Center for
European,
Governance and
Economic Development Research

Discussion Papers

Number 178 – December 2013

EXTERNALITIES OF NATIONAL PHARMACEUTICAL POLICY WHEN MARKETS ARE INTEGRATED THROUGH PARALLEL TRADE

Laura Birg

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

ISSN: 1439-2305

Externalities of national pharmaceutical policy when markets are integrated through parallel trade

Laura Birg*

December 2013

Abstract

This paper studies externalities of nationally determined cost-sharing systems, in particular coinsurance rates (patients pay a percentage of the price), under pharmaceutical parallel trade in a two-country model with a vertical distributor relationship. Parallel trade generates a price-decreasing competition effect in the destination country and a price-increasing double marginalization effect in the source country. An increase of the coinsurance rates in the destination country of the parallel import mitigates the double marginalization effect in the source country. An increase of the coinsurance rate in the source country reinforces the competition effect in the destination country. This may be a case for policy coordination in the European Union.

JEL classification: F12, I11, I18

Keywords: externalities, spillovers, parallel trade, cost-sharing, coinsurance rates

1 Introduction

This paper studies externalities of national decisions on health policy, in particular changes in coinsurance rates (patients pay a percentage of the price), under pharmaceutical parallel trade, i.e. trade outside the manufacturer's authorized distribution channel, in a two-country model with a vertical distributor relationship.

This analysis is motivated by the conflict between the consequences of parallel trade, namely market integration, and national competence in price regulation and reimbursement rules in the European Union.

On the one hand, the prevalence of pharmaceutical parallel trade, i.e. wholesalers or parallel traders being allowed to import pharmaceuticals from other countries without the permission of the manufacturer, is the result of market integration, in the European Union the internal

*Department of Economics, University of Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, laura.birg@wiwi.uni-goettingen.de. I would like to thank Horst Raff, Annika Herr, Jürgen Zerth and seminar participants at Berlin, Bochum, London, Glasgow, Göttingen, and Düsseldorf for helpful comments and suggestions.

market. The European Union has adopted regional exhaustion of intellectual property rights, which implies that parallel imports are legal within the European Union[1], but excluded if coming from non-member states. The European Court of Justice "has upheld the right to resell legitimately procured goods within the Community as a required safeguard for completing the internal market" (Maskus, 2000). On the other hand, pharmaceutical parallel trade is also a driving force for market integration and completing the internal market for pharmaceuticals. If parallel trade is legal and wholesalers perform parallel trade, pricing decisions of the pharmaceutical manufacturer in different markets become interdependent, as a low price in one market may induce parallel imports to a market with a higher price. Pharmaceutical parallel trade is the exploitation of these price differences, which may emerge e.g. from the monopolistic power of pharmaceutical manufacturers, allowing them to price-discriminate between different countries and/or divergent wholesale prices (NERA, 1999; EU Commission, 2003; Enemark et al., 2006). Consequently, a simple response to parallel trade by a pharmaceutical manufacturer would be the attempt to limit these price differences.

Market integration, further steps towards the completion of the internal market, requires non-prohibitive trade costs. Engaging in pharmaceutical parallel trade, i.e. importing a drug of identical chemical composition, dosage form and strength from another country involves obtaining a license (approximately €1500 in most countries) (Kyle, 2009). In addition, the parallel trader incurs repackaging costs to provide a package label and an insert in the language of the destination country (Kyle, 2009). This is offset by many destination countries providing incentives for patients to purchase 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. Also, market integration not only requires access of parallel traders to pharmaceutical distribution chains in other countries, but also improves access to pharmaceuticals in the destination countries by providing a lower-priced alternative to a brand-name drug, especially if cost-sharing systems sensitize patients for pharmaceutical prices. Then, market integration may result in the reduction of cross-country pharmaceutical price differences, either by manufacturer's responses to parallel trade (raising the price in the source country and/or lowering the price in the destination country) or by competition from parallel trade in the destination country.

Price convergence is predicted by several theoretical models[2], e.g. Rey (2003) or Jelovac & Bordoy (2005). Empirical evidence on this is, however, mixed. The two main studies on the effects of parallel trade, namely the ability of parallel trade to generate savings for health care systems and patients in the destination countries are a study by Kanavos et al. (2004) (commonly referred to as the LSE-study) and a study by West & Mahon (2003) (commonly referred to as the York-study). Kanavos et al. (2004) examine six product categories (19 products accounting for 21% of the market) in Denmark, Germany, the Netherlands, Norway, Sweden, and the UK and find no evidence for price competition or price convergence. On the contrary, West

[1] More precisely, parallel trade is allowed within the European Economic Area, which includes the European Union plus Norway and Switzerland.

[2] Other theoretical models even assume uniform prices under parallel trade, e.g. Pecorino (2002), Valetti (2006).

& Mahon (2003) who included the top-selling products plus a random sample of 150 products in Denmark, Germany, the Netherlands, and the UK observe indirect competitive effects in the parallel importing countries.

A more recent study by Granlund & Köksal analyzes Swedish drug prices for the period 2003 to 2007. They find that on average, drugs facing competition from parallel imports are priced at 17%-21% less as compared to what their prices would be if they had never faced such competition (Granlund & Köksal, 2011). Examining 1994-2003 data on prices of molecules that treat cardiovascular disease in France, Germany, Italy, Spain, and the United Kingdom, Timur, Picone & DeSimone (2010) suggest that cross-country differences between Germany and three of four other sample countries (France, Italy, Spain) have declined. They conclude that the European Union has come closer to achieving a single pharmaceutical market. Kyle, Allsbrook & Schulman (2008) who study the prices over 1000 pharmaceutical products in 30 countries over a 12-year period (1993-2004) also find that price differences have decreased in the European Union, where parallel trade is legal. But they also find that price differences have decreased less in countries of the European Union than in non-European Union countries, where parallel trade is not allowed.

Irrespective of whether parallel trade results in price convergence or not, price differences are a precondition for parallel trade. The profitability of performing cross-country arbitrage depends on substantial price differences. As mentioned above, these price differences may stem from a pharmaceutical manufacturer's price discrimination between different countries and/or differences in wholesale prices. In addition, different national pharmaceutical regulations in the individual member states may give rise to pharmaceutical price differences (Kanavos et al., 2004; Enemark et al., 2006). The Treaty on the Functioning of the European Union (TFEU), Art. 168, provides for national competence of member states in determining health policy, which includes the general organization of health care systems as well as pharmaceutical price regulation and cost-sharing systems.

So far, harmonization of different European rules has primarily concerned drug authorization procedures (Kyle; 2009). In 1995, two procedures were established, the Mutual Recognition Procedure (following approval in one reference member state a firm may launch the drug in other member states without additional applications) and approval by the European Medicines Evaluation Agency (European Union-wide approval) (Kyle, 2009). With respect to pharmaceutical price regulation and reimbursement, Directive 89/195/EC (so called Price Transparency Directive) is the only existing measure (Hancher, 2004). Originally, it was intended as a first, but retrospectively is the last measure with the objective of harmonizing national price regulation and reimbursement rules (Hancher; 2004). It provides rules for the control of pharmaceutical prices (respective measures have to be efficient, transparent and fair) amongst others (Desogus, 2011; Hancher, 2004). Among the member states, agreements on further harmonization could not be reached, although price differences have been considered as distortions (Desogus, 2011; Hancher, 2004). Consequently, drugs pricing – pharmaceutical price regulation and reimbursement rules – remain under exclusive national competence (Desogus, 2011). Given that direct harmonization of these rules is politically impossible, the European Commission has switched to soft law tools in recent years, pursuing price harmonization indirectly through trade liberalization, while regulation remains at a national level (Desogus, 2011). From the European Court of Justice's line of case law, it is clear that differences between national health care systems are not considered as obstacle to the free movement of goods (Desogus, 2011).

Regional exhaustion of property rights and the free movement of goods allow wholesalers or parallel traders to import pharmaceuticals from other countries without the authorization of the manufacturer, while health policy, including pharmaceutical price regulation and cost-sharing instruments, is in the national competence of member states. This implies that not only the manufacturer's pricing decisions but also national decisions on health policy may be interdependent. Externalities may emerge, whereby decisions about pharmaceutical price regulation and cost-sharing instruments in one country have an effect on drug prices and public health expenditure in other countries as well.

The implications of policy choices at the national level in a setting of markets being integrated by parallel trade have been analyzed by Raimondos-Møller & Schmitt (2010) for tax systems. They examine the interaction between commodity taxes and parallel imports when governments decide non-cooperatively on tax rates. They show that for an increasing volume of parallel imports origin taxes converge, while destination taxes diverge.

This paper studies externalities of national decisions in pharmaceutical policy, when markets are integrated by parallel trade. In particular, it shows that changes in coinsurance rates (patients pay a proportion of the drug price, health insurance reimburses the remainder) in one country have an impact on patients and health insurances in another country. Coinsurance rates are a cost-sharing instrument, which intends to restrict moral hazard in utilization of health services. Their design takes several health policy objectives into account: public health expenditure should be limited, but access to pharmaceuticals should be granted and there should be no excessive financial exposure of patients. When regulatory bodies set coinsurance rates in a setting of markets being integrated by parallel trade, they do not only have to balance these objectives for the respective country, but should also consider the impact on patients and health expenditure in other countries.

I analyze these externalities of coinsurance changes in a two-country model inspired by Maskus & Chen (2002) and Chen & Maskus (2005). It assumes a manufacturer that sells an innovative drug in two markets. In the home market, consumers purchase the drug directly from the manufacturer. In the foreign market the manufacturer markets the drug through an intermediary, which may engage in parallel trade and re-sell the drug in the home market. Parallel trade occurs as a by-product of the vertical control structure in the foreign country and flows from the foreign country as the source country to the home country as the destination country. When there is no parallel trade, the manufacturer's optimal strategy is to set a low wholesale price and extract the wholesaler's profit via a fixed fee to avoid the double marginalization problem arising from the intermediary's market power. However, in the presence of parallel trade, a low wholesale price induces more parallel trade. Consequently, the manufacturer may want to set the wholesale price higher in order to limit competition from parallel trade. The optimal wholesale price reflects the trade-off between an intensified double marginalization problem in the foreign market for a high wholesale price and increased competition from parallel trade in the home market due to a low wholesale price.

Parallel trade generates a competition effect in the destination country, resulting in lower drug prices and a higher quantity sold. The higher wholesale price (as compared to segmented markets) creates a double-marginalization effect with a higher drug price and a lower quantity sold. These results are also in line with Ganslandt & Maskus (2007). Parallel trade results in market integration, as it makes pricing decisions with respect to the different markets interdependent. In this setting, national decisions on coinsurance rates affect the trade-off between the double marginalization effect and the competition effect. By changes in the wholesale price, externalities occur. An increase of the coinsurance rate in the destination country mitigates the double marginalization effect in the source country (lower drug price, higher quantity); an increase of the coinsurance rate in the source country reinforces the competition effect from parallel trade in the destination country (lower drug prices, higher quantities).

