

Parallel Trade of Pharmaceuticals. Conflicts in Health Policy
Objectives and Regulatory Externalities in the EU Internal
Market

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Diplom-Ökonomin Laura Birg
aus Duisburg

Kiel, 2012

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Professor Horst Raff, Ph.D.

Erstberichterstattender: Professor Horst Raff, Ph.D.

Zweitberichterstattender: Professor Dr. Till Requate

Tag der Abgabe der Arbeit: 15. Mai 2012

Tag der mündlichen Prüfung: 10. Juli 2012

Acknowledgements

The present thesis has strongly been influenced by and benefited from discussions with professors and fellow students at the University of Kiel as well as presentations at various national and international conferences, seminars, and workshops. I am very grateful to all who supported my work in that way.

In particular, I would like to thank my supervisor Professor Horst Raff, Ph.D. for his advice and support. I am also very thankful to my second supervisor Professor Dr. Till Requate for very helpful remarks.

I am thankful to Anita Hartz for organizational support. Moreover, I would like to thank my colleagues Eike Berner, Philipp Henze, and Markus Kelle, who all helped me to improve this thesis in various stages of its development.

I would like to thank Robert Birg, Charity Gomo, Nadine Heitmann, Wan-Hsin Liu, Markus Pape, Matthias Raddant, Jörn Schmidt, and Jan S. Voßwinkel, who supported and encouraged me.

Abstract

Health policy in the EU is characterized by two underlying conflicts: First, government interventions or pharmaceutical parallel trade, i.e. trade outside the manufacturer's authorized distribution channel, may induce a conflict between different health policy objectives such as expenditure reduction and distributive objectives. Second, health policy is in the competence of member states, but the EU internal market may generate externalities of national decisions.

This thesis addresses these conflicts of pharmaceutical regulation within the EU.

Initially, I compare a maximum price system (price cap regulation) and a reference price system (reimbursement limit) with respect to their performance in different health policy objectives. The reference price system reduces public pharmaceutical expenditure to a larger extent, but results in higher financial exposure of patients and lower access to pharmaceuticals.

The subsequent chapters investigate the link between pharmaceutical parallel trade and pharmaceutical regulation.

Chapter 4 illustrates that national decisions on health policy, in particular, changes in coinsurance rates, result in externalities under parallel trade. 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 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.

A subsequent chapter compares a coinsurance scheme (consumers pay a percentage of the

drug price out-of-pocket) and an indemnity insurance scheme (reimbursement is independent of the drug price) with respect to the consequences of parallel trade on health care systems, especially on changes of co-payments and changes of public pharmaceutical expenditure. In the destination country, co-payments for patients decrease to a larger extent under indemnity insurance, reductions of public pharmaceutical expenditure occur only under coinsurance. In the source country, co-payments increase less under coinsurance, health expenditure is reduced more under indemnity insurance.

The last chapter studies the effect of pharmaceutical regulation at the wholesale level, in particular, maximum wholesale margins (restriction of pricing by the intermediary) and mandatory rebates (restriction of the pricing by the manufacturer) on drug prices, quantities, and public pharmaceutical expenditure. Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. In a symmetric equilibrium, maximum wholesale margins of both countries partly offset each other, mandatory rebates reinforce each other.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Outline	4
1.3	Related Literature	5
2	Pharmaceutical Regulation and Health Policy Objectives	10
2.1	Introduction	10
2.2	The Model	17
2.3	Regulatory Scenarios	19
2.3.1	No Regulation	19
2.3.2	Maximum Price System	20
2.3.3	Reference Price System	23
2.4	Policy Objectives in Regulation	27
2.4.1	Price Reductions	27
2.4.2	Expenditure Reduction	31
2.4.3	Equity and Access	35
2.4.4	Competition in Pharmaceutical Markets	44
2.5	Welfare Analysis	48
2.6	Conclusion	50

3	The Model	52
3.1	Vertical Control Structure – Supply Side	52
3.2	Vertical Product Differentiation – Demand Side	53
3.3	Timing of the Game	58
4	Externalities of national pharmaceutical policy when markets are integrated through parallel trade	60
4.1	Introduction	61
4.2	Equilibrium without Parallel Trade	64
4.3	Equilibrium with Parallel Trade	66
4.4	The Effect of Parallel Trade	70
4.4.1	Competition Effect in the Destination Country	70
4.4.2	Double Marginalization Effect in the Source Country	71
4.4.3	Price Convergence vs. Divergence	72
4.5	Policy Interdependence under Parallel Trade	73
4.5.1	Change of the Coinsurance Rate in the Destination Country	73
4.5.2	Change of the Coinsurance Rate in the Source Country	77
4.6	Implications for Health Policy	81
4.7	Conclusion	83
5	Pharmaceutical Cost-sharing Systems and Savings for Health Care Systems from Parallel Trade	85
5.1	Introduction	86
5.2	Equilibrium without Parallel Trade	91
5.3	Equilibrium with Parallel Trade	93
5.4	The Effect of Parallel Trade on Health Care Systems	96
5.4.1	Changes in Co-payments and Public Pharmaceutical Expenditure in the Destination Country	97

5.4.2	Changes in Co-payments and Public Pharmaceutical Expenditure in the Source Country	101
5.5	Conclusion	105
6	Pharmaceutical Regulation at the Wholesale Level and Parallel Trade	109
6.1	Introduction	109
6.2	Equilibrium without Regulation	112
6.3	Equilibrium with Maximum Wholesale Margins	118
6.3.1	Maximum Wholesale Margin Regulation in the Destination Country . . .	119
6.3.2	Maximum Wholesale Margin Regulation in the Source Country	122
6.3.3	Symmetric Equilibrium under Maximum Wholesale Margins	124
6.4	Policy Alternative: Mandatory Rebates	125
6.4.1	Mandatory Rebates in the Destination Country	126
6.4.2	Mandatory Rebates in the Source Country	127
6.4.3	Symmetric Equilibrium under Mandatory Rebates	129
6.5	Conclusion	130
7	Conclusion	132
7.1	Summary of Results	132
7.2	Policy Implications	134
7.3	Future Research	137
	Bibliography	138
A	Appendix to Chapter 2	145
A.1	Generic Competition Paradox	145
A.2	Equivalence of Exogenous and Endogenous Quality Difference	148
A.3	Comparative Statics with respect to Quality, Coinsurance Rate, Product Differentiation, and Regulatory Parameter	152

A.4	Health Policy Objectives	157
A.4.1	Price Reductions	157
A.4.2	Expenditure Reduction	158
A.4.3	Equity and Access	162
A.4.4	Competition in Pharmaceutical Markets	169
A.5	Welfare Analysis	172
B	Appendix to Chapter 4	179
B.1	The Effect of Parallel Trade	179
B.1.1	Competition Effect in the Destination Country	179
B.1.2	Double Marginalization Effect in the Source Country	180
B.1.3	Price Convergence vs. Divergence	181
B.2	Change of Coinsurance Rates under Segmented Markets	181
B.2.1	Change of the Coinsurance Rate in the Destination Country	181
B.2.2	Change of the Coinsurance Rate in the Source Country	184
B.3	Change of Coinsurance Rates under Parallel Trade	185
B.3.1	Change of the Coinsurance Rate in the Destination Country	185
B.3.2	Change of the Coinsurance Rate in the Source Country	188
B.4	Change in Health Expenditure	190
B.5	Implications for Health Policy	192
C	Appendix to Chapter 5	195
C.1	Drug Quantities under Segmented Markets	195
C.2	Drug Prices and Quantities under Parallel Trade	196
C.3	The Effect of Parallel Trade	197
C.3.1	Changes in Co-payments and Public Pharmaceutical Expenditure in the Destination Country	200

C.3.2	Changes in Co-payments and Public Pharmaceutical Expenditure in the Source Country	204
C.4	Identical Co-payments as Standard of Comparison	210
C.5	Degree of Product Differentiation	211
D	Appendix to Chapter 6	215
D.1	The Effect of Parallel Trade	215
D.1.1	Demand	215
D.1.2	Equilibrium without Parallel Trade	216
D.1.3	Equilibrium with Parallel Trade	219
D.2	The Effect of Contract Choice on the Effectiveness of Regulation under Segmented Markets	227
D.2.1	Two-part tariff	227
D.2.2	Linear Pricing	231
D.3	Equilibrium with Maximum Wholesale Margins	236
D.3.1	Maximum Wholesale Margin Regulation in the Destination Country	237
D.3.2	Maximum Wholesale Margin Regulation in the Source Country	239
D.4	Mandatory Rebates	241
D.4.1	Mandatory Rebates in the Destination Country	242
D.4.2	Mandatory Rebates in the Source Country	244
E	Eidesstattliche Versicherung	247

List of Figures

2.1	Brand-name Price, $\tau = 0.1$	29
2.2	Brand-name Price, $\tau = 0.5$	31
2.3	Expenditure, $\tau = 0.1$	33
2.4	Expenditure, $\tau = 0.5$	35
2.5	Brand-name Co-payment, $\tau = 0.1$	37
2.6	Brand-name Co-payment, $\tau = 0.5$	38
2.7	Generic Co-payment, $\tau = 0.1$	39
2.8	Generic Co-payment, $\tau = 0.5$	40
2.9	Total Quantity, $\tau = 0.1$	42
2.10	Welfare, $\tau = 0.1$	49
3.1	Coinsurance Rates vs. Indemnity Insurance	58
4.1	Increase of Coinsurance Rate, $c > 0$	79
6.1	Best Response Functions for $w_D^* = w_S^* = 0$ and $w_D^* > 0, w_S^* > 0$	116
B.2.1	Increase of the Coinsurance Rate, $c = 0$	183

Chapter 1

Introduction

1.1 Motivation

Basically all European countries are confronted with rising health care spending, even outpacing GDP growth (Maynard & Bloor, 2003). Expenditure for pharmaceuticals represents a substantial and increasing proportion of it, varying roughly between 11.8 % in the United Kingdom and 20.5 % in Spain (OECD Health Data, 2010). Important factors contributing to high spending include an aging population, increasing incidence and duration of chronic diseases as well as technological improvements in health care (Ess, Schneeweiss & Szucs, 2003). In addition, pharmaceutical market imperfections such as agency imperfections, informational asymmetries and moral hazard, which create reduced price sensitivity on the demand side and a certain degree of market power on the supply side, intensify this cost trend (Mossialos & Le Grand, 1999; Hurley, 2001).

The continuous increase in public health expenditure has induced a considerable number of government interventions and most European countries have introduced regulatory instruments to contain public spending (Maynard & Bloor, 2003). On the supply side, maximum prices or price negotiations are intended to restrict monopoly pricing and reduce the prices of covered services. Demand side instruments such as co-payments and reference prices are intended to

increase price sensitivity on the demand side.

This thesis focuses on two recurring conflicts of pharmaceutical market regulation at the European level.

First, government interventions or pharmaceutical parallel trade, i.e. trade outside the manufacturer's authorized distribution channel, may induce a conflict between different health policy objectives. The main objective of pharmaceutical regulation is the reduction of public expenditure, as public insurance schemes bear the majority of these expenses (Danzon, 1997). But at the same time, distributional objectives are also a major concern (Hurley, 2003). This includes limiting the financial exposure of patients and ensuring broad access to pharmaceuticals. Furthermore, pharmaceutical regulation may also aim at stimulating competition on pharmaceutical markets to restrict monopoly pricing. These health policy objectives may conflict with each another. A government intervention working towards one health policy objective may harm another objective at the same time, e.g. reducing insurance coverage reduces public health expenditure, but increases financial exposure of patients and restricts access to pharmaceuticals. This dissertation addresses the conflict between different health policy objectives in chapters 2, 4, 5, and 6. Chapter 2 examines two regulatory instruments with respect to their performance in the above mentioned health policy objectives. Chapter 4 illustrates that co-payment changes may generate a conflict between changes in financial exposure of patients and changes in public pharmaceutical expenditure, chapter 5 studies the effect of parallel trade on health care systems, with emphasizing the conflict between savings for patients and health insurances. Chapter 6 shows that the restriction of pricing of intermediaries may reduce drug prices, but also restricts competition at the same time.

Second, in the European Union, there may be a conflict between the national competence of member states in health policy and the internal market. In the European Economic Area¹, regional exhaustion of property rights and the free movement of goods allows intermediaries or parallel traders to import pharmaceuticals from other countries without the authorization

¹The European Economic Area includes the European Union plus Norway, Iceland, and Switzerland.

of the manufacturer, i.e. to engage in parallel trade. Tremendous price differences of up to 300 % percent between member states give rise to this kind of arbitrage (Glynn, 2009). These price differences may emerge from pharmaceutical manufacturers price-discriminating between different countries, different national pharmaceutical regulations in the individual member states and/or divergent wholesale prices (NERA, 1999; EU Commission, 2003; Enemark et al., 2006). Destination countries are high-price countries, such as Denmark, Germany, the Netherlands, Norway, Sweden, and the United Kingdom, where pricing is relatively free; source countries are characterized by strict price regulation, e.g. France, Greece, Italy, Portugal, and Spain (Kanavos & Costa-Font, 2005). Pharmaceutical parallel trade, the exploitation of these price differences, had a volume of € 4.8 bn. in the European Union in 2007 (Glynn, 2009). In 2009, the share of parallel imports in pharmacy market sales ranged between 20 % in Denmark and 1.3 % in Finland (EFPIA, 2011). At the same time, health policy, including pharmaceutical price regulation and cost-sharing instruments, falls in the member states' competence. So far, harmonization of different European rules has primarily concerned drug authorization procedures (Kyle, 2009). With respect to pharmaceutical price regulation and reimbursement, the 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 (so far) the last measure with the objective of harmonizing national price regulation and reimbursement rules (Hancher, 2004). It includes rules for the control of pharmaceutical prices (respective measures have to be efficient, transparent and fair) (Desogus, 2011; Hancher, 2004). Among the member states, agreements on further harmonization could not be reached (Desogus, 2011; Hancher, 2004). Consequently, drug pricing – pharmaceutical price regulation and reimbursement rules – remain under exclusive national competence (Desogus, 2011). This conflict inherent in the institutional setting implies that not only the manufacturer's pricing decisions but also national decisions on health policy are interdependent.

Externalities may emerge, whereby decisions about pharmaceutical price regulation and cost-sharing instruments in one country have an effect on other countries as well. At the same

time, the national competence in health policy allows member states to take their own specific characteristics, their health care system and distributive preferences into account. Chapter 4, 5, and 6 cover the issue of national competence in health policy and the externalities emerging under market integration. Chapter 4 studies externalities of national decisions on health policy, in particular changes of coinsurance rates. Chapter 5 shows that national competence in health policy allows member states to design their cost-sharing systems according to their status as a source or destination country of parallel trade and thereby maximize benefits from parallel trade. Chapter 6 analyzes the externalities of pricing restrictions at the wholesale level.

1.2 Outline

This introduction is completed by a short overview of the related literature.

Chapter 2 studies two instruments of pharmaceutical regulation in a vertical differentiation model with a brand-name drug and a generic. It compares a maximum price system (price cap regulation) and a reference price system (reimbursement limit) with respect to their performance in achieving certain health policy objectives, namely the reduction of public expenditure, the limitation of financial exposure of patients, the improvement of access to pharmaceuticals, and the stimulation of competition. Results show that for identical price reductions of the brand-name drug, the lower reimbursement amount under the reference price system results in lower health expenditure, but higher financial exposure of patients.

Chapter 3 presents the two country model following Maskus & Chen (2002), Chen & Maskus (2005), based on which the link between parallel trade and health policy is investigated.

Chapter 4 examines externalities of national decisions on health policy, in particular, changes in coinsurance rates under market integration by parallel trade. Results show that parallel trade generates a price-decreasing competition effect in the destination country and a price-increasing double marginalization effect in the source country. Changes in the coinsurance rate in one country affect the magnitude of these effects caused by parallel trade in the respective other country.

Chapter 5 analyzes the consequences of parallel trade on health care systems. It compares two cost-sharing systems – coinsurance (patients pay a percentage of the price) and indemnity insurance (reimbursement is price-independent) – with respect to price effects and performance in reducing public pharmaceutical expenditure. Results show that the price effects of parallel trade are independent of the cost-sharing system. In the destination country, savings for patients from lower priced parallel imports are higher under indemnity insurance, whereas reductions of public pharmaceutical expenditure occur only under coinsurance. In the source country, patients co-payments increase less under coinsurance, while reductions in public pharmaceutical expenditure are higher under indemnity insurance.

Chapter 6 explores the effects of maximum wholesale margins (restriction of pricing by intermediaries) on drug prices and quantities under parallel trade. Results show that maximum wholesale margins enhance the manufacturer’s ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. In the symmetric equilibrium with both countries applying maximum wholesale margins, regulatory instruments exhibit an offsetting effect. Mandatory rebates are a policy alternative that restricts pricing of the manufacturer. In the symmetric equilibrium with both countries applying mandatory rebates, regulatory instruments exhibit a reinforcing effect with respect to prices.

Finally, Chapter 7 summarizes the results of the preceding chapters, discusses policy implications and presents approaches for future research.

1.3 Related Literature

The analysis in this thesis relates to two strands of literature, the literature on pharmaceutical (price) regulation and the literature on pharmaceutical parallel trade.

