\gamma_D\,(\tau+3)^2},\ \phi_S^{\psi} = \frac{\left(1 - w_S^{\psi}\gamma_S\right)^2}{4\gamma_S} \tag{141}$$

Substituting (141), and equilibrium prices into (140) and maximizing with respect to w_S^{ψ} gives the wholesale price:

$$w_S^{\psi} = \frac{2\tau\,(1-\tau)\,(5-\tau)}{\left(4\gamma_D\,(3\tau+1) + \tau\gamma_S\,(1-\tau)\,(\tau+3)^2\right)} \tag{142}$$

The wholesale price is lower than under free pricing:

$$w_S^{\psi} - w_S = -\frac{8\,(1-\tau)^3\,(\gamma_D + \tau\gamma_S\,(1-\tau))}{\left(4\gamma_D\,(3\tau+1) + \tau\gamma_S\,(1-\tau)\,(\tau+3)^2\right)(4\gamma_D + \gamma_S\,(3\tau+1)\,(1-\tau))} < 0. \tag{143}$$

Equilibrium drug prices are given as:

$$\begin{aligned}
p_{b,D}^{\psi} &= \frac{2\tau(\tau+3)(\gamma_D+\tau\gamma_S(1-\tau))}{\gamma_D\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}, \\
p_{\beta,D}^{\psi} &= \frac{\tau(1-\tau)(8\gamma_D+\tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}, \\
p_{b,S}^{\psi} &= \frac{\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)\left(4\tau+\tau^2+19\right)\right)}{2\gamma_S\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}.
\end{aligned} \tag{144}$$

All prices are lower than under free pricing:

$$\begin{aligned}
p_{b,D}^{\psi}-p_{b,D} &= -\frac{4(1-\tau)(\gamma_D+\tau\gamma_S(1-\tau))(2\gamma_D(\tau+1)+\tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))}<0, \\
p_{\beta,D}^{\psi}-p_{\beta,D} &= -\frac{2(1-\tau)^2(\gamma_D+\tau\gamma_S(1-\tau))(4\gamma_D+\tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))}<0, \\
p_{b,S}^{\psi}-p_{b,S} &= -\frac{4(1-\tau)^3(\gamma_D+\tau\gamma_S(1-\tau))}{\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))}<0.
\end{aligned} \tag{145}$$

Equilibrium quantities are

$$\begin{aligned}
q_{b,D}^{\psi} &= \frac{2(\tau+3)(\gamma_D+\tau\gamma_S(1-\tau))}{\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}, \\
q_{\beta,D}^{\psi} &= \frac{(1-\tau)(\tau\gamma_S(\tau+3)-2\gamma_D)}{\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)} \\
q_{b,S}^{\psi} &= \frac{4\gamma_D(3\tau+1)-\tau\gamma_S(1-\tau)\left(1-8\tau-\tau^2\right)}{2\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}.
\end{aligned} \tag{146}$$

In the destination country, mandatory rebates shift demand from the parallel import to the locally sourced version. In the source country, the quantity is higher than under free pricing:

$$\begin{aligned}
q_{b,D}^{\psi}-q_{b,D} &= \frac{2(1-\tau)(\gamma_D+\tau\gamma_S(1-\tau))(8\gamma_D+\gamma_S(\tau+3)(1-\tau)(\tau+1))}{(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}>0, \\
q_{\beta,D}^{\psi}-q_{\beta,D} &= -\frac{2(1-\tau)(\gamma_D+\tau\gamma_S(1-\tau))(4\gamma_D+\gamma_S(\tau+3)(1-\tau))}{(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}<0 \\
q_{b,S}^{\psi}-q_{b,S} &= \frac{4\gamma_S(1-\tau)^3(\gamma_D+\tau\gamma_S(1-\tau))}{(4\gamma_D+\gamma_S(3\tau+1)(1-\tau))\left(4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2\right)}>0.
\end{aligned}$$

8.2 Mandatory Rebates in the Source Country

Consider now the case of the source country implementing marginal cost pricing and forcing the manufacturer to set the wholesale price to zero, i.e. $\psi_S = 1$ and $w_S^\psi = 0$. In the destination country, pricing is free.

The manufacturer's profit is given as:

$$\pi_M^\psi = w_D^\psi \left(1 - \frac{\gamma_D(p_{b,D} - p_{\beta,D})}{\tau}\right) + \phi_D^\psi + \phi_S^\psi, \tag{147}$$

with fixed fees given as

$$\phi_D^\psi = \frac{\left(2\tau + \gamma_D w_S^\psi - \gamma_D w_D^\psi (\tau+1)\right)^2}{\tau\gamma_D (\tau+3)^2}, \; \phi_S^\psi = \frac{\left(\gamma_D w_D^\psi (1-\tau) + (1-\tau)\tau - \gamma_D w_S^\psi (1+\tau)\right)^2}{\tau\gamma_D (1-\tau)(\tau+3)^2}. \tag{148}$$

Substituting (148), and equilibrium prices into (147) and maximizing with respect to w_D^ψ gives the wholesale price:

$$w_D^\psi = 2\tau \frac{(1-\tau)}{\gamma_D (3\tau+1)}. \tag{149}$$

The wholesale price is lower than under free pricing:

$$w_D^\psi - w_D = -\frac{2(\tau-1)^2}{(3\tau+1)(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} < 0 \tag{150}$$

Equilibrium drug prices are given as

$$\begin{aligned} p_{b,D}^\psi &= \frac{2\tau}{\gamma_D (3\tau+1)}, \\ p_{\beta,D}^\psi &= \frac{(1-\tau)\tau}{\gamma_D (3\tau+1)}, \\ \text{and } p_{b,S}^\psi &= \frac{1}{2\gamma_S}. \end{aligned} \tag{151}$$

All prices are lower than under free pricing:

$$\begin{aligned} p_{b,D}^\psi - p_{b,D} &= -\frac{2(1-\tau)}{(3\tau+1)(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} < 0, \\ p_{\beta,D}^\psi - p_{\beta,D} &= -\frac{2(1-\tau)(\tau+1)}{(3\tau+1)(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} < 0, \\ p_{b,S}^\psi - p_{b,S} &= -\frac{(1-\tau)}{(4\gamma_D + \gamma_S (3\tau+1)(1-\tau))} < 0. \end{aligned} \tag{152}$$

Equilibrium quantities are:

$$q_{b,D}^{\psi} = 2\frac{\tau}{3\tau+1},\ q_{\beta,D}^{\psi} = \frac{1}{3\tau+1},\ q_{b,S}^{\psi} = \frac{1}{2}. \tag{153}$$

In the destination country, mandatory rebates shift demand from the locally sourced version to the parallel import. In the source country, the quantity is higher than under free pricing:

$$\begin{aligned}
q_{b,D}^{\psi} - q_{b,D} &= -2\gamma_D \frac{(1-\tau)}{(3\tau+1)\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} < 0, \\
q_{\beta,D}^{\psi} - q_{\beta,D} &= 4\frac{\gamma_D}{(3\tau+1)\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} > 0, \\
q_{b,S}^{\psi} - q_{b,S} &= \gamma_S \frac{(1-\tau)}{\left(4\gamma_D + \gamma_S\left(3\tau+1\right)\left(1-\tau\right)\right)} > 0.
\end{aligned} \tag{154}$$