The rest of the paper is organized as follows. In the next section, the two-country model with a vertical distributor relationship is presented and the case of segmented markets, when parallel trade is not allowed, and the case of integrated markets, when parallel trade is possible, are analyzed. In Section 3, the effects of parallel trade with respect to drug prices and price convergence are studied. Section 4 examines the implications of changes of cost-sharing instruments in the destination country for the source country and vice versa, section 5 discusses implications for health policy. Section 6 concludes.

2 The Model

Following Maskus & Chen (2002), (2005), consider a (domestic) manufacturer M selling a brand-name drug b in two countries, in its home country and a foreign country. In the home country, the manufacturer sells directly to the consumers; in the foreign country, it sells through an independent intermediary I. The manufacturer follows a two-part pricing strategy, it charges the intermediary a wholesale price w and a fixed fee ϕ.

In a regime of international exhaustion of intellectual property rights, due to lack of complete vertical control, the intermediary may engage in parallel trade and resell the drug in the home country. The sales of the drug as a parallel import are denoted by β. 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 the source country S buy the drug from the intermediary, consumers in the destination country D have the choice between the locally sourced version b when purchasing from the manufacturer and the parallel import β when buying from the intermediary. 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 appearance 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, $\gamma_j \in (0, 1)$ 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 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).

Health insurance reimburses a fraction of the drug price, the remaining fraction γ is paid by the patient. Thus, the effective price of the drug to the patient amounts to the proportion γ of the market price set by the manufacturer or intermediary (Zweifel et al., 2009).

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 directly from the manufacturer (b) or not purchasing at all (0), has a gross valuation $\theta_D^{b,0}$, given by

$$\theta_D^{b,0} - \gamma_D p_{b,D} = 0 \Leftrightarrow \theta_D^{b,0} = \gamma_D p_{b,D}. \tag{2}$$

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

$$q_{b,D} = 1 - \gamma_D p_{b,D}. \tag{3}$$

If parallel trade takes place, consumers in country D have the choice between the locally sourced version (b) (directly) from the manufacturer or the parallel import (β) from the intermediary. 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

$$\theta_D^{b,\beta} - \gamma_D p_{b,D}^* = \theta_D^{b,\beta}(1-\tau) - \gamma_D p_{\beta,D}^* \Leftrightarrow \theta_D^{b,\beta} = \frac{\gamma_D \left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau}. \tag{4}$$

An asterisk is used to denote variables associated with parallel trade.

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

$$\theta_D^{\beta,0}(1-\tau) - \gamma_D p_{\beta,D}^* = 0 \Leftrightarrow \theta_D^{\beta,0} = \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}. \tag{5}$$

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

$$q_{b,D}^* = 1 - \frac{\gamma_D \left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} \text{ and } q_{\beta,D}^* = \frac{\gamma_D \left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}. \tag{6}$$

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

$$\theta_S^{b,0} - \gamma_S p_{b,S} = 0 \iff \theta_S^{b,0} = \gamma_S p_{b,S}. \tag{7}$$

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

$$q_{b,S} = 1 - \gamma_S p_{b,S}. \tag{8}$$

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 two-stage game: In the first stage, the manufacturer specifies a wholesale price w and fixed fee ϕ. In the second and final stage, the intermediary 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 manufacturer sets the price for the locally sourced version in country D (that is, $p_{b,D}$).

2.1 Equilibrium without Parallel Trade

First consider the case where parallel trade is not allowed and markets are segmented. Both pricing decisions by the manufacturer – the drug price in country D and the wholesale price w that determines the drug price in country S – are independent.

The manufacturers profit is given as

$$\pi_M = \underbrace{p_{b,D}\left(1-\gamma_D p_{b,D}\right)}_{\pi_{b,D}} + \underbrace{w\left(1-\gamma_S p_{b,S}\right)}_{\pi_{w_b}} + \phi, \tag{9}$$

where $\pi_{b,D}$ denotes the monopoly profit from direct sales in country D, π_{w_b} the wholesale profit from the intermediary's sales in market S, and ϕ the fixed fee that is used to extract the intermediary's profit.

The wholesaler's total profit is given as

$$\pi_I = \underbrace{\left(p_{b,S}-w\right)\left(1-\gamma_S p_{b,S}\right)}_{\pi_{b,S}} - \phi, \tag{10}$$

where $\pi_{b,S}$ denotes the profit from sales in country S.

In market D, the manufacturer M sets the monopoly drug price $p_{b,D} = \frac{1}{2\gamma_D}$.

In market S, the intermediary I charges the monopoly drug price $p_{b,S} = \frac{(1+w\gamma_S)}{2\gamma_S}$. The drug price $p_{b,S}$ increases in the wholesale price w.

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

$$\phi = \pi_{b,S} = \frac{\left(1-w\gamma_S\right)^2}{4\gamma_S} \tag{11}$$

in order to extract the intermediary's profit. 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. $w = 0$[3]. 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

$$p_{b,D} = \frac{1}{2\gamma_D} \text{ and } p_{b,S} = \frac{1}{2\gamma_S}. \tag{12}$$

Prices decrease in coinsurance rates. Effective prices for consumers ($\gamma_D p_{b,D} = \frac{1}{2}$, $\gamma_S p_{b,S} = \frac{1}{2}$) are equivalent to prices without insurance coverage ($p_{b,D} = \frac{1}{2}$, $p_{b,S} = \frac{1}{2}$). That is, the effect

[3]Substituting (11) and equilibrium prices into (9) and maximizing with respect to w results in $w = 0$.

from reimbursement by health insurance is completely appropriated by the manufacturer. Price differences across countries result from differences in health care systems, i.e. coinsurance rates, only:

$$p_{b,D} - p_{b,S} = \frac{\gamma_S - \gamma_D}{2\gamma_D \gamma_S}. \tag{13}$$

That is, if $\gamma_S > \gamma_D$, country D is the high price-country and country S is the low price-country.

Equilibrium quantities are

$$q_{b,D} = \frac{1}{2},\ q_{b,S} = \frac{1}{2}. \tag{14}$$

Quantities are independent of coinsurance rates, as the effect from reimbursement completely accrues to the manufacturer. Health insurance refunds the fraction $(1 - \gamma_D)$ of the monopoly drug price $p_{b,D}$ per drug. Accordingly, in country D, public health expenditure amounts to

$$E_D(\gamma) = (1 - \gamma_D)\, p_{b,D} q_{b,D}. \tag{15}$$

Similarly, in country S, the fraction $(1 - \gamma_S)$ of the drug price $p_{b,S}$ is reimbursed per drug and public health expenditure is given as

$$E_S(\gamma) = (1 - \gamma_S)\, p_{b,S} q_{b,S}. \tag{16}$$

2.2 Equilibrium with Parallel Trade

If parallel trade is allowed, the manufacturer's pricing decisions – the drug price in country D and the wholesale price charged in country S – are no longer independent. A low wholesale price induces parallel imports sold by the intermediary in country D (the wholesale price constitutes the lower price bound for the intermediary). Increasing the wholesale price in response creates and aggravates a double marginalization problem in country S. Consequently, if parallel trade is allowed, the choice of the wholesale price reflects the trade-off between an aggravated double marginalization problem in country S and intensified competition from parallel trade in country D.

The manufacturer's profit is given as

$$\pi_M^* = \underbrace{p_{b,D}^* \left(1 - \frac{\gamma_D \left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau}\right)}_{\pi_b^*} + \underbrace{w^* \left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{w_b}^*} + \underbrace{w^* \left(\frac{\gamma_D \left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}\right)}_{\pi_{w_\beta}^*} + \phi^*, \tag{17}$$

where π_b^* denotes the profit from direct sale in D, $\pi_{w_b}^*$ the wholesale profit from the intermediary's sales in market S, $\pi_{w_\beta}^*$ the wholesale profit from the intermediary's sales as parallel imports in market D, and ϕ^* the fixed fee. Again, an asterisk is used to denote variables associated with parallel trade.

Parallel trade affects the manufacturer's profit in three ways: First, he faces competition by the intermediary in market D. Second, for a given wholesale price, the fixed fee extracted from the intermediary is higher, as it now also contains the intermediary's profit from parallel trade. Third, the intermediary's sales as reimports result in additional wholesale profit for the manufacturer.

The intermediary's profit is given by

$$\pi_I^* = \underbrace{\left(p_{b,S}^* - w^*\right)\left(1 - \gamma_S p_{b,S}^*\right)}_{\pi_{b,S}^*} + \underbrace{\left(p_{\beta,D}^* - w^*\right)\left(\frac{\gamma_D\left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}\right)}_{\pi_{\beta,D}^*} - \phi^*, \tag{18}$$

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

In country D, the manufacturer M maximizes (17) with respect to $p_{b,D}^*$. The first order condition of this problem is

$$\underbrace{\left(1 - \frac{\gamma_D\left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau}\right)}_{I} + \underbrace{p_{b,D}^*\left(-\frac{\gamma_D}{\tau}\right)}_{II} + \underbrace{w^*\left(\frac{\gamma_D}{\tau}\right)}_{III} = 0, \tag{19}$$

which yields the best response function $p_{b,D}^* = \frac{\tau}{2\gamma_D} + \frac{1}{2}\left(p_{\beta,D}^* + w^*\right)$. Compared to the first order condition for segmented markets, part I and consequently $p_{b,D}^*$ are higher (lower) under parallel trade, if $p_{b,D}^* < \frac{p_{\beta,D}^*}{(1-\tau)}$ ($p_{b,D}^* > \frac{p_{\beta,D}^*}{(1-\tau)}$). 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. Part III illustrates the indirect effect of competition from parallel trade: A larger volume of parallel imports results in a higher wholesale profit. A higher wholesale price results in a higher price for the locally sourced version, as it leads to less competition from parallel trade.

The intermediary maximizes (18) with respect to $p_{\beta,D}^*$ and $p_{b,S}^*$. The first order condition with respect to $p_{\beta,D}^*$ is

$$\left(\frac{\gamma_D\left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}\right) + \left(p_{\beta,D}^* - w^*\right)\left(-\frac{\gamma_D}{\tau} - \frac{\gamma_D}{1-\tau}\right) = 0 \tag{20}$$

and the best response function is $p_{\beta,D}^* = \frac{1}{2}\left(w^* + p_{b,D}^*(1-\tau)\right)$. Solving for equilibrium prices in country D results in $p_{b,D}^* = \frac{2\tau + 3w^*\gamma_D}{\gamma_D(3+\tau)}$ and $p_{\beta,D}^* = \frac{\tau(1-\tau) + w^*\gamma_D(3-\tau)}{\gamma_D(\tau+3)}$.

In country S, the intermediary maximizes (18) with respect to $p_{b,S}^*$. The first order condition to this maximization problem is

$$\left(1-\gamma_S p_{b,S}^*\right)+\left(p_{b,S}^*-w^*\right)\underbrace{\left(-\gamma_S\right)}_{\frac{\partial q_{b,S}^*}{\partial p_{b,S}^*}}=0, \tag{21}$$

resulting in the price $p_{b,S}^* = \frac{1+w^*\gamma_S}{2\gamma_S}$. The first order condition is identical to the first order condition, if parallel trade is not allowed. Note that as $p_{b,S}^*$ increases in the wholesale price w^*, $p_{b,S}^*$ will be higher under parallel trade, if $w^* > 0$.