The literature on pharmaceutical regulation has mainly focused on the reference price system. Empirically, the reference price system has been analyzed extensively (see Puig-Junoy, 2010 for a survey). The few theoretical models arrive at very different results. Miraldo (2007) analyzes the impact of the reference price system on drug prices. Her results suggest that if drugs are

of equal quality, reference pricing may lead to higher prices. Mestre-Ferrándiz (2003) identifies a certain interval, in which the reference price has to be set to stimulate price competition and lead to lower pharmaceutical expenditure. Merino-Castello (2003) shows that the reference price system induces significantly lower brand-name prices, while generic prices "remain more or less constant". She also finds that the generic market share "remains constant or even decreases" under the reference price system. Cabrales (2003) compares the reference price system with a binding price ceiling. His results suggest that the price of the high quality drug may be higher under duopoly than under monopoly. Also, he finds that the a lower binding price ceiling results in a higher market share of the high quality drug. Brekke, Königbauer & Straume (2006) contrast a generic reference price system, in which the reference price is set for a group of chemically equivalent drugs (a brand-name drug and the corresponding generic version), with a therapeutic reference price system, in which the reference price applies to a group of therapeutically equivalent drugs. They find that the therapeutic reference price system induces stronger competition, thus causing lower drug prices and pharmaceutical expenditures than the generic reference price system, which distorts drug choices more. Brekke, Holmas & Straume (2010) compare price cap regulation with reference pricing. Their model suggests – and is confirmed by empirical evidence – that reference pricing results in price reductions, which are higher for brand-name drugs – and correspondingly induces stronger generic competition and lower brand-name market shares.

The analysis of this thesis is linked to three aspects of the literature on pharmaceutical parallel trade. First, several studies examine, whether the exploitation of arbitrage opportunities through parallel trade is able to contain (public) pharmaceutical expenditure in the destination countries of parallel imports. Empirical evidence on this is ambiguous. Three recent studies have presented contradictory results with respect to the ability of parallel trade to generate savings for health insurance funds and patients. On the one hand, a study by Kanavos et al. (2004) carried out at the LSE Health and Social Care Research Centre at the London School of Economics (commonly referred to as the LSE-study) finds only modest direct savings accruing

to health insurances (a total of € 45 m. for Denmark, Germany, the Netherlands, Norway, Sweden, and the UK) and no (measurable) patient benefits. They suggest that parallel traders are the main beneficiaries of parallel trade. On the other hand, a study by West & Mahon (2003) conducted at the York Health Consortium at the University of York (commonly referred to as the York-study) concludes that parallel trade generates considerable savings. It estimates total direct savings from parallel trade accruing to both health insurances and patients at € 635 m. for Denmark, Germany, the Netherlands, and the UK. A third study, conducted at the University of Southern Denmark by Enemark et al. (2006) also concludes that parallel trade gives rise to significant savings, both direct to patients and health insurances, amounting to € 466 m. in Denmark, Germany, Sweden, and the UK.² The different results reflect differences in methodology and in the range of products covered, but they agree on one thing: The effects of parallel trade, that is, the level of savings and cross-countries differences in savings and the split of savings between health insurances and patients depend to a large extent on the co-payment structure (Kanavos et al., 2004; Enemark et al., 2006). However, the importance of cost-sharing systems for the consequences of parallel trade has not attracted much attention in the theoretical literature on parallel trade. Bordoy & Jelovac (2005) and Köksal (2009) examine the effect of cost-sharing structures on parallel trade. Bordoy & Jelovac (2005) investigate the implications of cross-country differences in coinsurance rates for the welfare effects of parallel trade in a vertical differentiation model with horizontal arbitrage. If parallel trade is driven by differences in coinsurance rates, it reduces welfare, as it reallocates drugs from patients with a higher valuation of drug consumption to patients with a lower valuation. On the contrary, parallel trade increases welfare, if it is based on differences in health needs. Based on the Bordoy & Jelovac (2005)-model, Köksal (2009) compares price effects caused by parallel trade under coinsurance and reference pricing. Under reference pricing, price reductions from parallel

²Other empirical investigations of the effects of parallel trade are limited to Sweden and Finland in coverage. Based on a sample of 6 drugs in Sweden, Persson et al. 2001 estimate savings of parallel trade to € 13 m (quoted in Enemark et al. 2006). Linoosma et al. 2003 document savings of € 4.9 m for 169 drugs in Finland. Ganslandt & Maskus 2004 find for 50 top-selling drugs in Sweden that parallel trade gives rise to price reductions of up to 19%.

trade in the destination country are higher than under coinsurance. Furthermore, reference pricing does not affect the drug price in the exporting country.

Second, several papers study whether market integration through parallel trade results in the erosion of price differences. Several theoretical models predict price convergence, e.g. Rey (2003) or Jelovac & Bordoy (2005). Other theoretical models even assume uniform prices under parallel trade, e.g. Pecorino (2002), Valetti (2006). Empirical evidence on price convergence, however, is mixed. Kanavos et al. (2004) find no evidence for price competition or price convergence. On the contrary, West & Mahon (2003) 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 % to 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 of 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. 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.

Third, the influence of parallel trade on pharmaceutical price regulation has been examined in the literature. On the one hand, parallel trade may restrict policy choices, as Rey (2003) suggests. By leading to a price convergence towards lower prices, parallel imports thwart a government's efforts to contribute more to R&D by granting higher prices. On the other, parallel trade may strengthen the manufacturer's position in price negotiations. Pecorino (2002) shows that if parallel trade to an unregulated market is possible, a manufacturer will make less concessions in drug price bargaining in a potential source country of parallel imports. Königbauer (2004) and

Grossman & Lai (2006) find that the possibility of parallel trade limits the scope for international free riding on R&D contributions by providing weaker patent protection and imposing price caps. Parallel trade may even mitigate the negative impact of price controls on pharmaceutical innovation and result in more highly innovative and less me-too drugs (Schlaepfer, 2008).

Chapter 2

Pharmaceutical Regulation and Health Policy Objectives

Abstract

This chapter analyzes a maximum price system (price cap regulation) and a reference price system (limitation of reimbursement) in a vertical differentiation model with a brand-name drug and a generic. In particular, both instruments are compared with respect to their performance in reducing public expenditure, limiting financial exposure of patients, improving access to pharmaceuticals, and stimulating competition. For identical regulatory prices, free pricing under the reference system tends to result in a higher price for the brand-name drug. For identical price reductions of the brand-name drug, the lower reimbursement amount under the reference price system results in lower health expenditure, but higher financial exposure of patients.

2.1 Introduction

This chapter compares a maximum price system and a reference price system with respect to their performance in reducing public expenditure, limiting financial exposure of patients, improving access to pharmaceuticals, and stimulating competition.

This analysis is motivated by the following observations: The two regulatory instruments – the maximum price system and the reference price system – examined in this chapter are applied in almost all Western European countries. Pharmaceutical markets are characterized by patients not paying the full price of pharmaceuticals out-of-pocket, but only a co-payment, while the health insurance reimburses the remaining part. In a system of public insurance, reimbursement brings about public expenditure. An increase of co-payments would relieve the public purse, but contradict distributive objectives in the supply of pharmaceuticals. Consequently, regulatory instruments such as the maximum price system or the reference price system are introduced to reduce public health expenditure.

Maximum price systems, in which the regulatory body sets a price cap that can be charged for a drug, are applied in Austria, Belgium, Finland, France, Greece, Ireland, Italy, Norway, Portugal, and Spain (Espin & Rovira, 2007). The reference price system, in which the regulator sets a ceiling for the amount reimbursable (reference price) for a group of pharmaceuticals (cluster), has first been introduced in Germany in 1989, other states followed. Nowadays, most Western European countries apply reference pricing: Belgium, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, Portugal, Spain, and the United Kingdom¹(Puig-Junoy, 2010; Espin & Rovira, 2007). Sweden and Norway have implemented reference price systems, but have eliminated them again (Espin & Rovira, 2007). In reference price systems, the group of pharmaceuticals, for which the reference price is the maximum reimbursement, is defined in terms of interchangeability (López-Casasnova & Puig-Junoy, 2000). This can be understood in a chemical (drugs contain the same active ingredient), pharmacological (drugs belong to the same therapeutic category), or therapeutic (drugs have the same therapeutic function) way² (López-Casasnova & Puig-Junoy, 2000). Firms remain free to charge prices (Danzon, 2001). If the

¹Some states, e.g. Finland, apply both the maximum price and the reference price system. In Finland, price control applies to wholesale prices, which are calculated on basis of wholesale prices in several other European countries (Espin & Rovira, 2007). Reference pricing then limits reimbursement of the retail price. That is, in this case, both instruments are applied at different levels, the maximum price system at the wholesale level and reference pricing at the retail level. Note that both instruments cannot be applied at the same level for the same class of pharmaceuticals, as they are mutually exclusive.

²See López-Casasnova & Puig-Junoy (2000) for a survey of limitations with respect to the interchangeability of drugs within the same reference group.

manufacturer's price exceeds the reference price, the patient has to pay the difference between the market price and the reference price him/herself (Danzon, 2001). That is, the reference price system involves an additional co-payment, which can be considered avoidable in the sense that purchasing a drug which is priced at or below the reference price does not involve the additional co-payment (López-Casasnová & Puig-Junoy, 2000).

Besides potential differences in performance with respect to health policy objectives, the choice between both regulatory instruments is determined by different economic rationales. Supply side measures are commonly thought to limit the market power of pharmaceutical firms stemming from patients' and physicians' price insensitivity (Scherer, 1996). Reference pricing aims at increasing market transparency and allowing consumers to compare a drug's price in relation to prices of suitable substitutes (Danzon, 2001; López-Casasnová & Puig-Junoy, 2000). Thereby reference pricing introduces an element of price-sensitivity and producers may only maintain prices above the reference price, if additional quality or value is associated with the respective drug (Espin & Rovira, 2007). By making demand more price elastic, reference pricing creates incentives to substitute less expensive generics for higher priced brand-name drugs.

As discussed in the introduction, health policy objectives may include reducing public expenditure, limiting financial exposure of patients, ensuring broad access to pharmaceuticals, and stimulating competition on pharmaceutical markets.

An aging population with growing health needs, technological progress, and pharmaceutical market imperfections result in high expenditure for pharmaceuticals. Public insurance schemes bear the majority of these expenses (Danzon, 1997). Pharmaceutical markets are characterized by agency imperfections, informational asymmetries and moral hazard, that create reduced price sensitivity on the demand side and a certain degree of market power on the supply side (Mossialos & Le Grand, 1999; Hurley, 2001). As stability of contribution rates is a sociopolitical aim, regulatory instruments have been introduced to contain spending for pharmaceuticals.

In an effort to reduce moral hazard in health care utilization, co-payments have been introduced in basically all Western European countries, mostly in the form of coinsurance rates, where

patients pay a percentage of the price (Mossialos & Le Grand, 1999). In the European Union, various forms of coinsurance (with coinsurance rates fixed or depending on drug classes or price levels) are applied in Belgium, Cyprus, Denmark, Estonia, Finland, France, Germany, Greece, Latvia, Lithuania, Poland, Portugal, Slovakia, Slovenia, Spain, and Sweden (Espin & Rovira, 2007)³. Co-payments are intended to make patients more aware of the prices of pharmaceuticals or health care services and to reduce the use of pharmaceuticals or services that are not really necessary⁴ (Robinson, 2002). Usually, co-payments are set too low to affect the use of pharmaceuticals or health services. In general, it is considered not equitable to increase co-payments to a level high enough to discourage the unnecessary use of drugs or services (Mossialos & Le Grand, 1999). The price elasticity of demand for health care is higher for those, who spend a larger share of their income on pharmaceuticals, i.e. the chronically ill and poor (Grootendorst, 2006). Evidence from the RAND Health Experiment suggests that utilization of pharmaceuticals or health care services is more responsive to co-payments for low-income groups (see Zweifel & Breyer, 2006 for a survey). Price elasticity of demand increases with the coinsurance rate, but as guaranteeing broad access to services is a policy objective, the coinsurance rate cannot constitute an instrument of cost containment. Several EU member states such as Belgium, Finland, France, Germany, and the Netherlands have introduced a ceiling for all co-payments (Mossialos & Le Grand, 1999). Also, empirically, co-payments have never been a preferred instrument of cost containment, nor has the extent, that is the percentage of the price to be paid by patients, increased significantly (Mossialos & Le Grand, 1999). Instead, co-payments tend to become more complex. The model used in this chapter features coinsurance rates, whereby co-payments result in out-of-pocket expenditure for patients and reimbursement in publicly funded health expenditure. This allows me to analyze also the effect of regulation on public pharmaceutical expenditure and financial exposure of patients.

³Of the remaining countries, the Netherlands and Malta have no co-payment, Austria, Italy, and the UK apply a flat rate (with a charge per service), and Ireland and Sweden apply a deductible (consumers have to pay the first x Euros, until insurance coverage begins) (Espin & Rovira, 2007).

⁴In addition, co-payments are also supposed to raise revenue, which is in partial conflict with the aim of reducing moral hazard (Mossialos & Le Grand, 1999).

Competition between off-patent brand-name drugs and generics is mainly characterized by first-mover advantages and consumer perception. Generics tend to enter the market at substantial discounts to the price prevailing before patent expiry of the brand-name drug (Scherer, 2001). Additional generic competitors then result in further price reductions, whereas the price of the brand-name drug remains stable or even increases (Scherer, 2001). The observation of increasing brand-name prices in response to generic entry has been labelled the generic competition paradox. For instance, Frank and Salkever (1997) describe a positive relationship between the price of an off-patent brand-name drug and the number of generic competitors in their analysis of price trends of 83 drugs between 1984 and 1987. They observe an increase of the brand-name drug price by 0.7 % for every additional market entrant. Grabowski & Vernon (1992) discover price increases with growth rates above the inflation rate for a sample of 18-off patent drugs between 1983 and 1987. Frank & Salkever (1992) explain the diverging price trends of brand-name drugs and generics with a segmented demand side of price-sensitive consumers willing to switch to generic and brand-loyal consumers willing to pay a higher price for the brand-name drug. As the remaining brand loyal demand is less elastic than the demand from both segments, the optimization behavior of the brand-name drug supplier causes the brand-name drug price to rise, as it is more profitable to serve the brand-loyal segment at a high price than to compete with generic producers at a low price for both segments. Indeed, empirical observations confirm different price elasticities for the demand for brand-name drugs and generics (Ellison et al., 1997). The existence of a brand-loyal segment implies that brand-name drugs and generics are not considered to be homogenous products by consumers, but rather differentiated products. In addition to objective differences between brand-name drugs with respect to additives, consumers sometimes associate a lower quality with generic versions (see Gaither et al., 2001 for a survey of consumers' perception of generics). Also, the inability to assess the quality of drugs before consumption and the risk of bad choices such as adverse side effects contributes to first-mover advantages for brand-name producers, allowing them to maintain market shares at substantial higher prices (Scherer, 1996). The model in this chapter therefore explicitly assumes a heteroge-

neous demand side, where consumers differ in their valuation of the drug, and a certain degree of product differentiation between the brand-name drug and the generic versions - either due to consumer perceptions or firms' investment decisions. An explanation of the generic paradox without the explicit assumption of brand-loyalty can be found in Appendix A.1.

The paper most closely related to this chapter is Brekke, Holmas & Straume (2010), who compare price cap regulation with reference pricing. Their model suggests that reference pricing results in price reductions, which are higher for brand-name drugs, and correspondingly it induces stronger generic competition and lower brand-name market shares (Brekke, Holmas & Straume, 2010). This is in line with empirical evidence from Pavcnik (2002), who studies prices for oral antidiabetics and antiulcers in Germany between 1986 and 1996. She finds major price reductions for both brand-name and generic drugs, with larger reductions for brand-name prices. With respect to price cap regulation, Brekke, Holmas & Straume (2010) find that a reduction of the maximum price reduces the generic market share, that is, stricter direct price regulation weakens generic competition. A study by the European Commission, which analyzes the prices of 122 active ingredients in 17 EU countries between 2000 and 2007, confirms that price cap regulation affects price competition negatively (European Commission, 2009).

Whereas the existing literature mainly analyzes firms' pricing behavior and the potential of reference pricing in reducing drug prices, this analysis takes a broader perspective and focuses more on the overall implications of these regulatory instruments in more than one dimension. In other words, as pharmaceutical regulation may exhibit a trend of inhibiting competition due to less generic entry, not only price reductions but also high generic market shares and a sufficient degree of competition are essential (Danzon, Wang & Wang, 2005). Moreover, price reductions may not be politically intended if they result in higher financial exposure of patients. Consequently, the purpose of this chapter is to analyze the performance of price caps and reference pricing with respect to several policy objectives. Furthermore, the explicit comparison of both instruments takes into account that both instruments constitute policy alternatives and that the relative importance of health policy objectives determines what instrument is chosen.