With

$$\phi^* = \underbrace{\frac{\left(1-w\gamma_S\right)^2}{4\gamma_S}}_{\pi_{b,S}^*} + \underbrace{\frac{\tau\left(1-2w\gamma_D-\tau\right)^2}{\gamma_D\left(1-\tau\right)\left(3+\tau\right)^2}}_{\pi_{\beta,D}^*} \tag{22}$$

the manufacturer extracts the intermediary's total profit. Substituting (22) and equilibrium prices into (17) and maximizing with respect to w^* gives the wholesale price:

$$w^* = \frac{2\left(1-\tau\right)\left(9-5\tau\right)}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2}. \tag{23}$$

For segmented markets, the manufacturer's optimal strategy to avoid the double marginalization problem resulting from vertical separation in imperfectly competitive markets is to set the wholesale price equal to marginal cost, i.e. $w = 0$. However, if parallel trade is allowed and results in market integration, a low wholesale price induces more parallel trade. Consequently, the manufacturer will set a higher wholesale price to limit competition from parallel trade in country D. The optimal wholesale price w reflects the trade-off between an aggravated double marginalization problem in country S and intensified competition in country D.

Equilibrium drug prices are

$$p_{b,D}^* = \frac{2\gamma_D\left(9-5\tau\right)+2\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)}{\gamma_D[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}, \tag{24}$$

$$p_{\beta,D}^* = \frac{\left(1-\tau\right)[2\gamma_D\left(9-5\tau\right)+\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)]}{\gamma_D[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}, \tag{25}$$

and

$$p_{b,S}^* = \frac{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(27-4\tau+\tau^2\right)}{2\gamma_S[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}. \tag{26}$$

Equilibrium quantities are

$$q_{b,D}^* = \frac{2[\gamma_D\left(9-5\tau\right)+\gamma_S\left(3+\tau\right)\left(1-\tau\right)]}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2}, \tag{27}$$

$$q_{\beta,D}^* = \frac{\gamma_S\left(3+\tau\right)\left(1-\tau\right)}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2}, \tag{28}$$

and

$$q^*_{b,S} = \frac{4\gamma_D (9 - 5\tau) - \gamma_S (1 - \tau) \left(9 - 16\tau - \tau^2\right)}{2[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]}. \tag{29}$$

In the destination country, public drug expenditure is given as

$$E^*_D = (1 - \gamma_D) \left(p^*_{b,D} q^*_{b,D} + p^*_{\beta,D} q^*_{\beta,D}\right),$$

and in the source country, public drug expenditure is given as

$$E^*_S = (1 - \gamma_S) \, p^*_{b,S} q^*_{b,S}. \tag{30}$$

3 The Effect of Parallel Trade

This section investigates the effect of parallel trade on drug prices and quantities in the destination country and the source country. In addition, it explores whether cross-country arbitrage results in the erosion of price differences, i.e. price convergence.

3.1 Competition Effect in the Destination Country

In country D, parallel trade induces a competition effect with lower drug prices and a higher quantity sold, see Appendix A for details.

Compared to segmented markets, competition from parallel trade reduces the price of the drug sold directly by the manufacturer:

$$\frac{p^*_{b,D}}{p_{b,D}} < 1, \tag{31}$$

with the price of the parallel import being lower than the price of the locally sourced version:

$$\frac{p^*_{\beta,D}}{p^*_{b,D}} < 1. \tag{32}$$

The difference between the prices of the two versions of the drug stems from (perceived) vertical product differentiation: The intermediary has to compensate consumers for the lower (perceived) quality by pricing at a certain discount from a given price of the locally sourced drug version[4].

Accordingly, under parallel trade, the prices of both versions of the drug are lower than the monopoly drug price under segmented markets.

The quantity of the locally sourced version is higher under parallel trade:

$$\frac{q^*_{b,D}}{q_{b,D}} > 1. \tag{33}$$

[4]Note that the intermediary's best response function is $p^*_{\beta,H} = \frac{1}{2}\left(w + p^*_{b,H}(1 - \tau)\right)$.

Consequently, the total quantity of the drug available, that is, the quantity of the locally sourced version plus the parallel import, is higher than the monopoly quantity under segmented markets.

3.2 Double Marginalization Effect in the Source Country

In country S, parallel trade generates a double marginalization effect with a higher drug price and a lower quantity due to an increase of the wholesale price, see Appendix A for details.

Compared to segmented markets, the wholesale price w^* is higher under parallel trade. As a low wholesale price induces more parallel trade and consequently enhances the competition from parallel trade in the destination country D, the manufacturer raises the wholesale price in order to deter parallel trade partially:

$$w^* > w = 0. \tag{34}$$

The increase of the wholesale price induced by parallel trade translates to an increase of the drug price:

$$\frac{p^*_{b,S}}{p_{b,S}} > 1 \tag{35}$$

and the higher price reduces the quantity sold:

$$\frac{q^*_{b,S}}{q_{b,S}} < 1. \tag{36}$$

3.3 Price Convergence vs. Divergence

Parallel trade results in price convergence if it goes from the ex-ante low price country to the ex-ante high price country (i.e. if the pre-parallel trade drug price in the source country S is lower than the pre-parallel trade price in the destination country D[5]), see Appendix A for details:

$$\frac{p^*_{b,D} - p^*_{b,S}}{p_{b,D} - p_{b,S}} < 0 \text{ if } \gamma_S > \gamma_{D}. \tag{37}$$

The intuition is quite simple: If parallel trade goes from the low-price to the high price country, the double marginalization effect results in a higher price in the low price country and the competition effect lowers the price in the destination country, both reducing the price spread. On the contrary, if parallel trade goes from a high price to a low price country, it results in price divergence, as the double marginalization effect contributes to an even higher price in the high price country and the competition effect lowers the low price in the destination country. Although there is also evidence for parallel trade from high-price to low-price countries, the bulk of parallel trade goes from low-price to high-price countries.

[5] Note that under segmented markets, $p_{b,H} > p_{b,F}$, i.e. country H is the high price country and country F is the low price country, if $\gamma_F > \gamma_H$.

4 Policy Interdependence under Parallel Trade

This section investigates externalities of national health policy decisions on prices and quantities in the respective other country. In other words, this section analyzes pharmaceutical policy interdependence under parallel trade. Under segmented markets, there are no externalities of changes in coinsurance rates, as the manufacturer's pricing decisions in both markets are independent, see Appendix B for details.

As the reduction of rising health expenditure is one of the main objectives of pharmaceutical policy in many European countries, I analyze changes of cost-sharing instruments with the aim to reduce public expenditure. This corresponds to the reduction of reimbursed amounts and increases of copayments, more specifically increases of coinsurance rates. In the case of reductions of copayments, i.e. reductions of coinsurance rates, price and quantity changes go in the opposite direction.

4.1 Change of the Coinsurance Rate in the Destination Country

Consider first a change of the cost-sharing instrument in the destination country and its implications for the source country.

An increase in the coinsurance rate in the destination country D raises effective consumer prices, lowers the quantity consumed, and reduces health expenditure in the destination country D and lowers effective consumer prices, increases the quantity consumed, and raises health expenditure in the source country S. For explicit expressions of changes in prices and quantities, see Appendix C.

In the destination country, the increase of copayments, i.e. the increase of the coinsurance rate results in lower drug prices and lower quantities sold.

An increase of the coinsurance rate in country D increases the price elasticity of demand. As willingness to pay decreases, demand for the locally sourced version of the drug decreases c.p.:

$$\frac{\partial q_{b,D}^{*}}{\partial \gamma_D} < 0. \tag{38}$$

Consequently, the manufacturer lowers the price for the locally sourced version of the drug, as illustrated by the best response function: $p_{b,D}^{*} = \frac{1}{2}\left(\frac{\tau}{\gamma_D} + p_{\beta,D}^{*} + w\right)$. For the parallel import, demand increases, if the price difference between the locally sourced version and the parallel import exceeds the quality difference:

$$\frac{\partial q_{\beta,D}^{*}}{\partial \gamma_D} > 0, \text{ if } p_{\beta,D}^{*} < (1-\tau)\, p_{b,D}^{*}. \tag{39}$$

The direct effect of the price for the locally sourced version on the price for the parallel import, however, leads to a decrease of the price for the parallel import as well[6]. This is demonstrated

[6] In addition, 39 implies that the intermediary has to lower $p_{\beta,H}^{*}$ in order to prevent a decrease of demand.

by the best response function: $p^*_{\beta,D} = \frac{1}{2}\left(w + p^*_{b,D}\left(1-\tau\right)\right)$.

Accordingly, in country D both drug prices decrease in the coinsurance rate:

$$\frac{\partial p^*_{b,D}}{\partial \gamma_D} < 0,\ \frac{\partial p^*_{\beta,D}}{\partial \gamma_D} < 0. \tag{40}$$

Higher price elasticity under parallel trade of demand limits the ability to increase prices in response to an increase of the coinsurance rate and consequently, effective consumer prices increase:

$$\frac{\partial \gamma_D p^*_{b,D}}{\partial \gamma_D} > 0,\ \frac{\partial \gamma_D p^*_{\beta,D}}{\partial \gamma_D} > 0. \tag{41}$$

As price decreases cannot compensate the effect of lower demand, quantities of both versions of the drug decrease in γ_D:

$$\frac{\partial q^*_{b,D}}{\partial \gamma_D} < 0,\ \frac{\partial q^*_{\beta,D}}{\partial \gamma_D} < 0. \tag{42}$$

Lower prices and lower quantities consumed reduce the public health expenditure:

$$\frac{\partial E^*_D}{\partial \gamma_D} < 0. \tag{43}$$

Spillovers of copayment changes in country D to the source country S occur via the wholesale price, as the manufacturer's pricing decisions are interdependent under parallel trade. Representing the intermediary's marginal cost, the wholesale price is the lower bound for the drug price in country S and the price of the parallel import in country D. With respect to country S, a lower wholesale price is preferable for the manufacturer (limiting the double marginalization effect), with respect to country D, a higher wholesale price is in the interest of the manufacturer (limiting competition from parallel trade). The resulting wholesale price represents a trade-off between competition in D and double marginalization effect in S, with competition in D inducing an upward influence on the wholesale price and the successive monopoly position of the manufacturer and the intermediary in country S exerting downward pressure on the wholesale price. Although decreasing drug prices in country D could be considered as intensifying competition, the decrease of total demand reduces the effect of competition. The double marginalization effect gains relative importance and, accordingly, the wholesale price is lowered:

$$\frac{\partial w^*}{\partial \gamma_D} < 0. \tag{44}$$

The drug price in country S is a mark-up over the intermediary's marginal cost, which is the wholesale price w. (The intermediary's best response function is $p^*_{b,S} = \frac{1+w\gamma_S}{2\gamma_S}$). A decrease of the wholesale price then results in drug price decreases:

$$\frac{\partial p^*_{b,S}}{\partial \gamma_D} < 0. \tag{45}$$

As the coinsurance rate in the source country S is unchanged, the effective drug price decreases:

$$\frac{\partial \gamma_S p^*_{b,S}}{\partial \gamma_D} < 0, \tag{46}$$

which increases the quantity consumed:

$$\frac{\partial q^*_{b,S}}{\partial \gamma_D} > 0. \tag{47}$$

Thus, for increasing copayments in country D, the decrease of total demand reduces the relative importance of competition by parallel trade and the wholesale price is lowered, translating to a lower drug price and higher quantity sold in the source country S. In other words, the reduction of the competition by parallel trade enables the manufacturer to more follow the optimal strategy of setting a low wholesale price to avoid excessive mark-ups in the successive monopoly of manufacturer and intermediary. That is, a copayment increase in the destination country D mitigates the double marginalization effect in the source country S.

Health expenditure increases, as the effect from a higher quantity consumed exceeds the effect of a lower drug price (see Appendix D):

$$\frac{\partial E^*_S}{\partial \gamma_D} > 0. \tag{48}$$

Consequently, an increase of the coinsurance rate in the destination country D decreases demand and accordingly the importance of the competition from parallel imports, which results in a decrease of the wholesale price. This reduces marginal cost for the intermediary, which translates to a price reduction for the drug in the source country S and increase of the quantity consumed. By reducing drug prices and increasing the quantities sold, a copayment increase in the source country mitigates the double marginalization effect in the source country S.

Proposition 1 summarizes the effect of an increase in the coinsurance rate in the destination country D:

Proposition 1 *An increase in the coinsurance rate in the destination country D i) raises effective consumer prices, lowers the quantity consumed, and reduces health expenditure in country D, ii) lowers effective consumer prices, increases the quantity consumed in, and raises health expenditure in country S.*

4.2 Change of the Coinsurance Rate in the Source Country

Consider now a change of the cost-sharing instrument in the source country and its implications for the destination country.

An increase in the coinsurance rate in the source country S raises effective consumer prices, lowers the quantity consumed and reduces health expenditure in the source country S and lowers effective consumer prices, increases the quantity consumed and lowers health expenditure in the destination country country D. For explicit expressions of changes in prices and quantities see Appendix C.

In the source country, the increase of copayments, i.e. an increase of the coinsurance rate results, similarly to the effects in the destination country, in lower drug prices and lower quantities sold.

As willingness to pay decreases, demand for the drug decreases c.p.:

$$\frac{\partial\left(1-\gamma_S p^*_{b,S}\right)}{\partial\gamma_S}<0. \tag{49}$$

The intermediary then reduces the drug price in response, as illustrated by the best response function $p^*_{b,S}=\frac{1+w^*\gamma_S}{2\gamma_S}$.

Accordingly, the drug price decreases in γ_S:

$$\frac{\partial p^*_{b,S}}{\partial\gamma_S}<0. \tag{50}$$

The effective drug price increases, as marginal cost is no longer zero[7]:

$$\frac{\partial\gamma_S p^*_{b,S}}{\partial\gamma_S}>0 \tag{51}$$

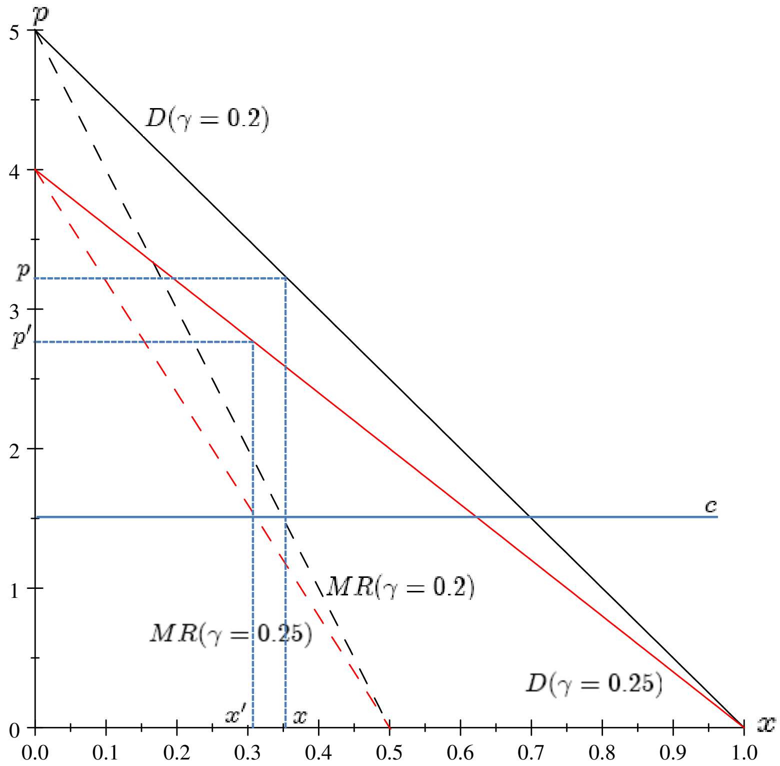
As the price decrease does not offset the effect of an increase of the copayment and thus, under coinsurance rates, the quantity consumed also decreases:

$$\frac{\partial q^*_{b,S}}{\partial\gamma_S}<0. \tag{52}$$

Graph 1 illustrates the effect of an increase of the coinsurance rate for marginal cost greater than zero. Let $D\left(\gamma=0.2\right)$ denote the demand curve for a coinsurance rate of $\gamma=0.2$ and $MR\left(\gamma=0.2\right)$ the corresponding marginal revenue curve. Similarly, let $D\left(\gamma=0.25\right)$ and $MR\left(\gamma=0.25\right)$ denote the demand curve and marginal revenue curve respectively for a coinsurance rate of $\gamma=0.25$. An increase of the coinsurance rate from $\gamma=0.2$ to $\gamma=0.25$ increases price elasticity of demand for all positive prices and quantities (inward turn of the demand curve) and makes the manufacturer lower the price from p to p'. As marginal cost is not zero, the price decrease cannot compensate the effect from higher price elasticity and the quantity sold decreases. The intersection of marginal cost and marginal revenue does not coincide with the x-axis, as marginal cost is greater than zero. Thus, the dimension of the intersection of marginal cost and marginal revenue depends on the coinsurance rate, i.e. the price elasticity of demand, as graph 1 shows. In other words, as the effective consumer price increases with the coinsurance rate, the quantity

[7]Note that $\frac{\partial\left(\gamma_F\frac{1+w^*\gamma_F}{2\gamma_F}\right)}{\partial\gamma_F}=\frac{1}{2}w^*$. That is, if $w=0$, the effective consumer price is independent of the coinsurance rate; if $w>0$, an increase of the coinsurance rate implies an increase of the effective consumer price.

consumed decreases.



Graph 1: Increase of coinsurance rate, $c > 0$.

Thus, similar to the effect of an increase of the copayment in country D on drug prices and quantities in country D, the increase of the copayment in country S results in a lower drug price and a lower quantity sold, which decreases health expenditure:

$$\frac{\partial E_S^*}{\partial \gamma_S} < 0. \tag{53}$$

Spillovers of copayment changes in country S to the destination country D again occur via the wholesale price. Since the effective drug price increases in the wholesale price and accordingly, the quantity sold decreases in the wholesale price, a higher wholesale price aggravates the double marginalization effect. Consequently, the manufacturer reduces the wholesale price:

$$\frac{\partial w^*}{\partial \gamma_S} < 0. \tag{54}$$

In country D, the price of the parallel import is a mark-up over the intermediary's marginal cost, which is the wholesale price w^*: $p_{\beta,D}^* = \frac{1}{2}\left(w^* + p_{b,D}^*\left(1-\tau\right)\right)$. Consequently, a decrease of the wholesale price results in a lower price for the parallel import. This induces the manufacturer to reduce also the price for the locally sourced version of the drug in order not to lose too many consumers to the parallel import, as illustrated by the best response function: $p^*_{b,D} = \frac{1}{2}\left(\tau + p^*_{\beta,D} + w^*\right)$. Accordingly, in country D, both drug prices decrease in the coinsurance rate in the source country:

$$\frac{\partial p^*_{b,D}}{\partial \gamma_S} < 0, \; \frac{\partial p^*_{\beta,D}}{\partial \gamma_S} < 0. \tag{55}$$

Effective drug prices decrease, as the coinsurance rate in destination country D is unchanged

$$\frac{\partial \gamma_D p^*_{b,D}}{\partial \gamma_S} < 0, \; \frac{\partial \gamma_D p^*_{\beta,D}}{\partial \gamma_S} < 0. \tag{56}$$

A drug price decrease and an unchanged coinsurance rate increases the quantity sold:

$$\frac{\partial q^*_{b,D}}{\partial \gamma_S} > 0, \; \frac{\partial q^*_{\beta,D}}{\partial \gamma_S} > 0. \tag{57}$$

As the effect of lower prices more than offsets the effect of a higher quantity, public health expenditure decreases (see Appendix D):

$$\frac{\partial E^*_D}{\partial \gamma_S} < 0. \tag{58}$$

Thus, an increase of the copayment in the source country S increases the extent and accordingly the importance of the double marginalization effect, which results in a decrease of the wholesale price. This reduces marginal cost for the intermediary, which translates to a price reduction for the parallel import and then, as prices are strategic complements, also to a price reduction for the locally sourced version of the drug. By reducing drug prices and increasing the quantities sold, a copayment increase in the source country reinforces the effect of competition by parallel trade in country D.

Proposition 2 summarizes the effect of an increase in the coinsurance rate in the source country S:

Proposition 2 *An increase in the coinsurance rate in the source country S i) raises effective consumer prices, lowers the quantity consumed and reduces health expenditure in country S, ii) lowers effective consumer prices, increases the quantity consumed and lowers health expenditure in country D.*

5 Implications for Health Policy

This section investigates the implications of the externalities of national decisions for health policy.

When markets are integrated through parallel trade and pricing decisions are interdependent, national decisions on coinsurance rates result in spillovers to the respective other country. By changing prices and volume, national decisions also have an effect on (potential) objectives of health policy[8], namely public pharmaceutical expenditure and consumer surplus.

A change of the coinsurance rate in the destination country D results in a mitigation of the double marginalization effect in the source country S. By lowering the drug price and increasing the quantity consumed, this increases consumer surplus and increases public pharmaceutical expenditure in the source country. A change of the coinsurance rate in the source country S reinforces the effect of competition by parallel trade in the destination country D, where drug prices are reduced and the quantity consumed is increased. This increases consumer surplus and reduces public pharmaceutical expenditure in the destination country.

Given that coinsurance rates are the result of a political optimization, balancing different health policy objectives, a change of the coinsurance rate in one country might induce a change of the coinsurance rate in the respective other country. Consequently, there might be an incentive for one government to modify the coinsurance rate, following a change of the coinsurance rate by the other government.

5.1 Non-coordinated Health Policy

Consider first the case of governments setting coinsurance rates in a non-coordinated way and without taking the externalities for the respective other country into account. Assume that countries set coinsurance rates to maximize local welfare.