The vertical product differentiation with a heterogenous demand side takes into account that both objective and subjective differences between brand-name drugs and generics exist and that consumers choose which version to buy based on their valuation of the drug. The result is an endogenous segmentation of the demand side, with a segment of consumers with high valuation of pharmaceuticals buying the brand-name drug and a segment of consumers with an intermediate valuation purchasing the generic. Consumers with a very low valuation will buy neither version of the drug, this allows me to analyze the effect of regulation on the access to pharmaceuticals, if it results in a change of market coverage. The maximum price system is modelled as a price cap amounting to the generic price in the benchmark case of no regulation plus a mark-up, the reference price system assumes a reference price as the weighted average of brand-name and generic price. Both constructions allow me to analyze different degrees of regulation explicitly.

For identical regulatory degrees, the endogenous specification of the reference price system captures the firms' strategic response to the introduction of a reimbursement limit and generates higher price reductions for the brand-name drug under the reference price system. For identical regulatory prices, the reference price system tends to result in higher drug prices, as free pricing associated with the reference price system enables the brand-name producer to skim off additional willingness to pay by setting a price above the reference price. The additional co-payment element (patients have to pay the difference between the market price of the brand-name drug and the reference price) results in higher financial exposure of patients, but also lower public expenditure. In addition, the effects of both regulatory instruments on prices and market shares are similar to the results of Brekke, Holmas & Straume (2010). Whereas the maximum price system does not change the brand-name price premium and results in a lower generic market share, the reference price system reduces the brand-name price premium and brings about a higher generic quantity, but *unchanged* generic market share⁵.

The rest of the chapter is organized as follows. The next section presents the vertical dif-

⁵Brekke, Holmas & Straume (2010) identify a higher market share of generics under reference pricing. They find that the market share of generics is increasing in the weight of the generic in the reference price, i.e. in the degree of regulation.

differentiation model with a brand-name drug producer and a generic drug producer. Section 2.3 analyzes the benchmark case of no pharmaceutical regulation, the case of regulation through a price cap (maximum price system), and the case of regulation through a reimbursement limit (reference price system). Section 2.4 compares the regulatory scenarios with respect to the health policy objectives of the reduction of public expenditure, the limitation of financial exposure of patients, the improvement of access to pharmaceuticals, and the stimulation of competition. Section 2.5 analyzes welfare and section 2.6 concludes.

2.2 The Model

Consider a therapeutic market with two competing drugs, an off-patent brand-name drug b and the corresponding generic version g . This corresponds to the duopolistic transition period after patent expiry, with the first generic having already entered the market.

Both drugs contain the same active ingredient, but differ in both objective and subjective terms. On the one hand, generics and brand-name drugs show considerable differences with respect to binders, fillers, preservatives and density of packing (bioequivalence⁶), which may affect therapeutic efficacy (Scherer, 1996). On the other hand, generics may be perceived as of lower quality (see Gaither et al. (2001) for a survey on the lower quality perception of generics). In addition, there is evidence that the price of a drug may serve as quality indicator (Waber et al., 2008). Also uncertainty with respect to whether the generic version is really equivalent to the brand-name version may contribute to a lower willingness to pay for the generic. The property of pharmaceuticals as experience goods, i.e. the difficulty of evaluating quality ex ante, and the risk of bad choices such as adverse side effects add to the uncertainty (Scherer, 1996).

Consumers differ with respect to their gross valuation of drug consumption θ , which is uniformly distributed on the interval normalized to unity. A consumer with a positive net utility of drug consumption will choose the most preferred drug version by trading off (objective

⁶Differences in bioequivalence may imply also differences in bioavailability, which refers to the rate and extent, at which the active ingredient is absorbed.

and perceived) drug quality against drug co-payment. The higher the gross valuation of drug treatment θ , the more the consumer is willing to pay in order to purchase the (high-quality) brand-name drug. The consumer heterogeneity can be interpreted as differences in willingness to pay for a brand-name, differences in risk aversion regarding the trial of substitutes, differences in the severity of the condition or the level of suffering or differences in prescription practices (see e.g. Brekke, Holmas & Straume, 2010). This results in an endogenous segmentation of the demand side.

The quality difference between the brand-name and the generic version may be considered as either an exogenous difference stemming from the different perception of brand-name drugs and generics or as an endogenous one emerging from the pharmaceutical firms' investment decision. Appendix A.2 shows the equivalence of exogenous and endogenous quality difference between both drugs with respect to mechanics of the model.

A consumer who buys a drug i obtains a net utility of

$$U(\theta, \tau, s, c_i) = \begin{cases} \theta s - c_i & \text{if } i = b \\ \theta \tau s - c_i & \text{if } i = g, \end{cases} \quad (2.1)$$

where s denotes the quality of the brand-name drug, the parameter τ captures subjective and/or objective differences between the brand-name drug and the generic version and c_i is the patient co-payment for drug i . The utility derived from no drug purchase is zero.

The marginal consumer who is indifferent between buying the brand-name drug b or the generic version g has a gross valuation θ^* , given by

$$\theta^* s - c_b = \theta^* s \tau - c_g, \text{ yielding } \theta^* = \frac{c_b - c_g}{s(1 - \tau)}$$

while a consumer who is indifferent between buying the generic and not buying at all has a gross valuation $\underline{\theta}$, given by

$$\underline{\theta} s \tau - c_g = 0, \text{ yielding } \underline{\theta} = \frac{c_g}{s \tau}.$$

Hence, demand for brand-name drug b and for the generic g is given by

$$q_b = 1 - \frac{c_g}{s\tau} \text{ and } q_g = \frac{c_g}{s\tau} - \frac{c_g}{s\tau}.$$

Production technologies exhibit constant marginal costs, which are normalized to zero for simplicity, such that profits are given as

$$\Pi_i = p_i q_i. \tag{2.2}$$

2.3 Regulatory Scenarios

2.3.1 No Regulation

Consider a system with no regulation as a benchmark. Consumers are partially insured, a co-payment in the form of a proportion of the price (coinsurance) applies. The remaining amount is reimbursed by the health insurance.

Coinsurance rates and cost-sharing in general are intended to make patients aware of the costs of services and deter them from using services that are not needed and thus reduce moral hazard (Robinson, 2002). Price elasticity of demand increases with the coinsurance rate, but as guaranteeing broad access to services is a policy objective, the coinsurance rate does not constitute an instrument of cost containment.

Co-payment for the brand-name drug and the generic is given as

$$c_i = \kappa p_i, \tag{2.3}$$

where κ is the coinsurance rate.

Thus, demand functions are given as

$$q_b = 1 - \frac{\kappa(p_b - p_g)}{s\tau} \text{ and } q_g = \frac{\kappa(p_b - p_g)}{s\tau} - \frac{\kappa p_g}{s(1 - \tau)}. \tag{2.4}$$

The firms' profits are given as

$$\Pi_b = p_b \left(1 - \frac{\kappa(p_b - p_g)}{s\tau} \right) \text{ and } \Pi_g = p_g \left(\frac{\kappa(p_b - p_g)}{s\tau} - \frac{\kappa p_g}{s(1 - \tau)} \right). \quad (2.5)$$

First-order conditions for profit-maximizing drug prices yield the best-response functions

$$p_b = \frac{1}{2\kappa} (s\tau + \kappa p_g) \text{ and } p_g = \frac{1}{2} p_b (1 - \tau).$$

Best-response functions are increasing, prices are strategic complements.

Resulting equilibrium prices are

$$p_b = \frac{2s\tau}{\kappa(3 + \tau)} \text{ and } p_g = \frac{s\tau(1 - \tau)}{\kappa(3 + \tau)}. \quad (2.6)$$

Thus, indifferent consumers are located at

$$\theta^* = \frac{1 + \tau}{3 + \tau} \text{ and } \underline{\theta} = \frac{\tau}{3 + \tau}$$

and equilibrium quantities are

$$q_b = \frac{2}{3 + \tau} \text{ and } q_g = \frac{1}{3 + \tau}. \quad (2.7)$$

Profits are given by

$$\Pi_b = \frac{4s\tau}{\kappa(3 + \tau)^2} \text{ and } \Pi_g = \frac{s\tau(1 - \tau)}{\kappa(3 + \tau)^2}. \quad (2.8)$$

2.3.2 Maximum Price System

The main form of supply-side regulation is direct price control, in which a maximum price (price ceiling) limits the capacity of firms to set prices. In the European Union, all countries except for Denmark, Germany, the Netherlands, Malta, Sweden, and the UK control pharmaceutical prices directly (Espin & Rovira, 2007).

In this model, it is assumed that the regulator sets a maximum price equal to the price of the generic plus a markup. Health policy makers often compare prices for brand-name drugs and generics, as both versions of the drug are considered to be equivalent.

A maximum price of

$$\hat{p} = \underbrace{\frac{s\tau(1-\tau)}{\kappa(3+\tau)}}_{p_g} + (1-\mu) \underbrace{\frac{s\tau(1+\tau)}{\kappa(3+\tau)}}_{m=p_b-p_g}, \quad (2.9)$$

corresponds to the generic price in the no regulation case ($p_g = \frac{s\tau(1-\tau)}{\kappa(3+\tau)}$) plus a fraction $1-\mu$, with $\mu \in (0, 1)$, of a markup $m = \frac{s\tau(1+\tau)}{\kappa(3+\tau)}$ ⁷. The case of $\mu = 0$ corresponds to no regulation (the brand-name drug producer is able to charge the optimal p_b), while the case of $\mu = 1$ corresponds to the strictest regulation possible (the price of the brand-name drug is set to the price of the generic version). The regulatory parameter μ is a measure for the strictness of regulation.

Patient co-payments are not affected by the maximum price system and are still given as $c_i = \kappa p_i$.

For a given price cap, the generic producer's best-response function is $p_g^\mu = \frac{1}{2}p_b^\mu(1-\tau)$.

Equilibrium prices are

$$p_b^\mu = \frac{s\tau(2-\mu(1+\tau))}{\kappa(3+\tau)} \text{ and } p_g^\mu = \frac{s\tau(2-\mu(1+\tau))(1-\tau)}{2\kappa(3+\tau)}. \quad (2.10)$$

Compared to the benchmark case of no regulation, both drug prices are lower under the maximum price system:

$$\begin{aligned} p_b - p_b^\mu &= \frac{s\tau\mu(1+\tau)}{\kappa(3+\tau)} > 0 \\ p_g - p_g^\mu &= \frac{s\tau\mu(1+\tau)(1-\tau)}{2\kappa(3+\tau)} > 0. \end{aligned}$$

The brand-name price is set to a lower amount by the regulator. The generic producer has to lower his price in response to the lower brand-name price as well, as he has to compensate

⁷This structure allows to analyze different degrees of regulation explicitly.

consumers for the lower (perceived) quality by pricing at a certain discount from a given brand-name price. This relationship is given by the generic producer's best-response function.

Equilibrium quantities are

$$q_b^\mu = \frac{4 + \mu(1 + \tau)^2}{2(3 + \tau)} \text{ and } q_g^\mu = \frac{2 - \mu(1 + \tau)}{2(3 + \tau)}. \quad (2.11)$$

The quantity of the brand-name drug is higher under the maximum price system, the quantity of the generic is lower:

$$q_b - q_b^\mu = -\frac{\mu(1 + \tau)^2}{2(3 + \tau)} < 0, \quad q_g - q_g^\mu = \frac{\mu(1 + \tau)}{2(3 + \tau)} > 0.$$

The brand-name drug is sold to more consumers under regulation, as the lower price makes it attractive also for consumers with an intermediate valuation θ who purchased the generic before. For the generic producer, this sales volume lost to the brand-name drug is larger than the volume gained from consumers who have not purchased before and now purchase the generic due to its lower price. That is, the consumer indifferent between the brand-name drug and the generic moves away from one by a larger distance than the consumer indifferent between the generic and not buying moves towards zero. The locations of indifferent consumers under the maximum price system are

$$\theta^{*\mu} = \frac{(2 - \mu(1 + \tau))(1 + \tau)}{2(3 + \tau)} \text{ and } \underline{\theta}^\mu = \frac{(2 - \mu(1 + \tau))\tau}{2(3 + \tau)},$$

of which $\underline{\theta}^\mu$ is closer to its counterpart under no regulation than $\theta^{*\mu}$ is:

$$\theta^* - \theta^{*\mu} = \frac{\mu(1 + \tau)^2}{2(3 + \tau)} > \underline{\theta} - \underline{\theta}^\mu = \frac{\mu\tau(1 + \tau)}{2(3 + \tau)}.$$

Profits are given as

$$\Pi_b^\mu = \frac{s\tau(2 - \mu(1 + \tau))(4 + \mu(\tau + 1))^2}{2\kappa(3 + \tau)^2} \text{ and } \Pi_g^\mu = \frac{s\tau(1 - \tau)(2 - \mu(1 + \tau))^2}{4\kappa(3 + \tau)^2}. \quad (2.12)$$

Compared to the benchmark case of no regulation, both firms' profits are lower. For the brand-name producer, the profit-decreasing effect of a lower price exceeds the profit-increasing effect of a higher sales volume. The generic producer sells a lower quantity at a lower price.

2.3.3 Reference Price System

A common form of demand side or reimbursement regulation is the reference price system, in which the regulatory body sets a reimbursement ceiling ("reference price") for a group of drugs ("cluster"). These drugs, which are considered interchangeable, are not reimbursed based on their market price, but the reference price. Firms remain free to charge higher prices. If a consumer wishes to purchase a drug, which is priced above the reference price, he/she has to pay the difference between the market price of the drug and the reference price in addition to the usual co-payment. In Western Europe, Belgium, Denmark, Estonia, Germany, Greece, Hungary, Italy, the Netherlands, Poland, Portugal, and Spain use reference price systems (Espin & Rovira, 2007). Norway and Sweden have adopted reference price systems, but have abolished them again (Espin & Rovira, 2007).

The reference price is a linear function of both drug prices:

$$p_r^\varrho = \varrho p_g^\varrho + (1 - \varrho)p_b^\varrho, \quad (2.13)$$

where ϱ ($\varrho \in (0, 1)$) is an exogenous weight⁸. For $\varrho = 1$, the reference price and consequently the reimbursement amount corresponds to the price of the generic, for $\varrho = 0$, the reference price and reimbursement amount coincide with the price of the brand-name drug, which amounts to

⁸Note that a specification of the reference price as a convex combination of both drug prices ($p_r^\varrho = \varrho p_g^\varrho + (1 - \varrho)p_b^\varrho$) is equivalent to a specification of the reference price as the generic price plus a fraction $(1 - \varrho)$ of the markup of the brand-name over the generic price ($p_r^\varrho = p_g^\varrho + (1 - \varrho)(p_b^\varrho - p_g^\varrho)$).

the benchmark case of no regulation.

Note two important characteristics of this reference price: First, being a convex combination of the two drugs' market prices, this reference price implies that the generic drug is available without any additional co-payment, whereas for the brand-name drug an additional co-payment applies. Thus, the reference price system can be considered to impose an additional, but avoidable co-payment (López-Casasnova & Puig-Junoy, 2000). In this model, consumers will have to trade off the additional co-payment against the (perceived) loss in quality, as the generic drug is associated with lower quality.

Second, in this model, the reference price is determined endogenously⁹, which involves a reaction of the reference price to the firms' strategic response to the introduction of a reference price system¹⁰. This is also in line with the design of the reference price system in Germany, where the reference price is set in such a manner that one third of the drugs is available at the reference price (§ 35 German Social Security Code V).

Patient co-payments are given as

$$c_b^{\varrho} = \kappa p_r^{\varrho} + (p_b^{\varrho} - p_r^{\varrho}) \text{ and } c_g^{\varrho} = \kappa p_g^{\varrho}. \quad (2.14)$$

An increase of the reference price has two effects on the total co-payment for the brand-name drug: On the one hand, the co-payment determined by the coinsurance rate and the reference price (κp_R) increases, on the other the difference between the sales price and the reference price ($p_b^{RP} - p_R$) decreases for given market prices, as the reference price can be considered a subsidy (Mestre-Ferrándiz, 2003).

Demand functions are given as

$$q_b^{\varrho} = 1 - \frac{((1 - \varrho)\kappa + \varrho)(p_b^{\varrho} - p_g^{\varrho})}{s\tau} \text{ and } q_g^{\varrho} = \frac{((1 - \varrho)\kappa + \varrho)(p_b^{\varrho} - p_g^{\varrho})}{s\tau} - \frac{\kappa p_g^{\varrho}}{s(1 - \tau)}. \quad (2.15)$$

⁹See Brekke, Holmas & Straume 2008 for the different implications of exogenously and endogenously determined reference prices.

¹⁰Strictly speaking, the introduction of a reference price system causes firms to lower their prices, which in turn then decreases the reference price. This reaction then results in firms lowering their prices again.