In the destination country D, welfare is given as the sum of consumer surplus and the manufacturer's profit[9] net of public pharmaceutical expenditure:

$$W_D^* = CS_D^* + \pi_M^* - E_D^*, \tag{59}$$

in the source country S, welfare is consumer surplus net of public pharmaceutical expenditure:

$$W_S^* = CS_S^* - E_S^*. \tag{60}$$

Maximizing W_D^* with respect to γ_D and W_S^* with respect to γ_S yields the best response functions $\widehat{\gamma_D}(\gamma_S)$ and $\widehat{\gamma_S}(\gamma_D)$, see Appendix E. Numerical simulations yield no equilibrium for $\gamma_D \in (0,1)$ and $\gamma_S \in (0,1)$. This implies that a modification of the coinsurance rate in one country triggers a change in the respective other country as well. In addition, coinsurance rates are inefficient from a global perspective: Without taking the externalities for the source country into account, the country D-regulatory body sets a coinsurance rate not sufficiently high with respect to consumer surplus in the source country and it chooses a rate not sufficiently low with respect to public pharmaceutical expenditure in the source country. If the country S-regulatory body sets the

[8] As mentioned in the introduction, the reduction of public expenditure or distributive objectives, e.g. minimization of financial exposure of patients and guaranteeing broad access to pharmaceuticals, may enter the objective function.

[9] Note that the intermediary's profit is extracted via the fixed fee. That is, it is included in the manufacturer's profit.

coinsurance rate without considering the externality to the destination country D, it chooses a rate not sufficiently high. From a consumer surplus perspective, a coordination of pharmaceutical policy would imply higher coinsurance rates as compared to national pharmaceutical policy. From a public pharmaceutical expenditure perspective, the coordination of pharmaceutical policy would imply a lower coinsurance rate in the destination country D and a higher coinsurance rate in the source country S as compared to national pharmaceutical policy.

5.2 Coordinated Health Policy

Assume that governments in both countries set coinsurance rates to maximize total welfare. Total welfare, i.e. $W^* = W_D^* + W_S^*$, strictly increases in γ_D and decreases in γ_S, see Appendix E. Thus, total welfare is maximized for

$$\gamma_D = 1, \ \gamma_S = 0. \tag{61}$$

This implies that there is no reimbursement in the destination country and patients pay the full drug price out-of-pocket. This reduces public pharmaceutical expenditure, but it also reduces consumer surplus by increasing financial exposure and reducing access to pharmaceuticals. In the source country, patients pay only a very small fraction of the market price, which increases consumer surplus, but it also increases public pharmaceutical expenditure. This illustrates that the coordination of pharmaceutical policy does not imply identical coinsurance rates in both countries, i.e. coordination of pharmaceutical policy does not result in harmonization of coinsurance rates. Also, this implies that the conflict between different health policy objectives – reduction of public health expenditure and distributive objectives – remains and cannot be resolved through the coordination of pharmaceutical policy.

Starting from these coordinated coinsurance rates, regulatory bodies can improve local welfare by changing the coinsurance rate, $\widehat{\gamma_D}\,(\gamma_S = 0) \neq 1$ and $\widehat{\gamma_S}\,(\gamma_D = 1) \neq 0$, see Appendix E. This is, there is the incentive for regulatory bodies to deviate from the coordinated coinsurance rates and modify the coinsurance rate.

6 Conclusion

In this paper, I have studied the externalities of national decisions on health policy, more precisely changes in coinsurance rates.

Parallel trade generates a competition effect in the destination country, resulting in lower drug prices and a higher quantity sold. The higher wholesale price (as compared to segmented markets) creates a double-marginalization effect with a higher drug price and a lower quantity sold. Parallel trade results in market integration, as it makes pricing decisions with respect to the different markets interdependent. In this setting, national decisions on coinsurance rates affect the trade-off between the double marginalization effect and the competition effect. By changes in the wholesale price, externalities occur. An increase of the coinsurance rate in the destination country mitigates the double marginalization effect in the source country; an increase of the coinsurance rate in the source country reinforces the competition effect from parallel trade in the destination country.

The interdependence of pharmaceutical policy under parallel trade may pose a number of problems and questions which need to be addressed.

First, these externalities may generate a frequent and ongoing adjustment of coinsurance rates. For instance, if the destination country increases the coinsurance rate, this increases public pharmaceutical expenditure in the source country, which then may trigger a coinsurance rate increase by the respective regulator in the source country as well. Second, coinsurance rate decreases may have adverse effects: Coinsurance rate decreases in the destination country aggravate the double marginalization effect in the source country. Coinsurance rate decreases in the source country weaken the competition effect in the destination country. Third, the change of the coinsurance rate in the destination country induces a conflict between the health policy objectives of reduction of public pharmaceutical expenditure and distributive objectives in the source country: A decrease of the coinsurance rate in the destination country reduces public pharmaceutical expenditure in the source country, but increases financial exposure of patients and worsens access to pharmaceuticals at the same time.

Consequently, this may present a case for policy coordination in the European Union. In this model, there might be an incentive for one government to modify the coinsurance rate, following a change of the coinsurance rate by the other government. Thus, a multilateral agreement on pharmaceutical policy may be desirable. However, it must be taken into account that this may aggravate the conflict between the reduction of public pharmaceutical expenditure and distributive objectives in health policy. In addition, given that EU countries differ in income per capita, financing of health insurance, culture etc. this may pose additional problems.

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] Bordoy, C. & Jelovac, I. (2005): Pricing and welfare implications of parallel imports in the pharmaceutical industry. International Journal of Health Care Finance and Economics 5, 5 - 21.

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

[4] Desogus, C. (2011): Competition and innovation in the EU regulation of pharmaceuticals: the case of parallel trade. Cambridge.

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

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

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

[8] Granlund, D. & Köksal, M. Y. (2011): EU enlargement, parallel trade and price competition in pharmaceuticals: what's to blame? ; derogation or perception?. Göteborg University Working Paper No. 499, http://gupea.ub.gu.se/bitstream/2077/25500/1/gupea_2077_25500_1.pdf.

[9] Hancher, L. (2004): The European Community dimension: coordinating divergence, in: Mossialos, E.; Mrazek, M. & Walley, T.(eds.): Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality, Maidenhead, 55-79.

[10] 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.

[11] Kyle, M. K. (2009): Parallel Trade in Pharmaceuticals: Firm Responses and Competition Policy, in: Hawk, B. E. (ed): International Antitrust Law & Policy, ?, 339-358.

[12] Kyle, M. K., Allsbrook J. S. & Schulman K. A. (2008): Does Reimportation Reduce Price Differences for Prescription Drugs? Lessons from the European Union. Health Service Research 43, 1308-1324.

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

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

[15] NERA (1999): The economic consequences of the choice of a regime of exhaustion in the area of trademarks. Final report for DGXV of the European Commission.

[16] Pecorino, P. (2002): "Should the US allow prescription drug reimports from Canada?. Journal of Health Economics, Elsevier 21, 699-708.

[17] Raimondos-Møller, P. & Schmitt, N. (2010): Commodity taxation and parallel imports. Journal of public economics 94, 153-162.

[18] Rey, P. (2003): The Impact of Parallel Imports on Prescription Medicines. mimeo, University of Toulouse.

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

[20] Timur, A.; Picone, G. A. & DeSimone, J. (2010): Has the European Union achieved a single pharmaceutical market?. NBER Working Paper No.16261, http://www.nber.org/papers/w16261.pdf.

[21] Valletti, T. M. (2006): Parallel trade, international exhaustion and intellectual property rights: a welfare analysis. The Journal of Industrial Economics 54, 499-526

[22] 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.

[23] West, J. & Mahon, P. (2003): Benefits to Payers and Patients from Parallel Trade. York Health Economics Consortium, University of York.

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

Appendix A: The Effect of Parallel Trade

Competition Effect in the Destination Country

In country D, parallel trade induces a competition effect with lower drug prices and a higher quantity sold.

Compared to segmented markets, competition from parallel trade reduces the price of the drug sold directly by the manufacturer:

$$\frac{p^*_{b,D}}{p_{b,D}} = \frac{4\gamma_D\left(9-5\tau\right)+4\tau\gamma_S\left(1-\tau\right)\left(3+\tau\right)}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2} < 1, \tag{62}$$

with the price of the parallel import being lower than the price of the locally sourced version:

$$\frac{p^*_{\beta,D}}{p^*_{b,D}} = \left(1-\tau\right)\frac{2\gamma_D\left(9-5\tau\right)+\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)}{2\gamma_D\left(9-5\tau\right)+2\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)} < 1. \tag{63}$$

The difference between the prices of the two versions of the drug stems from (perceived) vertical product differentiation: The intermediary has to compensate consumers for the lower (perceived) quality by pricing at a certain discount from a given price of the locally sourced drug version[10].

Accordingly, under parallel trade, the prices of both versions of the drug are lower than the monopoly drug price under segmented markets.

The quantity of the locally sourced version is higher under parallel trade:

$$\frac{q^*_{b,D}}{q_{b,D}} = \frac{4\gamma_D\left(9-5\tau\right)+4\gamma_S\left(1-\tau\right)\left(3+\tau\right)}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2} > 1. \tag{64}$$

Consequently, the total quantity of the drug available, that is, the quantity of the locally sourced version plus the parallel import, is higher than the monopoly quantity under segmented markets.

Double Marginalization Effect in the Source Country

In country S, parallel trade generates a double marginalization effect with a higher drug price and a lower quantity due to an increase of the wholesale price.

Compared to segmented markets, the wholesale price w^* is higher under parallel trade. As a low wholesale price induces more parallel trade and consequently enhances the competition from parallel trade in the destination country D, the manufacturer raises the wholesale price in order to deter parallel trade partially:

$$w^* > w = 0. \tag{65}$$

The increase of the wholesale price induced by parallel trade translates to an increase of the drug

[10]Note that the intermediary's best response function is $p^*_{\beta,H} = \frac{1}{2}\left(w + p^*_{b,H}\left(1-\tau\right)\right)$.

price:

$$\frac{p^*_{b,S}}{p_{b,S}} = \frac{4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\left(27-4\tau+\tau^2\right)}{\left[4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\,(3+\tau)^2\right]} > 1 \tag{66}$$

and the higher price reduces the quantity sold:

$$\frac{q^*_{b,S}}{q_{b,S}} = \frac{4\gamma_D\,(9-5\tau)-\gamma_S\,(1-\tau)\left(9-16\tau-\tau^2\right)}{\left[4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\,(3+\tau)^2\right]} < 1. \tag{67}$$

Price Convergence vs. Divergence

Parallel trade results in price convergence, if it goes from the ex-ante low price country to the ex-ante high price country (i.e. if the pre-parallel trade drug price in the source country S is lower than the pre-parallel trade price in the destination country D[11]):

$$\begin{aligned}\frac{p^*_{b,D}-p^*_{b,S}}{p_{b,D}-p_{b,S}} &= \\ &\frac{(\gamma_S-\gamma_D)\left[4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\,(3+\tau)^2\right]}{(\gamma_S-\gamma_D)\left[4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\,(3+\tau)^2\right]} \\ &\qquad -\frac{\gamma_S\,(1-\tau)\left[2\gamma_D\,(9-5\tau)+3\gamma_S\,(\tau+3)\,(1-\tau)\right]}{(\gamma_S-\gamma_D)\left[4\gamma_D\,(9-5\tau)+\gamma_S\,(1-\tau)\,(3+\tau)^2\right]} < 1 \\ &\text{if } \gamma_S > \gamma_D\end{aligned} \tag{68}$$

Appendix B: Change of coinsurance rates under segmented markets

Change in the destination country

Consider a change of the coinsurance rate in the destination country and its implications for the source country.

An increase of the coinsurance rate in the destination country D decreases the demand for the drug, as price elasticity increases:

$$\frac{\partial q_{b,D}}{\partial \gamma_D} = -p_{b,D} < 0. \tag{69}$$

As a result, the manufacturer lowers the drug price:

$$\frac{\partial p_{b,D}}{\partial \gamma_D} = -\frac{1}{2\gamma_D^2} < 0, \tag{70}$$

[11] Note that under segmented markets, $p_{b,H} > p_{b,F}$, i.e. country H is the high price country and country F is the low price country, if $\gamma_F > \gamma_H$.

leaving the effective consumer price unchanged

$$\frac{\partial \gamma_D p_{b,D}}{\partial \gamma_D} = 0^{12}. \tag{71}$$

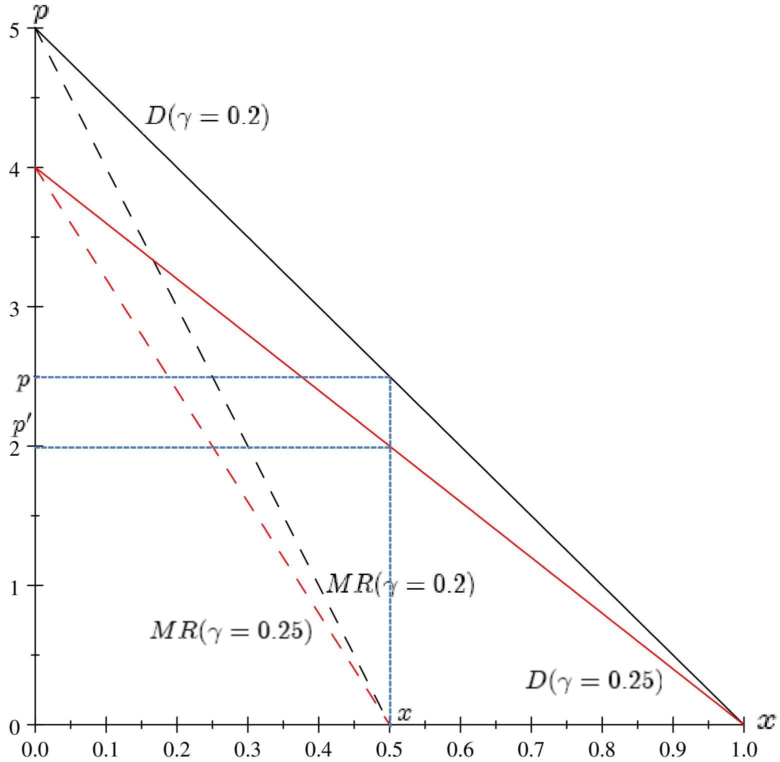
Consequently, the quantity consumed is unchanged

$$\frac{\partial q_{b,D}}{\partial \gamma_D} = 0.$$

Graph 2 illustrates the effect of an increase of the coinsurance rate under segmented markets, i.e. monopoly, and for marginal cost of zero. Let $D(\gamma = 0.2)$ denote the demand curve for a coinsurance rate of $\gamma = 0.2$ and $MR(\gamma = 0.2)$ the corresponding marginal revenue curve. Similarly, let $D(\gamma = 0.25)$ and $MR(\gamma = 0.25)$ denote the demand curve and marginal revenue curve respectively for a coinsurance rate of $\gamma = 0.25$. An increase of the coinsurance rate from $\gamma = 0.2$ to $\gamma = 0.25$ increases price elasticity of demand (inward turn of the demand curve) and makes the manufacturer lower the price from p to p'. This compensates the increase in the coinsurance rate completely and quantity consumed remains unchanged. Marginal cost of zero implies that the manufacturer sells a quantity up to a marginal revenue of zero. This corresponds to the intersection of marginal revenue curve and the x-axis, which is independent of changes in the coinsurance rate.[13] In other words, as the effective consumer price is independent of the coinsurance rate, so is the quantity consumed.

[12] Note that the effective consumer price $\gamma_H p_{b,H} = \frac{1}{2}$ is independent of the coinsurance rate.

[13] As an increase of the coinsurance turns the demand curve and does not affect the quantity demanded at a price of zero (intersection of the x-axis and the demand curve), also the intersection of the marginal revenue curve and the x-axis (marginal cost of zero) remains unchanged.



Graph 2: Increase of the coinsurance rate, $c = 0$.

A lower drug price at an unchanged quantity consumed reduces public health expenditure:

$$\frac{\partial E_D}{\partial \gamma_D} = -\frac{1}{4\gamma_D^2} < 0. \tag{72}$$

As the manufacturer's pricing decisions are independent under segmented markets, the drug price and the quantity consumed in the source country S are independent of (changes of) the coinsurance rate in the destination country D:

$$\frac{\partial p_{b,S}}{\partial \gamma_D} = 0, \; \frac{\partial q_{b,S}}{\partial \gamma_D} = 0. \tag{73}$$

In other words, there are no spillovers of changes in the destination country to the source country.

Change in the source country

Consider now a change of the coinsurance rate in the source country and its implications for the destination country.

Similarly, an increase of the coinsurance rate in the destination country S increases price elasticity of demand and thus decreases demand for the drug:

$$\frac{\partial q_{b,S}}{\partial \gamma_S} = -p_{b,S} < 0. \tag{74}$$

Consequently, the intermediary lowers the drug price:

$$\frac{\partial p_{b,S}}{\partial \gamma_S} = -\frac{1}{2\gamma_S^2} < 0, \tag{75}$$

leaving the effective consumer price unchanged:

$$\frac{\partial \gamma_S p_{b,S}}{\partial \gamma_S} = 0^{14}. \tag{76}$$

Also the quantity consumed is unchanged:

$$\frac{\partial q_{b,S}}{\partial \gamma_S} = 0. \tag{77}$$

A lower drug price reduces public health expenditure:

$$\frac{\partial E_S}{\partial \gamma_S} = -\frac{1}{4\gamma_S^2} < 0. \tag{78}$$

As the manufacturer's pricing decisions are independent under segmented markets, the drug price and the quantity consumed in the destination country D are independent of (changes of) the coinsurance rate in the source country S:

$$\frac{\partial p_{b,S}}{\partial \gamma_D} = 0, \ \frac{\partial q_{b,S}}{\partial \gamma_D} = 0. \tag{79}$$

In other words, there are no spillovers of changes in the source country to the destination country.

To summarize, without parallel trade, an increase in the coinsurance rate in either country has no effect on effective consumer prices and the quantity consumed, but reduces health expenditure, and has no effect on consumers or health expenditure in the other country.

Appendix C: Change of Coinsurance Rates under Parallel Trade

Change in the destination country

An increase in the coinsurance rate in the destination country D raises effective consumer prices, lowers the quantity consumed, and reduces health expenditure in the destination country D and

[14] Note that the effective consumer price $\gamma_H p_{b,H} = \frac{1}{2}$ is independent of the coinsurance rate.

lowers effective consumer prices, increases the quantity consumed, and raises health expenditure in the source country S.

In the destination country, the increase of the coinsurance rate results in lower drug prices and lower quantities sold.

An increase of the coinsurance rate in country D decreases demand for the locally sourced version of the drug c.p.:

$$\frac{\partial q_{b,D}^*}{\partial \gamma_D} = -\frac{\left(p_{b,D}^* - p_{\beta,D}^*\right)}{\tau} < 0. \tag{80}$$

For the parallel import, demand increases, if the price difference between the locally sourced version and the parallel import exceeds the quality difference:

$$\begin{aligned} \frac{\partial q_{\beta,D}^*}{\partial \gamma_D} &= \frac{(1-\tau)\, p_{b,D}^* - p_{\beta,D}^*}{\tau\,(1-\tau)} > 0, \\ \text{if } p_{\beta,D}^* &< (1-\tau)\, p_{b,D}^*. \end{aligned} \tag{81}$$

The direct effect of the price for the locally sourced version on the price for the parallel import, however, leads to a decrease of the price for the parallel import as well[15]. This is demonstrated by the best response function: $p_{\beta,D}^* = \frac{1}{2}\left(w + p_{b,D}^*\,(1-\tau)\right)$.

Accordingly, in country D both drug prices decrease in the coinsurance rate:

$$\begin{aligned} \frac{\partial p_{b,D}^*}{\partial \gamma_D} &= -\frac{2[4\gamma_D^2\,(9-5\tau)^2 + 8\tau\gamma_S\gamma_D\,(1-\tau)\,(\tau+3)\,(9-5\tau) + \tau\gamma_S^2\,(1-\tau)^2\,(\tau+3)^3]}{\gamma_D^2[4\gamma_D\,(9-5\tau) + \gamma_S\,(1-\tau)\,(3+\tau)^2]^2} < 0, \\ \frac{\partial p_{\beta,D}^*}{\partial \gamma_D} &= -\frac{(1-\tau)\,[8\gamma_D^2\,(9-5\tau)^2 + 8\tau\gamma_S\gamma_D\,(1-\tau)\,(\tau+3)\,(9-5\tau) + \tau\gamma_S^2\,(1-\tau)^2\,(\tau+3)^3]}{\gamma_D^2[4\gamma_D\,(9-5\tau) + \gamma_S\,(1-\tau)\,(3+\tau)^2]^2} \end{aligned} \tag{82}$$

Competition and higher price elasticity under parallel trade of demand limits the ability to increase prices in response to an increase of the coinsurance rate and consequently, effective consumer prices increase:

$$\begin{aligned} \frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_D} &= \frac{6\gamma_S\,(1-\tau)^2\,(9-5\tau)\,(3+\tau)}{[4\gamma_D\,(9-5\tau) + \gamma_S\,(1-\tau)\,(3+\tau)^2]^2} > 0, \\ \frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_D} &= \frac{2\gamma_S\,(1-\tau)^2\,(3-\tau)\,(\tau+3)\,(9-5\tau)}{[4\gamma_D\,(9-5\tau) + \gamma_S\,(1-\tau)\,(3+\tau)^2]^2} > 0. \end{aligned} \tag{83}$$

As price decreases cannot compensate the effect of lower demand, quantities of both versions

[15] In addition, 39 implies that the intermediary has to lower $p_{\beta,H}^*$ in order to prevent a decrease of demand.

of the drug decrease in γ_D:

$$\begin{aligned}
\frac{\partial q^*_{b,D}}{\partial \gamma_D} &= -\frac{2\gamma_S (1-\tau)^2 \left(27 - 6\tau - 5\tau^2\right)}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0, \\
\frac{\partial q^*_{\beta,D}}{\partial \gamma_D} &= -\frac{\gamma_S (1-\tau)(\tau+3)\, 4 (9-5\tau)}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (84)
\end{aligned}$$