First-order conditions for profit-maximizing drug prices yield the best-response functions

$$p_b^{\varrho} = \frac{1}{2} \frac{s\tau + p_g^{\varrho}((1-\varrho)\kappa + \varrho)}{((1-\varrho)\kappa + \varrho)} \text{ and } p_g^{\varrho} = \frac{p_b^{\varrho}(1-\tau)((1-\varrho)\kappa + \varrho)}{2(\kappa + \varrho(1-\tau)(1-\kappa))}.$$

Resulting equilibrium prices are

$$\begin{aligned} p_b^{\varrho} &= \frac{2s\tau(\kappa + \varrho(1-\tau)(1-\kappa))}{(\kappa + \varrho(1-\kappa))(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} \\ \text{and } p_g^{\varrho} &= \frac{s\tau(1-\tau)}{\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)}, \end{aligned} \quad (2.16)$$

which are both lower than in the no regulation case:

$$\begin{aligned} p_b - p_b^{\varrho} &= \frac{2s\tau\varrho(1-\kappa)(\kappa(3+\tau^2) + 3\varrho(1-\tau)(1-\kappa))}{\kappa(3+\tau)(\kappa + \varrho(1-\kappa))(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} > 0 \\ p_g - p_g^{\varrho} &= \frac{3s\tau\varrho(1-\kappa)(1-\tau)^2}{\kappa(3+\tau)(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} > 0. \end{aligned}$$

The brand-name producer has to lower his price in order to counterbalance the effect of the increased co-payment. The generic producer then lowers his price in response to the lower brand-name price as well. Price elasticity of demand increases for both drugs:

$$\begin{aligned} \eta_b^{\varrho} &= -\frac{(1-\varrho)\kappa + \varrho p_b^{\varrho}}{s\tau} \frac{p_b^{\varrho}}{q_b^{\varrho}} > -\frac{\kappa p_b^{\varrho}}{s\tau} \frac{p_b^{\varrho}}{q_b^{\varrho}} = \eta_b, \\ \eta_g^{\varrho} &= -\frac{\kappa + \varrho(1-\tau)(1-\kappa)}{s\tau(1-\tau)} \frac{p_g^{\varrho}}{q_g^{\varrho}} > -\frac{\kappa}{s\tau(1-\tau)} \frac{p_g^{\varrho}}{q_g^{\varrho}} = \eta_g. \end{aligned}$$

Indifferent consumers are located at

$$\begin{aligned} \theta^{*\varrho} &= \frac{(\kappa + \varrho(1-\kappa)) + \tau(\kappa - \varrho(1-\kappa))}{(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} \\ \text{and } \underline{\theta}^{\varrho} &= \frac{\kappa\tau}{(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))}, \end{aligned}$$

and equilibrium quantities are given as

$$q_b^e = \frac{2(\kappa + \varrho(1-\tau)(1-\kappa))}{\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)} \text{ and } q_g^e = \frac{(\kappa + \varrho(1-\tau)(1-\kappa))}{\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)}. \quad (2.17)$$

Both firms' quantities are higher than under no regulation:

$$\begin{aligned} q_b - q_b^e &= -\frac{2\tau\varrho(1-\kappa)(1-\tau)}{(3+\tau)(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} < 0, \\ q_g - q_g^e &= -\frac{\tau\varrho(1-\kappa)(1-\tau)}{(3+\tau)(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} < 0. \end{aligned}$$

The brand-name drug producer lowers his price by more than what would be needed to sell the same quantity as under no regulation. The generic producer attracts more consumers from lowering his price than he loses to the brand-name producer:

$$\begin{aligned} \theta^* - \theta^{*e} &= \frac{2\tau\varrho(1-\kappa)(1-\tau)}{(3+\tau)(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))} \\ &< \frac{\underline{\theta} - \underline{\theta}^e}{(3+\tau)(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))}. \end{aligned}$$

The reference price is given as

$$\begin{aligned} p_r^e &= \varrho p_g^e + (1-\varrho)p_b^e \\ &= \frac{s\tau((2-\varrho)(\kappa + \varrho(1-\kappa)) - (2 - (\kappa + \varrho(1-\kappa)))\tau\varrho)}{(\kappa + \varrho(1-\kappa))(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))}. \end{aligned} \quad (2.18)$$

Firms' profits are

$$\begin{aligned} \Pi_b^e &= \frac{4s\tau(\kappa + \varrho(1-\tau)(1-\kappa))^2}{(\kappa + \varrho(1-\kappa))(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))^2} \\ \text{and } \Pi_g^e &= \frac{s\tau(1-\tau)(\kappa + \varrho(1-\tau)(1-\kappa))}{(\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa))^2}. \end{aligned} \quad (2.19)$$

Both firms' profits are lower under the reference price system, as the profit-decreasing effect of lower prices dominates the profit-increasing effect of higher quantities.

2.4 Policy Objectives in Regulation

The two regulatory instruments analyzed in this paper, direct price control and reimbursement regulation, are policy alternatives that are mutually exclusive¹¹. The choice of a regulatory instrument is determined by the objectives of regulators (Maynard & Bloor, 2003). For example, the intention to reduce public expenditure, to improve access to pharmaceuticals, and to stimulate competition are commonly articulated by policymakers. Whether an instrument is able to achieve these objectives and to outperform the alternatives, determines which instrument is chosen from a set of alternatives.

Two main factors drive the performance of the two regulatory systems with respect to the health policy objectives: Price reductions and the reimbursement amount. For a given reimbursement amount, higher price reductions result in lower public pharmaceutical expenditure, lower financial exposure of patients and accordingly better access to pharmaceuticals. A lower reimbursement amount at given price reductions leads to lower health expenditure, but higher financial exposure of patients and worse access to pharmaceuticals.

2.4.1 Price Reductions

When comparing price reductions under both regulatory systems, two concepts have to be distinguished: The *regulatory degree* determines what percentage of the markup of the brand-name over the generic is included in the price cap and/or in the reimbursement limit under the reference price system respectively. The *regulatory price* specifies the price cap or reimbursement limit that is realized eventually. Under the maximum price system, these two concepts are equivalent: If e.g. the regulatory body allows the brand-name producer to charge 50 % of the markup over the generic, then this corresponds to setting a price cap including 50 % of this markup. Under the reference price system, the endogenous specification of the reference price (the reference price is not defined in terms of unregulated prices, but is instead a function of (current) market prices) results in a mismatch between the two concepts: If e.g. the regulatory

¹¹By definition, the reference price which allows free pricing, is not compatible with a price cap.

body sets a reimbursement limit covering 50 % of the markup of brand-name over the generic¹², then an exogenous specification would result in a reference price including 50 % of this markup, but the endogenous specification yields a lower reference price, as it captures the strategic price decrease of both firms following the introduction of a reimbursement limit. The introduction of a reference price system causes firms to lower their prices, which in turn then decreases the reference price and consequently, the realized reference price is lower than what is specified by the degree of regulation. The endogenous specification of the reference price is an important factor driving price reductions under the reference price system: Starting from a certain regulatory degree it increases the strictness of regulation by generating a lower regulatory price.

In addition, two other factors determine relative price reductions under the two regulatory systems: First, the reference price generates a higher price elasticity of demand for the brand-name drug and thus a higher incentive for the brand-name producer to lower the price. Second, opposed to this, the regulatory price corresponds to the price of the brand-name drug under the maximum price system per definition, whereas under the reference price system, free pricing enables the brand-name producer to set a price above the regulatory price to skim off additional willingness to pay¹³. Based on an identical regulatory degree, the higher price elasticity and the dynamics of the endogenous specification of the reference price result in higher price reductions for the brand-name under reference pricing, whereas, for a given regulatory price, price reductions are higher under the maximum price system due to free pricing under the reference price system.

Figure 2.1 illustrates the comparison of brand-name prices under both regulatory systems for a low degree of product differentiation ($\tau = 0.1$) and $\kappa = 0.1$.

Let $\psi(m)$ denote the regulatory degree-equivalence line, all combinations of ϱ and μ that

¹²Note that a specification of the reference price as a function of the brand-name and generic price weighted by $(1 - \varrho)$ and ϱ respectively is equivalent to a specification as the generic price plus a fraction $(1 - \varrho)$ of the markup of the brand-name over the generic: $p_r^e = \varrho p_g^e + (1 - \varrho)p_b^e = p_g^e + (1 - \varrho)(p_b^e - p_g^e)$.

¹³Note that also the endogenous specification of the reference price prevents the brand-name producer from setting a price equal or below the reference price. But also for an exogenously determined reference price the brand-name price above the reference price is profit maximizing, as it allows the brand-name producer to skim off additional willingness to pay.

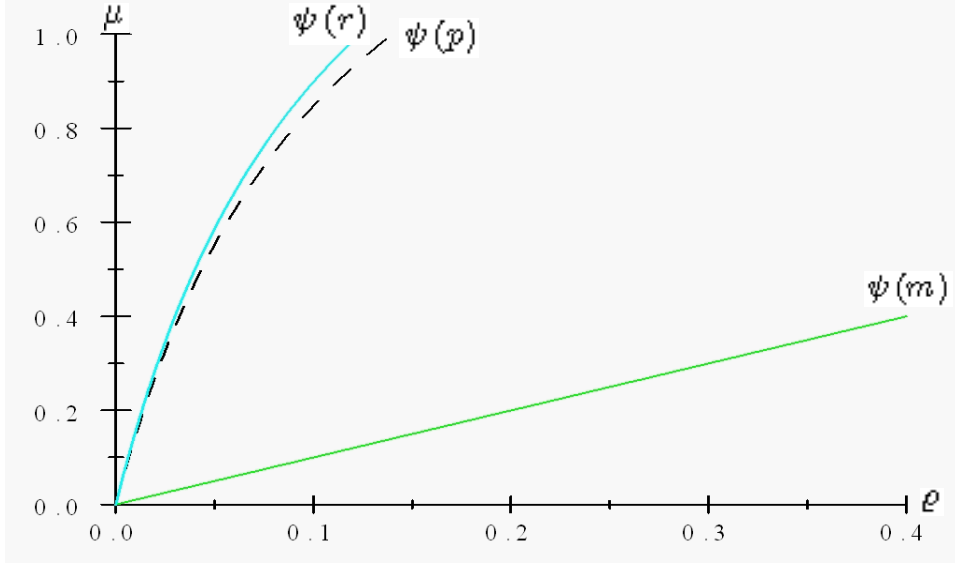


Figure 2.1: Brand-name Price, $\tau = 0.1$

result in identical percentages of the markup of the brand-name over the generic included in the price cap under the maximum price system and in the reimbursement limit under the reference price system respectively. This holds for $\varrho = \mu$. Let $\psi(r)$ denote the regulatory price-equivalence line, all combinations of ϱ and μ that result in the same price set by the regulatory authority, see Appendix A.4 for the specification of $\psi(r)$. That is, along this line, the price cap under the maximum price system is identical to the reimbursement limit under the reference price system, the market intervention is based on the same regulatory price. The slope of $\psi(r)$ is greater than 1 due to the endogenous specification of the reference price, which captures the firms' reaction to the reference price, i.e. the price decrease. This increases the strictness of regulation beyond the measure specified by the regulatory degree. For a higher ϱ , the slope of $\psi(r)$ becomes smaller. The price for the brand-name drug decreases by more than the price of the generic for a higher ϱ ($\frac{\partial p_b^{\varrho}}{\partial \varrho} < 0$), but is included in the reference price to a lower extent (a higher ϱ gives less weight to the brand-name relative to the generic). Compared to the endogenous specification of the reference price, the higher weight of the lower-priced generic in the reference price becomes a relatively more important factor in decreasing the reference price (but has a weaker impact

than the aforementioned). Note that the regulatory degree-equivalence line $\psi(m)$ depicts both factors determining price reductions under the reference price system – the increase of strictness due to the endogenous specification of the reference price and free price setting, whereas the regulatory price-equivalence line $\psi(r)$ – by ignoring the effect from the endogenous specification of the reference price – focuses on the effect of free price setting.

Let $\psi(p)$ denote the price-equivalence line, which represents all combinations of regulatory parameters that result in identical prices for the brand-name drug, see Appendix A.4 for specification of $\psi(p)$. To the left of $\psi(p)$, price reductions are higher under the maximum price system, to the right of $\psi(p)$, price reductions are higher under the reference price system. The regulatory degree-equivalence line $\psi(m)$ runs to the right of the price-equivalence line $\psi(p)$, i.e. for identical percentages of the markup of the brand-name over the generic permitted under the two regulatory systems, due to the endogenous specification of the reference price, price reductions are higher under the reference price system. As the effect of endogenous specification of the reference price is weaker for a high ϱ and accordingly the price of brand-name drug decreases less, the difference between $\psi(m)$ and $\psi(p)$ becomes smaller for a high ϱ . The regulatory price-equivalence line $\psi(r)$ runs to the left of the price-equivalence line $\psi(p)$, i.e. for identical regulatory prices, price reductions are higher under the maximum price system, as under the maximum price system, the regulatory price corresponds to the price for the brand-name drug, whereas under the reference price system, the brand-name producer sets a price above the reference price. For a high ϱ , the difference becomes larger. Under the maximum price system, a decrease in the regulatory price is equivalent to a (commensurate) decrease in the price for the brand-name drug. Under the reference price system, a decrease of the regulatory price results in a less than proportional price reduction for the brand-name drug.

If the brand-name drug and the generic are more remote substitutes, i.e. the degree of product differentiation is higher, also the willingness to pay for the brand-name drug is higher. Consequently, the brand-name producer sets the price further above the reference price. This additional incentive also weakens the effect from the endogenous specification of the reference

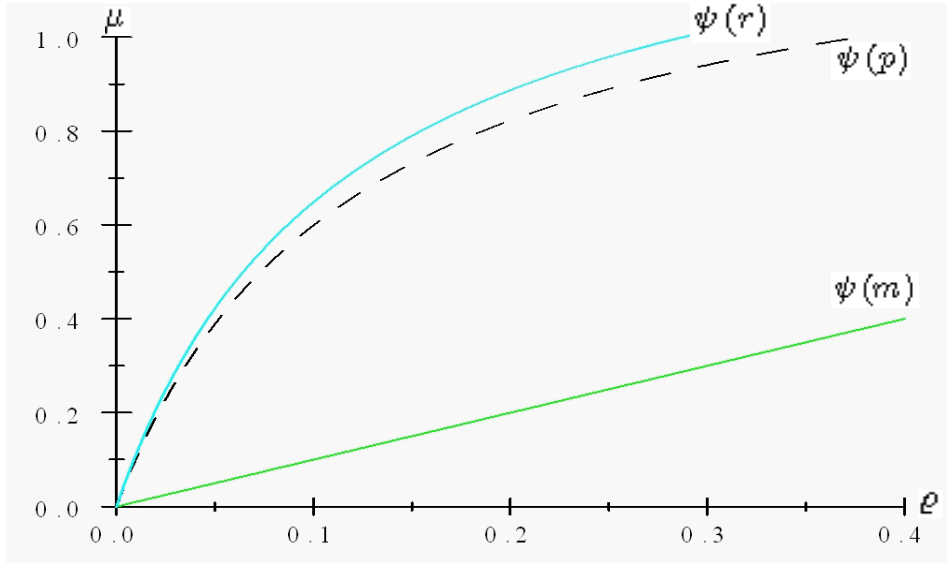


Figure 2.2: Brand-name Price, $\tau = 0.5$

price. Consider Figure 2.2 for a visualization of the comparison of brand-name prices for an intermediate degree of product differentiation ($\tau = 0.5$) and $\kappa = 0.1$. Due to the aforementioned incentive to increase the price the distance between $\psi(m)$ and $\psi(p)$ is smaller, while the distance between $\psi(r)$ and $\psi(p)$ increases.

Further increases in the degree of product differentiation increase this effect, see Appendix A.4 for a comparison of brand-name prices for $\tau = 0.9$.

In the following, the comparison of the two regulatory systems with respect to the health policy objectives will be based on regulatory prices only, the regulatory degree will be ignored.

2.4.2 Expenditure Reduction

Following from the government's role in funding health expenditure the primary objective in regulation is controlling public spending on pharmaceuticals (Danzon, 1997).

If the pharmaceutical market remains unregulated, public pharmaceutical expenditure is

given as the reimbursed fraction $(1 - \kappa)$ of total expenditure (unregulated price \times quantity):

$$E = (1 - \kappa) (p_b q_b + p_g q_g). \quad (2.20)$$

Under the maximum price system, the lower prices of both drugs reduce expenditure to

$$E^\mu = (1 - \kappa) (p_b^\mu q_b^\mu + p_g^\mu q_g^\mu), \quad E - E^\mu > 0, \quad (2.21)$$

and under the reference price system, both lower prices and a lower basis for reimbursement (not the market price, but the reference price is the basis for reimbursement) contribute to lower expenditure of:

$$E^e = (1 - \kappa) (p_r^e q_b^e + p_r^e q_g^e), \quad E - E^e > 0. \quad (2.22)$$

Thus, both regulatory instruments succeed in reducing pharmaceutical expenditure, see Appendix A.4 for details.

For a direct comparison of the two regulatory instruments with respect to their performance in expenditure reduction, two factors are crucial: Price reductions and the reimbursement amount. Whether price reductions are higher under the maximum price system or the reference price system, depends on the standard of comparison¹⁴. The reimbursement amount is lower under the reference price system (the brand-name drug is reimbursed based on reference price instead of the higher market price). The latter effect dominates and independent of the standard of comparison, expenditure is lower under the reference price system. This also implies that for given prices, expenditure is lower under the reference price system¹⁵.

Consider Figure 2.3 for a visualization of the comparison of expenditure under both systems for a low degree of product differentiation ($\tau = 0.1$) and $\kappa = 0.1$. Let E^* denote the expenditure-

¹⁴Based on the regulatory degree, price reductions are higher under the reference price system; whereas based on the regulatory price, price reductions are higher under the maximum price system.