Lower prices and lower quantities consumed reduce the public health expenditure:

$$\begin{aligned}
\frac{\partial E^*_D}{\partial \gamma_D} = & -\frac{16\gamma_D^3 (9-5\tau)^3 + \tau\gamma_S^3 (5-\tau)(1-\tau)^3 (\tau+3)^4}{\gamma_D^2 \left(4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2\right)^3} \\
& - \frac{2\gamma_S\gamma_D (1-\tau)(9-5\tau)(\tau+3)^2 \left(6\gamma_D (9-5\tau) + \gamma_S\gamma_D (9-5\tau)(1-\tau)^2\right)}{\gamma_D^2 \left(4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2\right)^3} \\
& - \frac{2\gamma_S\gamma_D (1-\tau)(9-5\tau)(\tau+3)^2 \left(6\tau\gamma_S (1-\tau)(5-\tau)\right)}{\gamma_D^2 \left(4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2\right)^3} < 0. \qquad (85)
\end{aligned}$$

Spillovers of copayment changes in country D to the source country S occur via the wholesale price. As the decrease of total demand in country D reduces the effect of competition, the double marginalization effect gains relative importance and, accordingly, the manufacturer lowers the wholesale price:

$$\frac{\partial w^*}{\partial \gamma_D} = -\frac{8 (9-5\tau)^2 (1-\tau)}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (86)$$

A decrease of the wholesale price then results in drug price decrease in country S:

$$\frac{\partial p^*_{b,S}}{\partial \gamma_D} = -\frac{4 (1-\tau)(9-5\tau)^2}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (87)$$

As the coinsurance rate in the source country S is unchanged, the effective drug price decreases:

$$\frac{\partial \gamma_S p^*_{b,S}}{\partial \gamma_D} = -\frac{4\gamma_S (1-\tau)(9-5\tau)^2}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0, \qquad (88)$$

which increases the quantity consumed:

$$\frac{\partial q^*_{b,S}}{\partial \gamma_D} = \frac{4\gamma_S (1-\tau)(9-5\tau)^2}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} > 0. \qquad (89)$$

Health expenditure increases, as the effect from a higher quantity consumed exceeds the effect of a lower drug price (see Appendix C):

$$\frac{\partial E_S^*}{\partial \gamma_D} = \frac{8\gamma_S (1-\tau)^2 (1-\gamma_S)(9-5\tau)^3}{\left(4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2\right)^3} > 0. \tag{90}$$

Change in the source country

Consider now a change of the cost-sharing instrument in the source country and its implications for the destination country.

An increase in the coinsurance rate in the source country S raises effective consumer prices, lowers the quantity consumed and reduces health expenditure in the source country S and lowers effective consumer prices, increases the quantity consumed and lowers health expenditure in the destination country country D.

In the source country, the increase of copayments, i.e. an increase of the coinsurance rate results, similarly to the effects in the destination country, in lower drug prices and lower quantities sold.

Demand for the drug decreases c.p.:

$$\frac{\partial \left(1 - \gamma_S p_{b,S}^*\right)}{\partial \gamma_S} = -\gamma_S p_{b,S}^* < 0. \tag{91}$$

Accordingly, the drug price decreases in γ_S:

$$\frac{\partial p_{b,S}^*}{\partial \gamma_S} = -\frac{16\gamma_D^2 (9-5\tau)^2 + \gamma_S (1-\tau)(\tau+3)^2 [8\gamma_D (9-5\tau) + \gamma_S (1-\tau)\left(27 - 4\tau + \tau^2\right)]}{2\gamma_S^2 [4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \tag{92}$$

The effective drug price increases, as marginal cost is no longer zero[16]:

$$\frac{\partial \gamma_S p_{b,S}^*}{\partial \gamma_S} = \frac{4\gamma_D (1-\tau)(9-5\tau)^2}{\left(4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2\right)^2} > 0 \tag{93}$$

As the price decrease does not offset the effect of an increase of the copayment and thus, under coinsurance rates, the quantity consumed also decreases:

$$\frac{\partial q_{b,S}^*}{\partial \gamma_S} = -\frac{4\gamma_D (1-\tau)(5\tau-9)^2}{[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \tag{94}$$

Similar to the effect of an increase of the copayment in country D on drug prices and quantities

[16] Note that $\frac{\partial\left(\gamma_F \frac{1+w^*\gamma_F}{2\gamma_F}\right)}{\partial \gamma_F} = \frac{1}{2}w^*$. That is, if $w = 0$, the effective consumer price is independent of the coinsurance rate; if $w > 0$, an increase of the coinsurance rate implies an increase of the effective consumer price.

in country D, the increase of the copayment in country S results in a lower drug price and a lower quantity sold, which decreases health expenditure:

$$\begin{aligned}
\frac{\partial E_S^*}{\partial \gamma_S} &= -\frac{64\gamma_D^3 (9-5\tau)^3 - \gamma_S^3 \left(9-16\tau-\tau^2\right)\left(27-4\tau+\tau^2\right)(\tau+3)^2 (1-\tau)^3}{4\gamma_S^2 \left(4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(\tau+3)^2\right)^3} \\
&\quad - \frac{4\gamma_S\gamma_D (1-\tau)(9-5\tau)\left(\gamma_S (1-\tau)\left(567-36\tau+262\tau^2+36\tau^3+3\tau^4-8\gamma_S (5\tau-9)^2\right)\right)}{4\gamma_S^2 \left(4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(\tau+3)^2\right)^3} \\
&\quad - \frac{4\gamma_S\gamma_D (1-\tau)(9-5\tau)\left(12\gamma_D (9-5\tau)(\tau+3)^2\right)}{4\gamma_S^2 \left(4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(\tau+3)^2\right)^3} < 0. \qquad (95)
\end{aligned}$$

Spillovers of copayment changes in country S to the destination country D again occur via the wholesale price.

As the quantity reduction increases in the wholesale price, the manufacturer reduces the wholesale price:

$$\frac{\partial w^*}{\partial \gamma_S} = -\frac{2(1-\tau)^2 (3+\tau)^2 (9-5\tau)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (96)$$

In country D, the decrease of the wholesale price results in a lower price for the parallel import. This induces the manufacturer to reduce also the price for the locally sourced version of the drug in order not to lose too many consumers to the parallel import. Accordingly, in country D, both drug prices decrease in the coinsurance rate in the source country:

$$\begin{aligned}
\frac{\partial p_{b,D}^*}{\partial \gamma_S} &= -\frac{6(1-\tau)^2 \left(27-6\tau-5\tau^2\right)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} < 0, \\
\frac{\partial p_{\beta,D}^*}{\partial \gamma_S} &= -\frac{2(1-\tau)^2 (3-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (97)
\end{aligned}$$

Effective drug prices decrease, as the coinsurance rate in destination country D is unchanged

$$\begin{aligned}
\frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_S} &= -\frac{6\gamma_D (1-\tau)^2 (3+\tau)(9-5\tau)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} < 0, \\
\frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_S} &= -\frac{2\gamma_D (1-\tau)^2 (3-\tau)(\tau+3)(9-5\tau)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} < 0. \qquad (98)
\end{aligned}$$

A drug price decrease and an unchanged coinsurance rate increases the quantity sold:

$$\begin{aligned}
\frac{\partial q_{b,D}^*}{\partial \gamma_S} &= \frac{2\gamma_D (1-\tau)^2 \left(27-6\tau-5\tau^2\right)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} > 0, \\
\frac{\partial q_{\beta,D}^*}{\partial \gamma_S} &= \frac{4\gamma_D (1-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D (9-5\tau)+\gamma_S (1-\tau)(3+\tau)^2]^2} > 0. \qquad (99)
\end{aligned}$$

As the effect of lower prices more than offsets the effect of a higher quantity, public health expenditure decreases:

$$\frac{\partial E_D^*}{\partial \gamma_S} = -\frac{2\gamma_S \left(1-\tau\right)^3 \left(1-\gamma_D\right) \left(27-6\tau-5\tau^2\right)^2}{[4\gamma_D \left(9-5\tau\right) + \gamma_S \left(1-\tau\right) \left(\tau+3\right)^2]^3}. \tag{100}$$

Appendix D: Change in Health Expenditure

Under segmented markets, public drug expenditure in the destination country is given as

$$E_D \left(\gamma\right) = \left(1-\gamma_D\right) p_{b,D} q_{b,D} = \frac{\left(1-\gamma_D\right)}{4\gamma_D}. \tag{101}$$

and public health expenditure in the source country is given as:

$$E_S \left(\gamma\right) = \left(1-\gamma_S\right) p_{b,S} q_{b,S} = \frac{\left(1-\gamma_S\right)}{4\gamma_S}. \tag{102}$$

Under parallel trade, public drug expenditure in the destination country amounts to

$$\begin{aligned} E_D^* &= \left(1-\gamma_D\right) \left(p_{b,D}^* q_{b,D}^* + p_{\beta,D}^* q_{\beta,D}^*\right) \\ &= \left(1-\gamma_D\right) \left(\begin{array}{c} \frac{4\gamma_D^2 (9-5\tau)^2 + 2\gamma_S \gamma_D (1-\tau)(9-5\tau)(\tau+3)^2}{\gamma_D [4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2} \\ + \frac{\tau \gamma_S^2 (5-\tau)(3+\tau)^2 (1-\tau)^2}{\gamma_D [4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2} \end{array}\right). \end{aligned} \tag{103}$$

and public health expenditure in the source country is given as:

$$\begin{aligned} E_S^* &= \left(1-\gamma_S\right) p_{b,S}^* q_{b,S}^* \\ &= \left(1-\gamma_S\right) \left(\begin{array}{c} \frac{16\gamma_D^2 (9-5\tau)^2 + 8\gamma_S \gamma_D (1-\tau)(9-5\tau)(3+\tau)^2}{4\gamma_S [4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2} \\ - \frac{\gamma_S^2 (1-\tau)^2 \left(27-4\tau+\tau^2\right)\left(9-16\tau-\tau^2\right)}{4\gamma_S [4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2} \end{array}\right). \end{aligned} \tag{104}$$

Increase of health expenditure in the source country following from an increase of the coinsurance rate in the destination country

$$\frac{\partial\left(E_S^*(\gamma)\right)}{\partial\gamma_D}=\frac{\partial\left(1-\gamma_S\right)p_{b,S}^*q_{b,S}^*}{\partial\gamma_D}=\left(1-\gamma_S\right)\left(\frac{\partial p_{b,S}^*}{\partial\gamma_D}q_{b,S}^*+p_{b,S}^*\frac{\partial q_{b,S}^*}{\partial\gamma D}\right)>0,$$
$$\text{since }\underbrace{\frac{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(27-4\tau+\tau^2\right)}{2\gamma_S[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}}_{p_{b,S}^*}\underbrace{\frac{4\gamma_S\left(1-\tau\right)\left(9-5\tau\right)^2}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{\frac{\partial q_{b,S}^*}{\partial\gamma_D}}$$
$$>\underbrace{\frac{4\gamma_D\left(9-5\tau\right)-\gamma_S\left(1-\tau\right)\left(9-16\tau-\tau^2\right)}{2[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}}_{q_{b,S}^*}\underbrace{\frac{4\left(1-\tau\right)\left(9-5\tau\right)^2}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{\frac{\partial p_{b,S}^*}{\partial\gamma_D}}\tag{105}$$