¹⁵Two other factors point in the opposite direction: the reference price system exhibits a larger generic share and a lower brand-name premium, that is, a larger generic volume at a higher relative price is included in the expenditure under the reference price system. But these latter effects are dominated by the expenditure-reducing effect from a lower reimbursement amount.

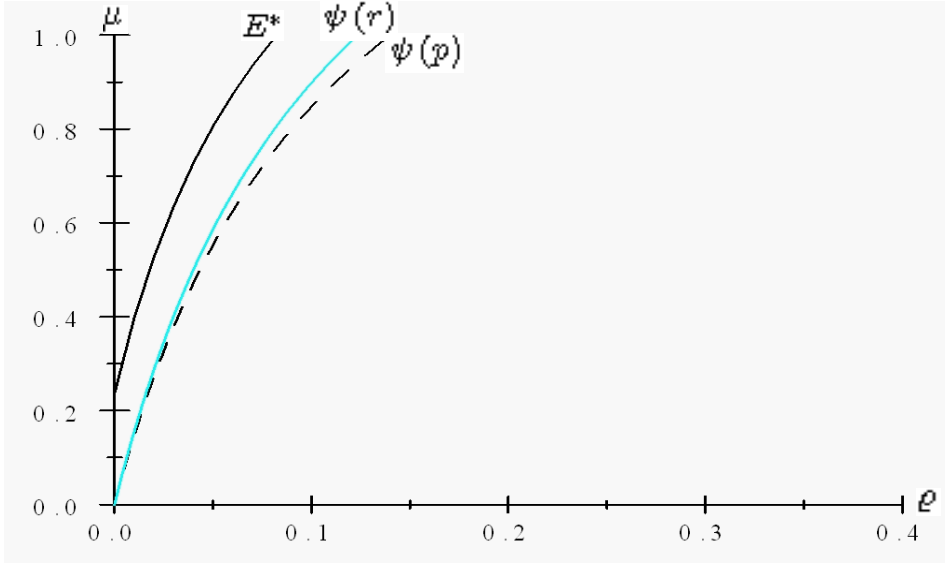


Figure 2.3: Expenditure, $\tau = 0.1$

equivalence line, i.e. all combinations of the regulatory parameters ϱ and μ that result in identical expenditure under the maximum price and the reference price system, see Appendix A.4 for the specification of E^* . To the right of this line, expenditure is higher under the maximum price system, to its left, expenditure is higher under the reference price system. The slope of the expenditure-equivalence line is greater than 1. This implies that a given reduction in expenditure, i.e. an upwards move on the expenditure-equivalence line, requires a greater increase in μ under the maximum price system than the increase in ϱ it requires under the reference price system. In other words, based on the same point of initial expenditure, an identical decrease of expenditure under the maximum price system and the reference price system is associated with a higher decrease of the price ceiling under the maximum price system than with the decrease of the reimbursement amount under the reference price system. On the contrary, a small decrease of ϱ , i.e. a small raise of the reimbursement limit, already causes the same increase in expenditure than a higher increase in μ under the maximum price system would yield.

Let $\psi(r)$ denote the regulatory price-equivalence line, and $\psi(p)$ the price-equivalence line.

Comparing expenditures under both systems for identical regulatory prices, i.e. an identical price cap under the maximum price system and reimbursement limit under the reference price system, the reference price system reduces expenditure to a greater extent because of a lower reimbursement amount despite higher price reductions under the maximum price system. This is illustrated by $\psi(r)$ running to the right of E^* .

For identical price reductions, expenditure is lower under the reference price system, as the reimbursement amount is lower. Put differently, for identical expenditure under the maximum price system and the reference price system, the reference price system exhibits higher market prices of brand-name drugs.

Note that the regulatory degree-equivalence line $\psi(m)$ and the regulatory price-equivalence line $\psi(r)$ depict both factors determining expenditure, the reimbursement amount and price reductions, whereas the price reduction-equivalence line $\psi(p)$ – by ignoring the effect from different price reductions under the two systems – focuses on the effect of a lower reimbursement amount.

Also with an increasing degree of product differentiation, the reference price system reduces expenditure to a larger extent than the maximum price system, see Figure 2.4 for an illustration. For identical price reductions, the relative advantage of the reference price system in reducing expenditure is higher (the distance between $\psi(p)$ and E^* is higher). That is, for higher degrees of product differentiation the relative importance of price reductions decreases, whereas the effect from a lower reimbursement amount increases.

Further increases of the degree of product differentiation increase this effect, see Appendix A.4 for a comparison of expenditure for $\tau = 0.9$.

Proposition 2.1 summarizes the performance of both regulatory instruments with respect to the reduction of expenditure:

Proposition 2.1. *Suppose that price reductions for the brand-name drug are identical under the maximum price system and the reference price system. Then public pharmaceutical expenditure is lower under the reference price system.*

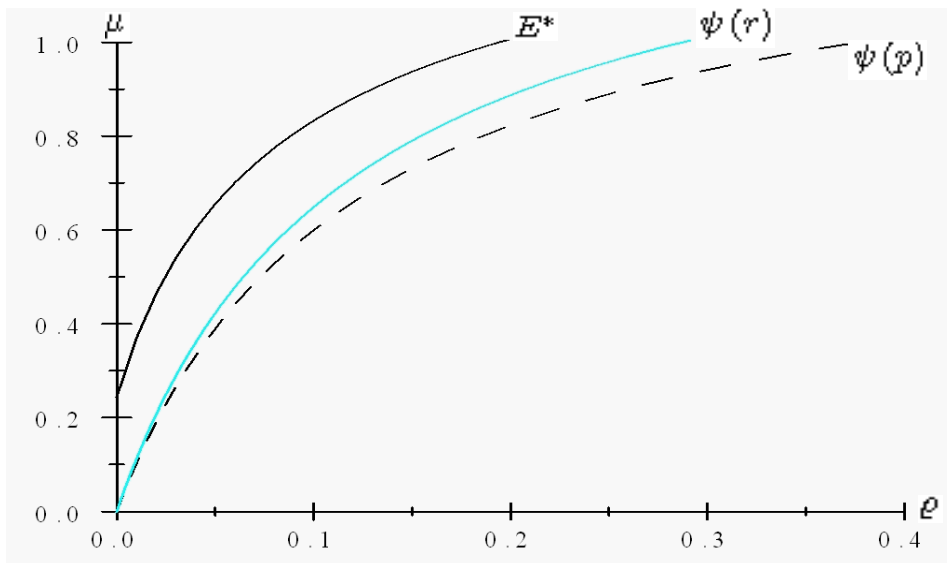


Figure 2.4: Expenditure, $\tau = 0.5$

2.4.3 Equity and Access

Equity, the concept of fairness and justice, is one of the major concerns in health policy (Hurley, 2003). Two dimensions of equity are relevant when comparing the two regulatory instruments: financial exposure and access. First, the analysis of out-of-pocket expenditure under the different scenarios illustrates financial exposure of patients. Under the maximum price system, co-payments rules do not change; consequently, consumers benefit fully from lower prices. The reference price system, however, introduces an additional co-payment element, patients also have to pay the difference between the market price and reference price. Therefore, out-of-pocket expenditure under the reference price system needs to be compared carefully with payments under no regulation. Second, the analysis of quantities and the uncovered part of the market (consumers with a low valuation θ) gives an idea of access to pharmaceuticals. In general, lower drug prices improve access, as also consumers with lower valuation can now afford the generic. Third, the concept of consumer surplus as a measure for well-being in the aggregate combines the aspects of financial exposure and access.

Co-payments

If the market is unregulated, co-payments for the brand-name drug and the generic, respectively, are:

$$c_b = \kappa p_b = \frac{2s\tau}{3 + \tau} \text{ and } c_g = \kappa p_g = \frac{s\tau(1 - \tau)}{3 + \tau}. \quad (2.23)$$

Under the maximum price system, lower drug prices reduce co-payments to

$$c_b^\mu = \frac{s\tau(2 - \mu(1 + \tau))}{(3 + \tau)} \text{ and } c_g^\mu = \frac{s\tau(2 - \mu(1 + \tau))(1 - \tau)}{2(3 + \tau)}, \quad (2.24)$$

$$c_b - c_b^\mu > 0 \text{ and } c_g - c_g^\mu > 0.$$

Co-payments for the brand-name drug and the generic, respectively, under the reference price system are given as:

$$c_b^\rho = \kappa p_r^\rho + p_b^\rho - p_r^\rho = \frac{s\tau(2\kappa + \rho(1 - \tau)(1 - \kappa))}{(\kappa(3 + \tau) + 3\rho(1 - \tau)(1 - \kappa))} \quad (2.25)$$

$$\text{and } c_g^\rho = \frac{\kappa s\tau(1 - \tau)}{\kappa(3 + \tau) + 3\rho(1 - \tau)(1 - \kappa)}, \quad (2.26)$$

$$c_b - c_b^\rho > 0 \text{ and } c_g - c_g^\rho > 0.$$

which are both lower than under no regulation. The co-payment-decreasing effect of a lower brand-name price dominates the co-payment-increasing effect of a lower reimbursement amount (the reference price instead of the market price is the basis for reimbursement). The co-payment for the generic is unambiguously lower under the reference price system, as the drug price is lower and co-payment rules do not change.

When comparing co-payments for the brand-name drug directly, two factors determine, whether out-of-pocket expenditure is higher under the maximum price or the reference price system: First, for identical regulatory prices, price reduction are higher under the maximum price system. The application of coinsurance rates implies that lower drug prices translate to lower co-payments. Second, in the reference price system health insurance reimburses the brand-

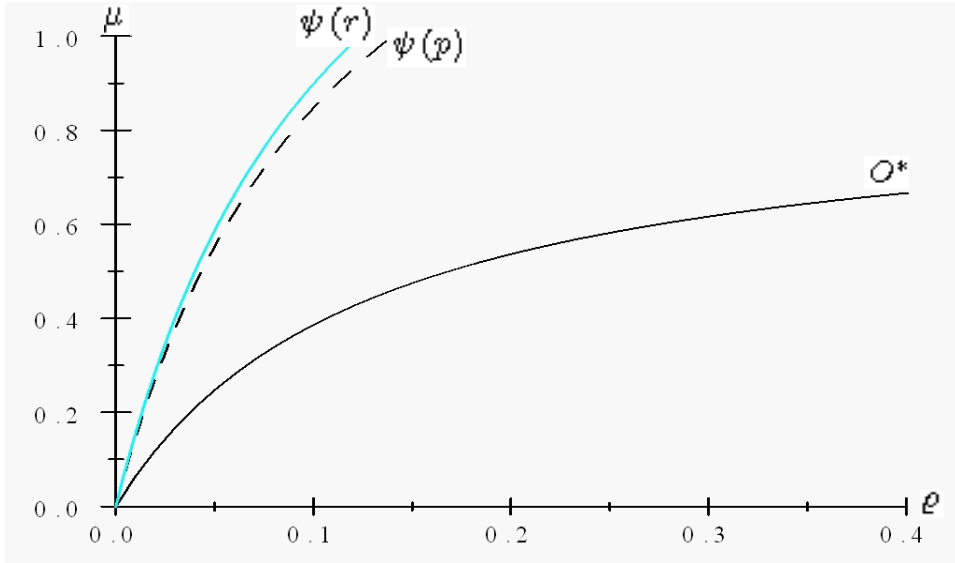


Figure 2.5: Brand-name Co-payment, $\tau = 0.1$

name drug based on the reference price. This involves an additional co-payment element – the difference between the market price of the brand-name drug and the reference price. Both factors result in a higher co-payment for the brand-name drug under the reference price system, as illustrated by Figure 2.5 for $\tau = 0.1$. Let O^* denote the out-of-pocket expenditure-equivalence line, i.e. all combinations of the regulatory parameters ρ and μ that give identical co-payments for the brand-name drug under the maximum price and the reference price system, see Appendix A.4 for a specification. To the right of this line, out-of-pocket expenditure is higher under the maximum price system, to its left, out-of-pocket expenditure is higher under the reference price system. The regulatory price-equivalence line $\psi(r)$ runs to the left of the out-of-pocket expenditure-equivalence line O^* .

That is, the both higher price reductions under the maximum price system and a lower reimbursement amount under the reference price system result in higher co-payments under the reference price system. The price reduction-equivalence line $\psi(p)$ visualizes the isolated effect from a changed reimbursement amount. A lower reimbursement amount makes the co-payment for the brand-name drug always higher under the reference price system. This is illustrated by

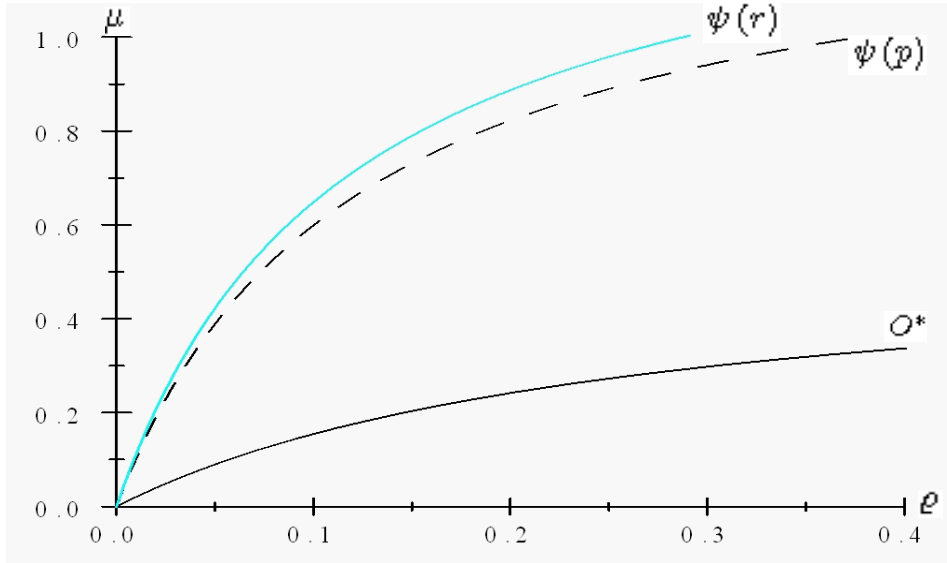


Figure 2.6: Brand-name Co-payment, $\tau = 0.5$

the price reduction-equivalence line $\psi(p)$ running to the left of the out-of-pocket expenditure-equivalence line O^* .

Also for an intermediate degrees of product differentiation, the co-payment for the brand-name drug is higher under the reference price system, see Figure 2.6. Similar to the comparison of relative performance in decreasing expenditure, the effect of a lower reimbursement amount becomes more important (the distance between $\psi(p)$ and O^* is higher) and the effect of price reductions is relatively less important for a higher degree of product differentiation.

Further increases of the degree of product differentiation increase this effect, see Appendix A.4 for a comparison of co-payments for $\tau = 0.9$.

With respect to the generic, only the magnitude of price reductions determines, whether co-payments are higher under the maximum price or the reference price system. Under both the maximum price system and the reference price system, the generic is reimbursed based on its market price. Accordingly, the co-payments for the generic under both regulatory systems are directly proportional to the market price.

Consider Figure 2.7 for a visualization of the comparison of co-payments for the generic under

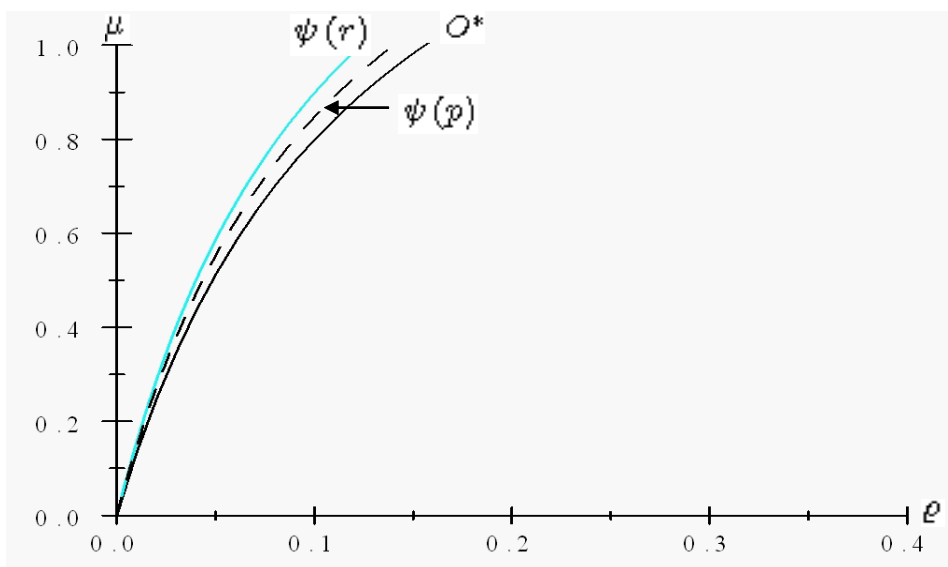


Figure 2.7: Generic Co-payment, $\tau = 0.1$

both systems for a low degree of product differentiation ($\tau = 0.1$) and $\kappa = 0.1$. Let O^* denote the out-of-pocket expenditure-equivalence line, i.e. all combinations of the regulatory parameters ρ and μ that give identical co-payments for the generic under the maximum price and the reference price system, see Appendix A.4 for specification of O^* . To the right of this line, co-payments for the generic are higher under the maximum price system, to its left, co-payments for the generic are higher under the reference price system. Since co-payments are directly proportional to market prices, O^* also represents all combinations of regulatory parameters that result in identical prices for the generic under both systems.

For identical regulatory prices, free pricing causes the price for the brand-name drug to be higher under the reference price system. Under the maximum price system, the generic producer prices at a higher discount from a given brand-name price than under the reference price system¹⁶. Conversely, this implies that the brand-name premium is lower under the reference

¹⁶Note that under the maximum price system, the best response function of the generic producer is given as $p_g^\mu = \frac{1}{2}p_b^\mu(1 - \tau)$, whereas under the reference price system, the best response function of the generic producer is given as $p_g^\mu = \frac{1}{2}p_b^\rho(1 - \tau) \frac{((1-\rho)\kappa+\rho)}{(\kappa+\rho(1-\tau)(1-\kappa))} , \frac{((1-\rho)\kappa+\rho)}{(\kappa+\rho(1-\tau)(1-\kappa))} > 1$.

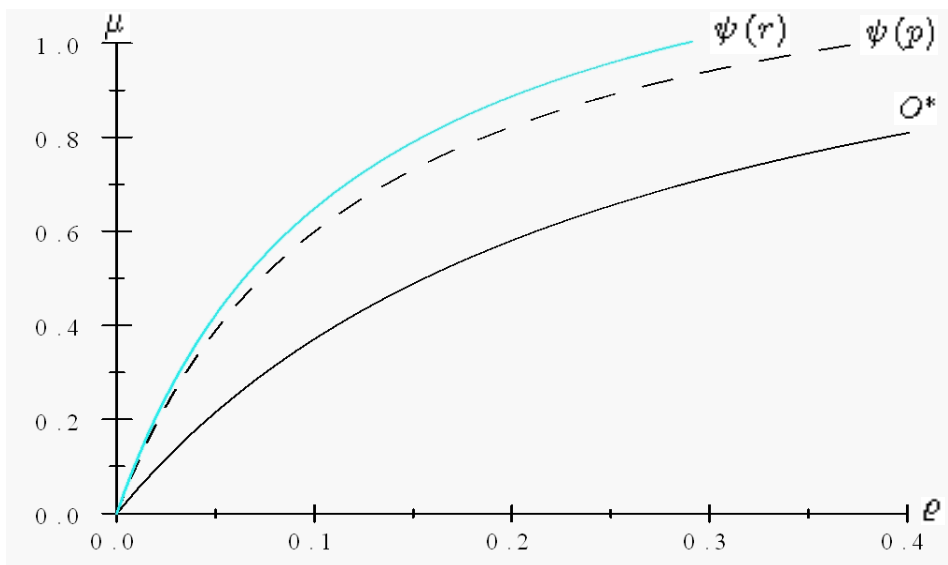


Figure 2.8: Generic Co-payment, $\tau = 0.5$

price system. That is, for identical regulatory prices, the difference between prices under the maximum price system and the reference price system is higher for generic prices as compared to brand-name prices. This is illustrated by O^* running to the right of $\psi(p)$. This implies that for identical regulatory prices and for identical prices of the brand-name drug, prices for the generic are lower under the maximum price system, and, consequently, co-payments for the generic are lower under the maximum price system.

For a higher degree of product differentiation, the discount of the generic price from the brand-name price increases more under the maximum price system than under the reference price system. This is illustrated by the difference between $\psi(p)$ and O^* becoming larger, as visualized in Figure 2.8. This implies that the relative advantage of the maximum price system in reducing out-of-pocket expenditure for generic users is higher for a higher degree of product differentiation.

That is, co-payments for both drugs are lower under the maximum price system. In other words, the financial exposure of patients to costs of pharmaceuticals is lower under the maximum price system.

Total Quantity (Access)

If the market is unregulated, the total quantity of both versions of the drug is

$$Q = \frac{3}{3 + \tau}. \quad (2.27)$$

Under the maximum price system, the effect from a higher brand-name sales volume exceeds the effect of a lower generic quantity so that the quantity of the drug increases to

$$Q^\mu = \frac{6 + \tau\mu(1 + \tau)}{2(3 + \tau)}, \quad Q - Q^\mu = -\frac{\tau\mu(1 + \tau)}{2(3 + \tau)} < 0. \quad (2.28)$$

Under the reference price system, the total sales volume of both versions of the drug is

$$\begin{aligned} Q^e &= \frac{3(\kappa + \varrho(1 - \tau)(1 - \kappa))}{\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)}, \\ Q - Q^e &= -\frac{3\tau\varrho(1 - \tau)(1 - \kappa)}{(\tau + 3)(\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa))} < 0. \end{aligned} \quad (2.29)$$

Compared to the unregulated market, more is sold of both the brand-name drug and the generic. Correspondingly, the total quantity is higher under the reference price system. Thus, as the quantity increases under both the maximum price and the reference price system, both regulatory instruments can be considered to improve access to pharmaceuticals.

When compared directly, three factors determine whether total quantity is higher under the maximum price or the reference price system. First, for identical regulatory prices, price reductions are higher under the maximum price system. Lower prices imply that more consumers are able and willing to buy the drug, hence higher price reductions translate to a higher quantity. Second, for identical price reductions, the quantity of the brand-name drug is higher under the maximum price system, since from a consumer perspective, less has to be paid for the drug under the maximum price system. Third, the generic quantity is higher under the reference price system. The first two effects exceed the latter and total quantity is higher under the maximum system.

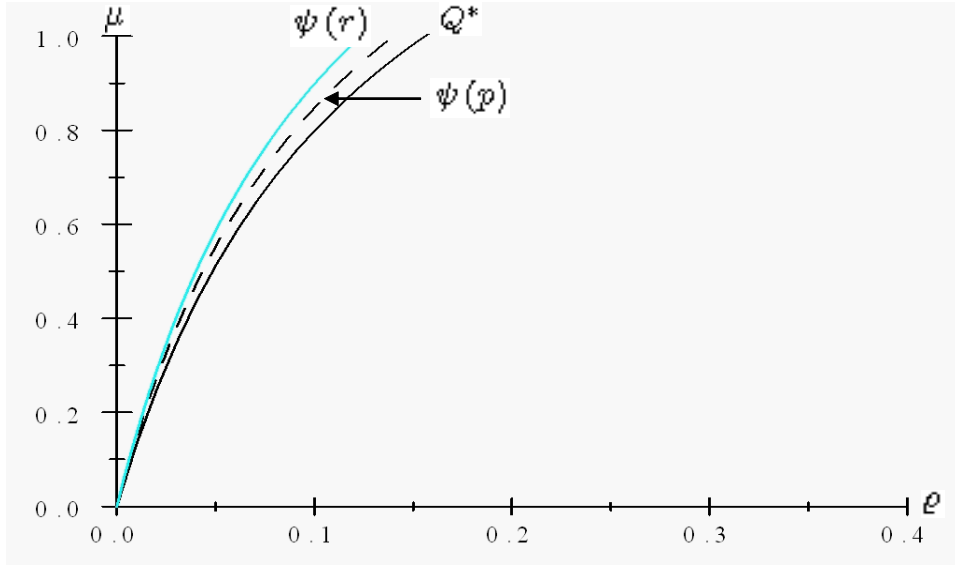


Figure 2.9: Total Quantity, $\tau = 0.1$

Consider Figure 2.9 for a visualization for $\tau = 0.1$. Let Q^* denote the quantity-equivalence line, i.e. all combinations of the regulatory parameters ρ and μ that give identical total sales volumes of both versions of the drug under the maximum price and the reference price system, see Appendix A.4 for specification of Q^* . To the right of this line, the total quantity of the drug is higher under the reference price system, to its left, the total sales volume is higher under the maximum price system. Q^* runs to the right of both $\psi(r)$ and $\psi(p)$, i.e. for identical regulatory prices and for identical prices of the brand-name, total quantity is higher under the maximum price system.

This also holds for a higher degree of product differentiation, see Appendix A.4.

Consumer Surplus

Consider consumer surplus as a measure for the well-being of consumers in the aggregate. If the market is not regulated, consumer surplus for the brand-name users is given as

$$CS_b = \int_{\theta^*}^1 (\theta s - \kappa p_b) d\theta = \frac{2s(2 - \tau)}{(\tau + 3)^2} \quad (2.30)$$

and for generic users as

$$CS_g = \int_{\underline{\theta}}^{\theta^*} (\theta(1 - \tau)s - \kappa p_g) d\theta = \frac{s(1 - \tau)}{2(\tau + 3)^2}. \quad (2.31)$$

Note that consumer surplus is higher for users of the brand-name drug. Both consumer surpluses increase in s , are independent of κ , and decrease in τ .

Under the maximum price system, consumer surplus for brand-name users is higher than in the benchmark case, as a larger quantity is consumed at a lower price:

$$CS_b^\mu = \int_{\theta^*}^1 (\theta s - \kappa p_b^\mu) d\theta, \quad CS_b - CS_b^\mu < 0. \quad (2.32)$$

Consumer surplus for generic users is lower than in the case of no regulation, as the effect of a lower quantity dominates the effect of a lower price on consumer surplus:

$$CS_g^\mu = \int_{\underline{\theta}}^{\theta^*} (\theta(1 - \tau)s - \kappa p_g^\mu) d\theta, \quad CS_g - CS_g^\mu > 0. \quad (2.33)$$

Thus, brand-name users benefit from the maximum price system, generic users lose from it.

Under the reference price system, consumer surplus for both brand-name users and generic users is higher than under no regulation, since drug prices are lower and higher quantities are consumed:

$$\begin{aligned} CS_b^e &= \int_{\theta^*}^1 (\theta s - \kappa p_R^e - (p_b^e - p_R^e)) d\theta, \quad CS_b - CS_b^e < 0, \\ CS_g^e &= \int_{\underline{\theta}}^{\theta^*} (\theta(1 - \tau)s - \kappa p_g) d\theta, \quad CS_g - CS_g^e < 0. \end{aligned} \quad (2.34)$$

Both groups of consumers benefit from the reference price system.

Consumer surplus for brand-name users is higher under the maximum price, as the co-payment for the brand-name drug is lower and the quantity is higher. For generic users, consumer surplus is higher under the reference price system, the co-payment for the generic is lower under

the maximum price system, but the quantity is higher under the reference price system.

Proposition 2.2 summarizes the performance of both regulatory instruments with respect to distributive objectives:

Proposition 2.2. *Suppose that price reductions of the brand-name drug are identical under the maximum price system and the reference price system. Then financial exposure of patients is lower and access to pharmaceuticals is higher under the maximum price system. Consumer surplus for brand-name users is higher under the maximum price system, consumer surplus for generic users is higher under the reference price system.*

2.4.4 Competition in Pharmaceutical Markets

The degree of competition between the brand-name producer and the generic producer is determined by the degree of product differentiation to a large extent, as it allows the brand-name producer to charge a higher price while maintaining a significant market share. The more remote substitutes the two versions of the drug are, the higher τ is, the more will prices diverge. In other words, decreasing the degree of product differentiation will stimulate competition between the two firms. For a given τ , regulation can also have an effect on competition. Against the background of the benchmark case of perfect competition, when $\tau = 0$ and both firms price at marginal cost, the analysis of competition between the two firms has several dimensions: First, the existence and extent of a brand-name price premium indicates whether both versions are considered and treated as close or remote substitutes. Note that both regulatory instruments assume equivalence of the brand-name and generic version. Second, the generic market share illustrates the extent to which the generic producer can prevail against the brand-name producer and generic competition occurs. Third, the relationship between profits and the degree of product differentiation gives an idea of incentives for firms to differentiate their products and to gain a competitive advantage. As higher product differentiation amounts to a lower degree of competition, this could be viewed as an anticompetitive behavior.

Brand-name Premium

If the market is not regulated, the brand-name price premium amounts to

$$\frac{p_b}{p_g} = \frac{2}{(1-\tau)}. \quad (2.35)$$

Under the maximum price system it is given as

$$\frac{p_b^\mu}{p_g^\mu} = \frac{2}{(1-\tau)}. \quad (2.36)$$

That is, under the maximum price system, price reductions for the brand-name drug are as high as for the generic drug.

Under the reference price system, the brand-name price premium amounts to

$$\frac{p_b^\varrho}{p_g^\varrho} = \frac{2}{(1-\tau)} \frac{(\kappa + \varrho(1-\tau)(1-\kappa))}{(\kappa + \varrho(1-\kappa))}, \quad (2.37)$$

which is lower than under no regulation, as $\frac{(\kappa + \varrho(1-\tau)(1-\kappa))}{(\kappa + \varrho(1-\kappa))} < 1$. Reference pricing reduces the brand-name price to a larger extent than the generic price. This is due to higher price elasticity of demand for the brand-name price.

Generic Market Share

In the benchmark case of no regulation, the generic market share is given as

$$\frac{q_g}{Q} = \frac{1}{3}. \quad (2.38)$$

Under the maximum price, the generic market share amounts to

$$\frac{q_g^\mu}{Q^\mu} = \frac{1 - \frac{\mu(1+\tau)}{2}}{3 + \frac{\mu\tau(1+\tau)}{2}}, \quad (2.39)$$

which is lower than under no regulation for $\mu > 0$. That is, by increasing the brand-name

quantity and decreasing the generic quantity, the maximum price system decreases the generic market share and accordingly, it weakens generic competition.

Under the reference price system, the generic market share is given as

$$\frac{q_g^e}{Q^e} = \frac{1}{3}, \quad (2.40)$$

which is as high as under no regulation. Thus, under the reference price system, brand-name and generic quantity increase by the same amount. Generic competition is not intensified under the reference price system.

Incentive for Product Differentiation

If the market is unregulated, brand-name and generic profit are given as

$$\Pi_b = \frac{4s\tau}{\kappa(\tau+3)^2} \text{ and } \Pi_g = \frac{s\tau(1-\tau)}{\kappa(\tau+3)^2}. \quad (2.41)$$

Brand-name profit increases with the degree of product differentiation

$$\frac{\partial \Pi_b}{\partial \tau} = \frac{4s(3-\tau)}{\kappa(\tau+3)^3} > 0.$$

Hence, there is an incentive for the brand-name producer to raise its profit by raising the degree of product differentiation. Generic profit increases in τ , if $\tau < \frac{3}{7}$ and decreases otherwise:

$$\frac{\partial \Pi_g}{\partial \tau} = \frac{s(3-7\tau)}{\kappa(\tau+3)^3} > 0 \text{ if } \tau < \frac{3}{7}.$$

This is, to some extent there is also incentive for the generic producer to raise τ . For a low degree of product differentiation, the positive strategic effect exceeds the negative direct effect. A certain degree of product differentiation allows the generic producer to attract additional consumers with a low valuation, but also forces it to lower its price.

Under the maximum price system, brand-name and generic profit are given as

$$\Pi_b^\mu = \frac{s\tau(2 - \mu(1 + \tau))(4 + \mu(\tau + 1))^2}{2\kappa(\tau + 3)^2} \text{ and } \Pi_g^\mu = \frac{s\tau(1 - \tau)(2 - \mu(1 + \tau))^2}{4\kappa(\tau + 3)^2}. \quad (2.42)$$

Both profits increase in τ for a low degree of regulation and decrease in τ for a high degree of regulation:

$$\frac{\partial \Pi_b^\mu}{\partial \tau} < 0, \text{ if } \mu > \mu_b^*.$$

$$\frac{\partial \Pi_g^\mu}{\partial \tau} < 0, \text{ if } \mu > \mu_g^*.$$

That is, for sufficiently strict regulation, there is no incentive to raise τ for both the brand-name producer and generic producer.

Under the reference price system, brand-name and generic profit are given as

$$\begin{aligned} \Pi_b^\varrho &= \frac{4s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))^2}{(\kappa + \varrho(1 - \kappa))(\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa))^2} \\ \text{and } \Pi_g^\varrho &= \frac{s\tau(1 - \tau)(\kappa + \varrho(1 - \tau)(1 - \kappa))}{(\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa))^2}. \end{aligned} \quad (2.43)$$

The relationship between profits and the degree of product differentiation is determined by the degree of regulation and the degree of product differentiation. The brand-name profit decreases in τ for a high degree of regulation and a high degree of product differentiation:

$$\frac{\partial \Pi_b^\varrho}{\partial \tau} < 0, \text{ if } \varrho > \varrho_b^*.$$

The generic profit decreases in τ for a high degree of regulation:

$$\frac{\partial \Pi_g^\varrho}{\partial \tau} < 0, \text{ if } \varrho > \varrho_g^*.$$

The reference price system can only reduce the incentive to raise τ for the brand-name producer, if τ is already very high and if regulation is sufficiently strict. In this case, there is also no

incentive to raise τ for the generic producer.