Decrease of health expenditure in the destination country following from an increase of the coinsurance rate in the source country

$$\frac{\partial\left(E_D^*(\gamma)\right)}{\partial\gamma_S}=\frac{\partial\left(\left(1-\gamma_D\right)\left(p_{b,D}^*q_{b,D}^*+p_{\beta,D}^*q_{\beta,D}^*\right)\right)}{\partial\gamma_S}$$
$$=\left(1-\gamma_D\right)\left(\frac{\partial p_{b,D}^*}{\partial\gamma_S}q_{b,D}^*+p_{b,D}^*\frac{\partial q_{b,D}^*}{\partial\gamma_S}+\frac{\partial p_{\beta,D}^*}{\partial\gamma_S}q_{\beta,D}^*+p_{\beta,D}^*\frac{\partial q_{\beta,D}^*}{\partial\gamma_S}\right)<0,$$
$$\text{since }\underbrace{\frac{2\gamma_D\left(9-5\tau\right)+2\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)}{\gamma_D[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}}_{p_{b,D}^*}\underbrace{\frac{2\gamma_D\left(1-\tau\right)^2\left(27-6\tau-5\tau^2\right)}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{\frac{\partial q_{b,D}^*}{\partial\gamma_S}}$$
$$+\underbrace{\frac{\left(1-\tau\right)[2\gamma_D\left(9-5\tau\right)+\tau\gamma_S\left(3+\tau\right)\left(1-\tau\right)]}{\gamma_D[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]}}_{p_{\beta,D}^*}\underbrace{\frac{4\gamma_D\left(1-\tau\right)\left(3+\tau\right)\left(9-5\tau\right)}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{\frac{\partial q_{\beta,D}^*}{\partial\gamma_S}}$$
$$<\underbrace{\frac{6\left(1-\tau\right)^2\left(27-6\tau-5\tau^2\right)}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{q_{b,D}^*}\underbrace{\frac{2[\gamma_D\left(9-5\tau\right)+\gamma_S\left(3+\tau\right)\left(1-\tau\right)]}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2}}_{\frac{\partial p_{b,D}^*}{\partial\gamma_S}}$$
$$+\underbrace{\frac{2\left(1-\tau\right)^2\left(3-\tau\right)\left(3+\tau\right)\left(9-5\tau\right)}{[4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2]^2}}_{q_{\beta,D}^*}\underbrace{\frac{\gamma_S\left(3+\tau\right)\left(1-\tau\right)}{4\gamma_D\left(9-5\tau\right)+\gamma_S\left(1-\tau\right)\left(3+\tau\right)^2}}_{\frac{\partial p_{\beta,D}^*}{\partial\gamma_S}}.\tag{106}$$

Appendix E: Implications for Health Policy

Total welfare in the destination country is given as:

$$W_D^*=CS_D^*+\pi_M^*-E_D^*,\tag{107}$$

with

$$\begin{aligned} CS_D^* &= \int_{\theta_D^{b,\beta}}^{1} (\theta - \gamma_D p_{b,D}) d\theta + \int_{\theta_D^{\beta,0}}^{\theta_D^{b,\beta}} (\theta (1-\tau) - \gamma_D p_{\beta,D}) d\theta \\ &= \frac{(9-5\tau)\,[4\gamma_D^2 (9-5\tau) + 4\gamma_D \gamma_S (1-\tau)\left(9-\tau^2\right) + \gamma_S^2 (3+\tau)^2 (1-\tau)^2]}{2[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2}, \end{aligned} \tag{108}$$

$$\pi_M^* = \frac{4\gamma_D^2 (9-5\tau) + \gamma_S[\gamma_D \left[(1-\tau)(3+\tau)^2 + 4(9-5\tau)\right] + 4\tau\gamma_S (1-\tau)(5-\tau)]}{4\gamma_D\gamma_S[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]}, \tag{109}$$

and

$$\begin{aligned} E_D^* &= (1-\gamma_D)\left(p_{b,D}^* q_{b,D}^* + p_{\beta,D}^* q_{\beta,D}^*\right) \\ &= \frac{\begin{array}{c}(1-\gamma_D)\,[4\gamma_D^2 (5\tau-9)^2 \\ +2\gamma_S\gamma_D (1-\tau)(9-5\tau)(\tau+3)^2 \\ +\tau\gamma_S^2 (5-\tau)(\tau+3)^2 (1-\tau)^2]\end{array}}{\gamma_D[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2}. \end{aligned}$$

Maximizing W_D^* with respect to γ_D yields

$$\widehat{\gamma_D} = \frac{2\left((9-5\tau)^2 - \tau\gamma_S (1-\tau)(\tau+3)^2\right)}{3(\tau+3)(9-5\tau)}$$

.

Total welfare in the source country is given as:

$$W_S^* = CS_S^* - E_S^*. \tag{110}$$

$$\begin{aligned} CS_S^* &= \int_{\theta_S^{\beta,0}}^{1} (\theta - \gamma_S p_{b,S}) d\theta \\ &= \frac{[4\gamma_D (9-5\tau) - \gamma_S (1-\tau)\left(9-16\tau-\tau^2\right)]^2}{8[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(3+\tau)^2]^2}, \end{aligned} \tag{111}$$

and

$$\begin{aligned} E_S^* &= (1-\gamma_S)\, p_{b,S}^* q_{b,S}^* \\ &= \frac{\begin{array}{c}(1-\gamma_S)\,[16\gamma_D^2 (9-5\tau)^2 \\ +8\gamma_S\gamma_D (1-\tau)(9-5\tau)(3+\tau)^2 \\ -\gamma_S^2 (1-\tau)^2 \left(27-4\tau+\tau^2\right)\left(9-16\tau-\tau^2\right)]\end{array}}{4\gamma_S[4\gamma_D (9-5\tau) + \gamma_S (1-\tau)(\tau+3)^2]^2}. \end{aligned} \tag{112}$$

Maximizing W_S^* with respect to γ_S yields

$$\begin{aligned}\widehat{\gamma_S} &= \sqrt[3]{-\frac{(27c-9ab+2a^3)}{54}+\sqrt{\frac{(4a^3c-a^2b^2+4b^3+27c^2-18abc)}{108}}} \\ &\quad -\frac{1}{3}a+\frac{\frac{(-3b+a^2)}{9}}{\sqrt[3]{-\frac{(27c-9ab+2a^3)}{54}+\sqrt{\frac{(4a^3c-a^2b^2+4b^3+27c^2-18abc)}{108}}}}\end{aligned}$$

with a=$\frac{4\gamma_D(1-\tau)(9-5\tau)\left(8\gamma_D(5\tau-9)^2-(1-\tau)\left(-36\tau+262\tau^2+36\tau^3+3\tau^4+567\right)\right)}{(27-4\tau+\tau^2)(1-\tau)^2\left((1-\tau)(9-16\tau-\tau^2)(\tau+3)^2+8\gamma_D(5\tau-9)^2\right)}$,
b=$\frac{48\gamma_D^2(\tau-1)(\tau+3)^2(5\tau-9)^2}{(27-4\tau+\tau^2)(1-\tau)^2\left((1-\tau)(9-16\tau-\tau^2)(\tau+3)^2+8\gamma_D(5\tau-9)^2\right)}$
and c=$\frac{64\gamma_D^3(5\tau-9)^3}{(27-4\tau+\tau^2)(1-\tau)^2\left((1-\tau)(9-16\tau-\tau^2)(\tau+3)^2+8\gamma_D(5\tau-9)^2\right)}$.

Total welfare for both countries is given as:

$$W=\frac{\left[\begin{array}{c}96\gamma_D^2(9-5\tau)^2+8\gamma_S\gamma_D(1-\tau)(9-5\tau)\left(40\tau+3\tau^2+45\right)\\ -\gamma_S^2\left(81-1044\tau-230\tau^2-28\tau^3+5\tau^4\right)(1-\tau)^2\end{array}\right]}{8\left[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2\right]^2}. \tag{113}$$

Total welfare for both countries increases in γ_D:

$$\frac{\partial W}{\partial \gamma_D} \tag{114}$$

$$= \frac{2\gamma_S(1-\tau)(9-5\tau)\left[\begin{array}{c}2\gamma_D(9-5\tau)\left(9-4\tau+3\tau^2\right)\\ +\gamma_S(1-\tau)\left(243-207\tau+41\tau^2+15\tau^3+4\tau^4\right)\end{array}\right]}{\left[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(\tau+3)^2\right]^3}>0. \tag{115}$$

Total welfare for both countries decreases in γ_S:

$$\frac{\partial W}{\partial \gamma_S} \tag{116}$$

$$= \frac{-2\gamma_D(1-\tau)(9-5\tau)\left[\begin{array}{c}2\gamma_D(9-5\tau)\left(9-4\tau+3\tau^2\right)\\ +\gamma_S(1-\tau)\left(243-207\tau+41\tau^2+15\tau^3+4\tau^4\right)\end{array}\right]}{\left[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(\tau+3)^2\right]^3}<0. \tag{117}$$

$$\widehat{\gamma_D}(\gamma_S=0)=\frac{2(9-5\tau)}{3(\tau+3)},$$

$\widehat{\gamma_D}(\gamma_S=0)>1$ if $\tau<\frac{9}{13}$.

$$\widehat{\gamma_S}\left(\gamma_D=1\right) = \sqrt[3]{-\frac{(27c-9ab+2a^3)}{54}+\sqrt{\frac{(4a^3c-a^2b^2+4b^3+27c^2-18abc)}{108}}} -\frac{1}{3}a+\frac{\frac{\left(-3b+a^2\right)}{9}}{\sqrt[3]{-\frac{(27c-9ab+2a^3)}{54}+\sqrt{\frac{(4a^3c-a^2b^2+4b^3+27c^2-18abc)}{108}}}},$$

with $a=\frac{4(1-\tau)(9-5\tau)\left(81-117\tau-98\tau^2+226\tau^3+33\tau^4+3\tau^5\right)}{(1-\tau)^2(27-4\tau+\tau^2)(729-891\tau+194\tau^2+74\tau^3+21\tau^4+\tau^5)}$, $b=-\frac{48(1-\tau)(\tau+3)^2(9-5\tau)^2}{(1-\tau)^2(27-4\tau+\tau^2)(729-891\tau+194\tau^2+74\tau^3+21\tau^4+\tau^5)}$, $c=-\frac{64(9-5\tau)^3}{(1-\tau)^2(27-4\tau+\tau^2)(729-891\tau+194\tau^2+74\tau^3+21\tau^4+\tau^5)}$. $\widehat{\gamma_S}\left(\gamma_D=1\right)>0$, if $\tau<0.974$.