Proposition 2.3 summarizes the performance of both regulatory instruments with respect to stimulation competition:

Proposition 2.3. *The maximum price system does not change the brand-name price premium and results in a lower generic market share. The reference price system reduces the brand-name price premium and increases the generic quantity, but does not change the generic market share. Both regulatory instruments reduce the incentive for firms to increase product differentiation.*

2.5 Welfare Analysis

This section examines the welfare effects of the two regulatory instruments. Welfare is given as the sum of consumer surplus for the brand-name and generic users, respectively, and profits for the two firms net of public pharmaceutical expenditure:

$$W = CS_b + CS_g + \Pi_b + \Pi_g - E. \quad (2.44)$$

Both regulatory instruments increase welfare:

$$W - W^\mu < 0, \quad W - W^e < 0. \quad (2.45)$$

When comparing welfare effects for the two regulatory instruments directly, the performance with respect to the above-mentioned health policy objectives determines, whether welfare is higher under the maximum price or the reference price system: Consumer surplus for brand-name users is higher under the maximum price system, while consumer surplus for generic users is higher under the reference price system. Total profits are higher under the maximum price system, see Appendix A.5. Public pharmaceutical expenditure is lower under the reference price system. The effect of the maximum price system with respect to increasing consumer surplus for brand-name users and minimizing losses for firms exceeds the effect of the reference price

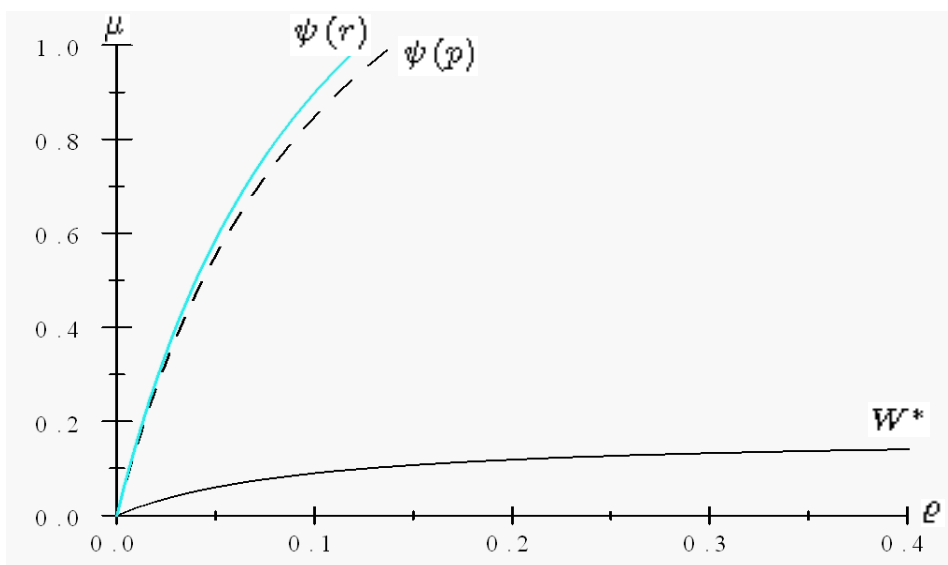


Figure 2.10: Welfare, $\tau = 0.1$

system with respect to increasing consumer surplus for generic users and decreasing public pharmaceutical expenditure. Total welfare is higher under the maximum price system.

Consider Figure 2.10 for a visualization of the comparison of welfare under both systems for a low degree of product differentiation ($\tau = 0.1$) and $\kappa = 0.1$. Let W^* denote the welfare-equivalence line, i.e. all combinations of the regulatory parameters μ and ρ that result in identical total welfare under the maximum price and the reference price system, see Appendix A.5 for the specification of W^* . To the right of this line, welfare is higher under the reference price system, to its left, welfare is higher under the maximum price system. W^* runs to the right of both $\psi(r)$ and $\psi(p)$, i.e. for identical regulatory prices and for identical prices of the brand-name, total welfare is higher under the maximum price system. This also holds for a higher degree of product differentiation, see Appendix A.5.

Taking marginal cost of raising public funds into account corresponds to giving a higher weight to public pharmaceutical expenditure. This shifts the result of the welfare comparison in favor of the reference price system, since the reference price system reduces public pharmaceutical expenditure to a larger extent. Proposition 2.4 summarizes the performance of both regulatory

instruments with respect to welfare:

Proposition 2.4. *Suppose that price reductions of the brand-name drug are identical under the maximum price system and the reference price system. Then welfare is higher under the maximum price system.*

2.6 Conclusion

In this model, both the maximum price and the reference price system result in a reduction of drug prices and pharmaceutical expenditure, as intended by regulators. Both instruments reduce financial exposure of patients and improve access to pharmaceuticals.

Under the reference price system, price reductions for the brand-name drug are driven by the endogenous specification of the reference price, which generates further price decreases by capturing the firms' strategic response to a reimbursement limit, and the pricing setting which enables the brand-name producer to skim off additional willingness to pay by setting a price above the reference price. The first factor generates higher price reductions under the reference price system for identical regulatory degrees, while the latter factor gives rise to lower price reductions as compared to the maximum price system for identical regulatory prices.

For identical price reductions of the brand-name drug, the lower reimbursement amount under the reference price system results in lower health expenditure, but higher financial exposure of patients. Access to pharmaceuticals is better under the maximum price system, although the generic quantity is higher under the reference price system. In the aggregate, consumer surplus for brand-name users is higher under the maximum price system, whereas consumer surplus for generic users is higher under the reference price system.

Whereas the maximum price system does not change the brand-name price premium and results in a lower generic market share, the reference price system reduces the brand-name price premium and brings about a higher generic quantity, but unchanged generic market share. Both regulatory instruments reduce the incentive for firms to gain a competitive advantage by

increasing product differentiation.

Consequently, there is a trade-off between important health policy objectives: The reference price system may be more appropriate to reduce public pharmaceutical expenditure or stimulate competition, but the maximum price system performs better for distributive objectives, such as limiting financial exposure of patients and guaranteeing access to pharmaceuticals.

In general, if both versions of the drug were considered equivalent and, accordingly, perfect substitutes, all health policy objectives could easily be achieved. Perfect competition would reduce public expenditure, minimize financial exposure of patients and maximize access to pharmaceuticals. Consequently, the main health policy challenge is to reduce the degree of product differentiation. With respect to objective product differentiation, this corresponds to reducing permitted bandwidths of equivalence of additives and the degree of bioavailability. In addition, information of the public and mandatory substitution as means for patients to gather experience with generics could help to reduce subjective product differentiation. However, it has to be considered that substitution is problematic in some classes of drugs, such as antiepileptics, as the optimal dose has to be determined at the individual patient level and divergent permitted degrees of bioavailability harm therapeutic success (Hopf, 2002). Adjustment costs or health costs may emerge, when regulation reduces compliance.

The long-term effects of pharmaceutical regulation are subject to further research. In Germany, the introduction of the reference price system in 1989 has reduced public pharmaceutical expenditure only in the following years. As a consequence, further regulatory instruments were added.

Furthermore, the impact of lower profits on the incentive to invest (in quality) or on market entry has to be studied further. If regulation reduces investments or inhibits entry, there is a trade-off between the static gains from cost containment on the one hand and potential dynamic losses from lower quality or reduced competition on the other hand.

Chapter 3

The Model

This chapter presents the model following Maskus & Chen (2002), Chen & Maskus (2005), which is used in the following chapters. In chapters 4 and 5, the link between cost-sharing and parallel trade is studied using this model. In chapter 6, a modification of the model is employed to study the effect of pharmaceutical regulation at the wholesale level under parallel trade.

3.1 Vertical Control Structure – Supply Side

Consider a domestic manufacturer M selling a brand-name drug b in two countries, in his home country D and a foreign country S .

In chapters 4 and 5, I assume that the manufacturer sells directly in his home country, but in the foreign country, it sells through an independent intermediary I . In chapter 6, I assume that in both countries, the manufacturer sells through an independent intermediary I_j ($j = D, S$)¹. With respect to the intermediaries, the manufacturer adopts a two-part pricing strategy, it charges each intermediary a wholesale price w_j per unit and a fixed fee ϕ_j .

Due to lack of complete vertical control, in a regime of international exhaustion of intellectual property rights, an intermediary may engage in parallel trade and resell the drug in the respective

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

other country. In chapters 4 and 5, the foreign intermediary takes advantage of this opportunity and resells the drug also in country D . In chapter 6, I assume that only the foreign intermediary I_S engages in parallel trade, but the domestic intermediary in country D , I_D does not (one-way parallel trade). Accordingly, the intermediary I_S exports the drug from country S and sells it in country D as a parallel import.

That is, the foreign country is the source country of the parallel import and the home country is the destination country. Therefore, the home country will be denoted as country D and the foreign country as country S . The sales of the drug as a parallel import are denoted by β .

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

3.2 Vertical Product Differentiation – Demand Side

In the source country, only the locally sourced version of the drug, supplied by the (foreign) intermediary is available. Consumers in the destination country D have the choice between the locally sourced version b and the parallel import β when buying from the (foreign) 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, 2000b). 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 their 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

$$U(\theta, \tau, c_i) = \begin{cases} \theta - c_{i,j} & \text{if } i = b \\ \theta\tau - c_{i,j} & \text{if } i = \beta \end{cases} \quad (3.1)$$

where $\tau \in [0, 1]$ reflects the perceived quality difference between both versions b and β of the drug and $c_{i,j}$ is the patient (price-dependent) co-payment for drug i in country j ($j = D, S$). 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 co-payment. 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, differences in the severity of the condition or the level of suffering or differences in prescription practices (see e.g. Brekke, Holmas & Straume, 2010).

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

$$\theta_D^{b,0} - c_{b,D} = 0 \Leftrightarrow \theta_D^{b,0} = c_{b,D}. \quad (3.2)$$

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

$$q_{b,D} = 1 - c_{b,D}. \quad (3.3)$$

If parallel trade is legal (international exhaustion of intellectual property rights), consumers in country D have the choice between the locally sourced version (b) or the parallel import (β). 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

$$\theta_D^{b,\beta} - c_{b,D} = \theta_D^{b,\beta} (1 - \tau) - c_{\beta,D} \Leftrightarrow \theta_D^{b,\beta} = \frac{c_{b,D} - c_{\beta,D}}{\tau}, \quad (3.4)$$

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

$$\theta_D^{\beta,0} (1 - \tau) - c_{\beta,D} = 0 \Leftrightarrow \theta_D^{\beta,0} = \frac{c_{\beta,D}}{(1 - \tau)}. \quad (3.5)$$

Consequently, in country D , if the parallel import is available, demand for the locally sourced version b and for the parallel import β is given by

$$q_{b,D}^* = 1 - \frac{c_{b,D} - c_{\beta,D}}{\tau} \text{ and } q_{\beta,D}^* = \frac{c_{b,D} - c_{\beta,D}}{\tau} - \frac{c_{\beta,D}}{(1 - \tau)}. \quad (3.6)$$

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

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

$$\theta_S^{b,0} - c_{b,S} = 0 \iff \theta_S^{b,0} = c_{b,S}. \quad (3.7)$$

Accordingly, in country S demand for b is given by

$$q_{b,S}^* = 1 - c_{b,S}. \quad (3.8)$$

In chapters 4 and 6, I assume co-payments in the form of coinsurance. In chapter 5, I compare two cost-sharing systems, coinsurance and indemnity insurance.

Coinsurance In the case of coinsurance, health insurance reimburses a fraction $0 < \kappa < 1$ of the drug price, the remaining fraction $1 - \kappa = \gamma$ 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). Consequently, co-payments are given as

$$c_{i,D} = \gamma_D p_{i,D} \text{ and } c_{i,S} = \gamma_S p_{i,S}. \quad (3.9)$$

If parallel trade is not allowed (regime of national exhaustion of intellectual property rights), the location of the consumer indifferent between the locally sourced version of the drug and not purchasing is given by

$$\theta_D^{b,0}(\gamma) = \gamma_D p_{b,D}. \quad (3.10)$$

If parallel trade is legal (international exhaustion of intellectual property rights), the location of the consumer indifferent between the locally sourced version of the drug and the parallel import is given by

$$\theta_D^{b,\beta}(\gamma) = \frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau}. \quad (3.11)$$

That is, for the choice between the two versions of the drug, the patient trades off the fraction γ_D of the price difference $p_{b,D}^* - p_{\beta,D}^*$ against then perceived quality difference τ .

In country S , the location of the consumer indifferent between the locally sourced version of the drug and not purchasing is given by

$$\theta_S^{b,0}(\gamma) = \gamma_S p_{b,S}, \text{ resp. } \theta_S^{b,0}(\gamma) = \gamma_S p_{b,S}^*. \quad (3.12)$$

Indemnity Insurance Indemnity insurance describes a form of lump-sum reimbursement in the event of drug purchase. Reimbursement is not tied to the effectively accrued cost of the drug (Zweifel et al., 2009). Patients are reimbursed a fixed amount δ , independent of their choice of

drug. Accordingly, co-payments are given as

$$c_{i,D}(\delta) = p_{i,D} - \delta_D \text{ and } c_{i,S}(\delta) = p_{i,S} - \delta_S. \quad (3.13)$$

If parallel trade is not allowed (regime of national exhaustion of intellectual property rights), the location of the consumer indifferent between the locally sourced version of the drug and not purchasing is given by

$$\theta_D^{b,0}(\delta) = p_{b,D} - \delta_D. \quad (3.14)$$

If parallel trade is legal (international exhaustion of intellectual property rights), the location of the consumer indifferent between the locally sourced version of the drug and the parallel import is given by

$$\theta_D^{b,\beta}(\delta) = \frac{p_{b,D}^* - p_{\beta,D}^*}{\tau}, \quad (3.15)$$

that is, for the choice between the two versions of the drug, consumers take the full price difference $p_{b,D}^* - p_{\beta,D}^*$ into account.

In country S , the location of the consumer indifferent between the locally sourced version of the drug and not purchasing is given by

$$\theta_S^{b,0}(\delta) = p_{b,S} - \delta_S, \text{ resp. } \theta_S^{b,0}(\delta) = p_{b,S}^* - \delta_S. \quad (3.16)$$

Compared to the original demand curve with no reimbursement by the health insurance, coinsurance rotates the demand curve, while indemnity insurance shifts the demand curve. Figure 3.1 illustrates this. The bold line is the original demand curve $p = 1 - q$, the thin dashed line is the demand curve with a coinsurance rate of $\gamma = 0.5$ and the solid line is the demand curve with a reimbursement amount of $\delta = 0.5$. Point A denotes the point with identical out-of-pocket expenditure for a price of $p = 1$.

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

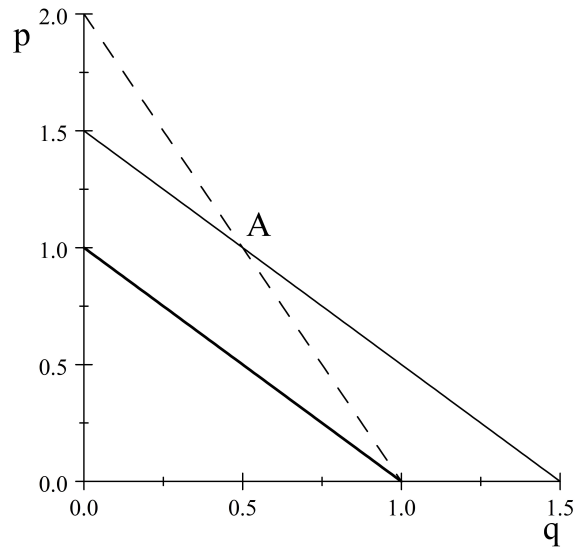


Figure 3.1: Coinsurance Rates vs. Indemnity Insurance

is sufficiently low:

$$\gamma_S \leq \gamma_D \frac{4}{(1 - \tau)(2 - 3\tau)}. \quad (3.17)$$

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

3.3 Timing of the Game

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_j and fixed fee ϕ_j . In the second and final stage, the manufacturer and the foreign intermediary (chapters 4 and 5), respectively the domestic intermediary I_D and the foreign intermediary I_S (chapter 6) compete in prices in country D , the (foreign) intermediary sets the price in country S .

Chapter 4

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

Abstract

This chapter studies externalities of national decisions on health policy under pharmaceutical parallel trade. In particular, it analyzes changes in coinsurance rates (patients pay a percentage of the price) with respect to the effect on drug prices, quantities, and public pharmaceutical expenditure in the destination and source country of parallel imports. 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.

4.1 Introduction

This chapter studies externalities of national decisions on health policy, in particular, changes in coinsurance rates, under pharmaceutical parallel trade.

This analysis is motivated by the conflict between the consequences of parallel trade, namely market integration, and the 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 the internal market. The European Union has adopted regional exhaustion of intellectual property rights, which implies that parallel imports are legal within the European Union¹, 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, 2000a). 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 requires 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 in many destination countries by providing incentives for patients to purchase parallel imports (via the cost-sharing mechanism) or legal requirements to dispense parallel imported drugs, which ensure 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.

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

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. 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.

Several theoretical models predict price convergence², e.g. Rey (2003) or Jelovac & Bordoy (2005). Empirical evidence on this is, however, mixed. 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.

Regional exhaustion of property rights and the free movement of goods allows wholesalers or parallel traders to import pharmaceuticals from other countries without the authorization of the

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

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 pharmaceutical 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 chapter 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. At the same time, their design takes several health policy objectives (as mentioned in the introduction) into account: public pharmaceutical 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), presented in the preceding chapter.

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 in the source country. 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; an increase of the coinsurance rate in the source country reinforces the competition effect from parallel trade in the destination country.

The rest of the chapter is organized as follows. The next two sections analyze the equilibrium without parallel trade and the equilibrium with parallel trade. In Section 4.4, the effects of parallel trade with respect to drug prices and price convergence are studied. Section 4.5 examines the externalities of changes in the coinsurance rate, section 4.6 discusses implications for health policy. Section 4.7 concludes.

4.2 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_S that determines the drug price in country S – are independent.

The manufacturers profit is given as

$$\pi_M = \underbrace{p_{b,D}(1 - \gamma_D p_{b,D})}_{\pi_{b,D}} + \underbrace{w_S(1 - \gamma_S p_{b,S})}_{\pi_{w_b}} + \phi_S, \quad (4.1)$$

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_S the fixed fee that is used to extract the intermediary's profit.

The wholesaler's total profit is given as

$$\pi_I = \underbrace{(p_{b,S} - w_S)(1 - \gamma_S p_{b,S})}_{\pi_{b,S}} - \phi_S, \quad (4.2)$$

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

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

$$\phi_S = \pi_{b,S} = \frac{(1 - w_S\gamma_S)^2}{4\gamma_S} \quad (4.3)$$

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. $w_S = 0$ 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

$$p_{b,D} = \frac{1}{2\gamma_D} \text{ and } p_{b,S} = \frac{1}{2\gamma_S}. \quad (4.4)$$

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 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,

³This result can also be obtained by substituting (4.3) and equilibrium prices into (4.1) and maximizing with respect to w .

only:

$$p_{b,D} - p_{b,S} = \frac{\gamma_S - \gamma_D}{2\gamma_D\gamma_S}. \quad (4.5)$$

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

$$q_{b,D} = \frac{1}{2}, \quad q_{b,S} = \frac{1}{2}. \quad (4.6)$$

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 pharmaceutical expenditure amounts to

$$E_D(\gamma) = (1 - \gamma_D)p_{b,D}q_{b,D}. \quad (4.7)$$

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

$$E_S(\gamma) = (1 - \gamma_S)p_{b,S}q_{b,S}. \quad (4.8)$$

4.3 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

$$\begin{aligned}
\pi_M^* &= \underbrace{p_{b,D}^* \left(1 - \frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_{\pi_{b,D}^*} + \underbrace{w_S^* (1 - \gamma_S p_{b,S}^*)}_{\pi_{w_b}^*} \\
&\quad + \underbrace{w_S^* \left(\frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1 - \tau)} \right)}_{\pi_{w_\beta}^*} + \phi_S^*, \tag{4.9}
\end{aligned}$$

where $\pi_{b,D}^*$ denotes the profit from direct sales 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_S^* the fixed fee. 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

$$\pi_I^* = \underbrace{(p_{b,S}^* - w_S^*) (1 - \gamma_S p_{b,S}^*)}_{\pi_{b,S}^*} + \underbrace{(p_{\beta,D}^* - w_S^*) \left(\frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1 - \tau)} \right)}_{\pi_{\beta,D}^*} - \phi_S^*, \tag{4.10}$$

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 (4.9) with respect to $p_{b,D}^*$. The first order

condition of this problem is

$$\underbrace{\left(1 - \frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau}\right)}_I + \underbrace{p_{b,D}^* \left(-\frac{\gamma_D}{\tau}\right)}_{II} + \underbrace{w_S^* \left(\frac{\gamma_D}{\tau}\right)}_{III} = 0, \quad (4.11)$$

which yields the best response function $p_{b,D}^* = \frac{\tau}{2\gamma_D} + \frac{1}{2} (p_{\beta,D}^* + w_S^*)$. 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 (4.10) with respect to $p_{\beta,D}^*$ and $p_{b,S}^*$. The first order condition with respect to $p_{\beta,D}^*$ is

$$\left(\frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)}\right) + (p_{\beta,D}^* - w_S^*) \left(-\frac{\gamma_D}{\tau} - \frac{\gamma_D}{1-\tau}\right) = 0 \quad (4.12)$$

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

In country S , the intermediary maximizes (4.10) with respect to $p_{b,S}^*$. The first order condition to this maximization problem is

$$(1 - \gamma_S p_{b,S}^*) + (p_{b,S}^* - w_S^*) \underbrace{(-\gamma_S)}_{\frac{\partial q_{b,S}^*}{\partial p_{b,S}^*}} = 0, \quad (4.13)$$

resulting in the price $p_{b,S}^* = \frac{1+w_S^* \gamma_S}{2\gamma_S}$. The first order condition under parallel trade is identical

to the first order condition under no parallel trade. Note that as $p_{b,S}^*$ increases in the wholesale price w_S^* , $p_{b,S}^*$ will be higher under parallel trade, if $w_S^* > 0$.

With

$$\phi_S^* = \underbrace{\frac{(1 - w_S^* \gamma_S)^2}{4\gamma_S}}_{\pi_{b,S}^*} + \underbrace{\frac{\tau (1 - 2w_S^* \gamma_D - \tau)^2}{\gamma_D (1 - \tau) (3 + \tau)^2}}_{\pi_{\beta,D}^*} \quad (4.14)$$

the manufacturer extracts the intermediary's total profit. Substituting (4.14) and equilibrium prices into (4.9) and maximizing with respect to w_S^* gives the wholesale price:

$$w_S^* = \frac{2(1 - \tau)(9 - 5\tau)}{4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(3 + \tau)^2}. \quad (4.15)$$

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. $w_S = 0$. 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_S^* reflects the trade-off between an aggravated double marginalization problem in country S and intensified competition in country D .

Equilibrium drug prices are

$$p_{b,D}^* = \frac{2\gamma_D(9 - 5\tau) + 2\tau\gamma_S(3 + \tau)(1 - \tau)}{\gamma_D[4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(3 + \tau)^2]}, \quad (4.16)$$

$$p_{\beta,D}^* = \frac{(1 - \tau)[2\gamma_D(9 - 5\tau) + \tau\gamma_S(3 + \tau)(1 - \tau)]}{\gamma_D[4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(3 + \tau)^2]}, \quad (4.17)$$

and

$$p_{b,S}^* = \frac{4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(27 - 4\tau + \tau^2)}{2\gamma_S[4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(3 + \tau)^2]}. \quad (4.18)$$

Equilibrium quantities are

$$q_{b,D}^* = \frac{2[\gamma_D(9-5\tau) + \gamma_S(3+\tau)(1-\tau)]}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2}, \quad (4.19)$$

$$q_{\beta,D}^* = \frac{\gamma_S(3+\tau)(1-\tau)}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2}, \quad (4.20)$$

and

$$q_{b,S}^* = \frac{4\gamma_D(9-5\tau) - \gamma_S(1-\tau)(9-16\tau-\tau^2)}{2[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}. \quad (4.21)$$

In the destination country, public drug expenditure is given as

$$E_D^* = (1 - \gamma_D) (p_{b,D}^* q_{b,D}^* + p_{\beta,D}^* q_{\beta,D}^*),$$

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

$$E_S^* = (1 - \gamma_S) p_{b,S}^* q_{b,S}^*. \quad (4.22)$$

4.4 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.

4.4.1 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 B.1 for details.

Compared to segmented markets, competition from parallel trade reduces the price of the

drug sold directly by the manufacturer:

$$\frac{p_{b,D}^*}{p_{b,D}} < 1, \quad (4.23)$$

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

$$\frac{p_{\beta,D}^*}{p_{b,D}^*} < 1. \quad (4.24)$$

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 version⁴.

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:

$$\frac{q_{b,D}^*}{q_{b,D}} > 1. \quad (4.25)$$

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.

4.4.2 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 B.1 for details.

Compared to segmented markets, the wholesale price w_S^* 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

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

order to deter parallel trade partially:

$$w_S^* > w_S = 0. \quad (4.26)$$

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

$$\frac{p_{b,S}^*}{p_{b,S}} > 1 \quad (4.27)$$

and the higher price reduces the quantity sold:

$$\frac{q_{b,S}^*}{q_{b,S}} < 1. \quad (4.28)$$

4.4.3 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 B.1 for details:

$$\frac{p_{b,D}^* - p_{b,S}^*}{p_{b,D} - p_{b,S}} < 0 \text{ if } \gamma_S > \gamma_D \quad (4.29)$$

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.

⁵Note 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 $\gamma_F > \gamma_H$.

4.5 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.2 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 co-payments, more specifically increases of coinsurance rates. In the case of reductions of co-payments, i.e. reductions of coinsurance rates, price and quantity changes go in the opposite direction.

4.5.1 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 B.3.

In the destination country, the increase in co-payments, i.e. the increase in the coinsurance rate results in lower drug prices and lower quantities sold.

An increase in 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.:

$$\frac{\partial q_{b,D}^*}{\partial \gamma_D} < 0. \quad (4.30)$$

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

$$\frac{\partial q_{\beta,D}^*}{\partial \gamma_D} > 0, \text{ if } p_{\beta,D}^* < (1 - \tau) p_{b,D}^*. \quad (4.31)$$

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 well⁶. This is demonstrated by the best response function: $p_{\beta,D}^* = \frac{1}{2} \left(w_S^* + p_{b,D}^* (1 - \tau) \right)$.

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

$$\frac{\partial p_{b,D}^*}{\partial \gamma_D} < 0, \quad \frac{\partial p_{\beta,D}^*}{\partial \gamma_D} < 0. \quad (4.32)$$

Competition from parallel trade of demand limits the ability to decrease prices sufficiently to compensate the increase in the coinsurance rate and consequently, effective consumer prices increase:

$$\frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_D} > 0, \quad \frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_D} > 0. \quad (4.33)$$

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

$$\frac{\partial q_{b,D}^*}{\partial \gamma_D} < 0, \quad \frac{\partial q_{\beta,D}^*}{\partial \gamma_D} < 0. \quad (4.34)$$

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

$$\frac{\partial E_D^*}{\partial \gamma_D} < 0. \quad (4.35)$$

⁶In addition, 4.31 implies that the intermediary has to lower $p_{\beta,H}^*$ in order to prevent a decrease of demand.

Spillovers of co-payment 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:

$$\frac{\partial w_S^*}{\partial \gamma_D} < 0. \quad (4.36)$$

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

$$\frac{\partial p_{b,S}^*}{\partial \gamma_D} < 0. \quad (4.37)$$

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

$$\frac{\partial \gamma_S p_{b,S}^*}{\partial \gamma_D} < 0, \quad (4.38)$$

which increases the quantity consumed:

$$\frac{\partial q_{b,S}^*}{\partial \gamma_D} > 0. \quad (4.39)$$

Thus, for increasing co-payments 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 co-payment 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 B.4):

$$\frac{\partial E_S^*}{\partial \gamma_D} > 0. \quad (4.40)$$

Consequently, an increase in 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 in the quantity consumed. By reducing drug prices and increasing the quantities sold, a co-payment increase in the source country mitigates the double marginalization effect in the source country S .

Proposition 4.1 summarizes the effect of an increase in the coinsurance rate in the destination country D :

Proposition 4.1. *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, and raises health expenditure in country S .*

4.5.2 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 D . For explicit expressions of changes in prices and quantities see Appendix B.3.

In the source country, the increase in co-payments, i.e. an increase in 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.:

$$\frac{\partial (1 - \gamma_S p_{b,S}^*)}{\partial \gamma_S} < 0. \quad (4.41)$$

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

Accordingly, the drug price decreases in γ_S :

$$\frac{\partial p_{b,S}^*}{\partial \gamma_S} < 0. \quad (4.42)$$

The effective drug price increases, as marginal cost is no longer zero⁷:

$$\frac{\partial \gamma_S p_{b,S}^*}{\partial \gamma_S} > 0 \quad (4.43)$$

The price decrease does not offset the effect of an increase in the co-payment and thus, under

⁷Note 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.

coinsurance rates, the quantity consumed also decreases:

$$\frac{\partial q_{b,S}^*}{\partial \gamma_S} < 0. \quad (4.44)$$

Figure 4.1 illustrates the effect of an increase in the coinsurance rate for marginal cost greater than 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 in 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 Figure 4.1 shows. In other words, as the effective consumer price increases with the coinsurance rate, the quantity consumed decreases.

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

$$\frac{\partial E_S^*}{\partial \gamma_S} < 0. \quad (4.45)$$

Spillovers of co-payment 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

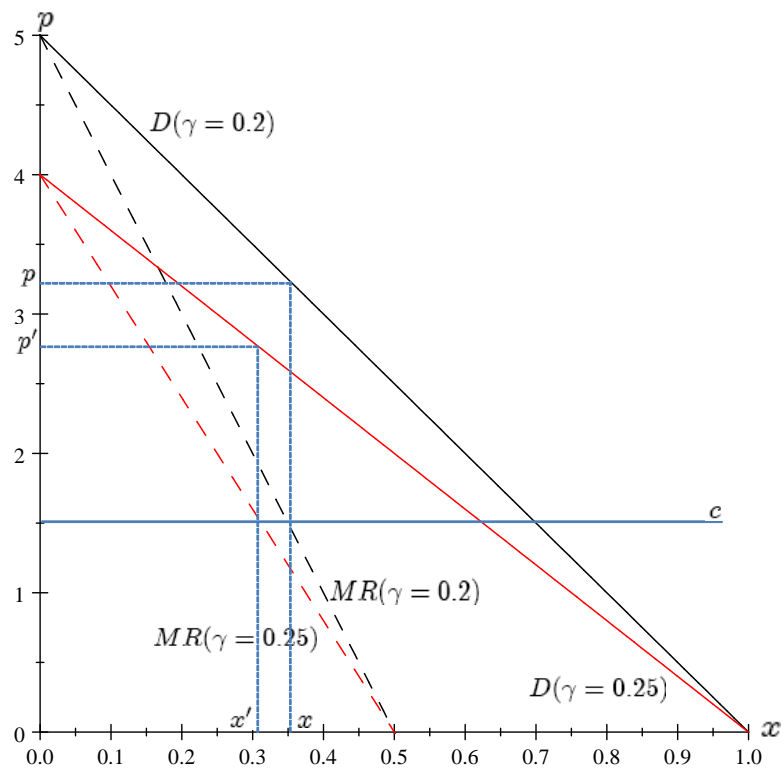


Figure 4.1: Increase of Coinsurance Rate, $c > 0$.

marginalization effect. Consequently, the manufacturer reduces the wholesale price:

$$\frac{\partial w_S^*}{\partial \gamma_S} < 0. \quad (4.46)$$

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_S^* : $p_{\beta,D}^* = \frac{1}{2} \left(w_S^* + p_{b,D}^* (1 - \tau) \right)$. 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: $p_{b,D}^* = \frac{1}{2} \left(\tau + p_{\beta,D}^* + w_S^* \right)$. Accordingly, in country D , both drug prices decrease in the coinsurance rate in the source country:

$$\frac{\partial p_{b,D}^*}{\partial \gamma_S} < 0, \quad \frac{\partial p_{\beta,D}^*}{\partial \gamma_S} < 0. \quad (4.47)$$

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

$$\frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_S} < 0, \quad \frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_S} < 0. \quad (4.48)$$

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

$$\frac{\partial q_{b,D}^*}{\partial \gamma_S} > 0, \quad \frac{\partial q_{\beta,D}^*}{\partial \gamma_S} > 0. \quad (4.49)$$

As the effect of lower prices more than offsets the effect of a higher quantity, public pharmaceutical expenditure decreases (see Appendix B.4):

$$\frac{\partial E_D^*}{\partial \gamma_S} < 0. \quad (4.50)$$

Thus, an increase in the co-payment 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 co-payment increase in the source country reinforces the effect of competition by parallel trade in country D .

Proposition 4.2 summarizes the effect of an increase in the coinsurance rate in the source country S :

Proposition 4.2. *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 .*

4.6 Implications 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.

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 effect increases consumer surplus and increases public pharmaceutical expenditure in the source country. If the coinsurance rate in the source country is the result of a political optimization, taking into account different health policy objectives⁸, a change of the coinsurance rate in the destination country induces a change of the coinsurance rate in the source country. At the same time, 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. In the destination country, consumer surplus is increased and public pharmaceutical expenditure is reduced. Thus, a change of the

⁸As 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.

coinsurance rate in the source country might trigger a change of the coinsurance rate in the destination 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.

If governments set coinsurance rates without taking the externalities for the respective other 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 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.

Assume that governments in both countries set coinsurance rates to maximize total welfare. In the destination country D , welfare is given as the sum of consumer surplus and the manufacturer's profit⁹ net of public pharmaceutical expenditure:

$$W_D^* = CS_D^* + \pi_M^* - E_D^*, \quad (4.51)$$

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

$$W_S^* = CS_S^* - E_S^*. \quad (4.52)$$

Total welfare, i.e. $W^* = W_D^* + W_S^*$, strictly increases in γ_D and decreases in γ_S , see Appendix

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

B.5. Thus, total welfare is maximized for

$$\gamma_D = 1, \gamma_S = \varepsilon.^{10} \tag{4.53}$$

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 also 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 also public pharmaceutical expenditure. This also illustrates that the coordination of pharmaceutical policy does not imply identical coinsurance rates in both countries. 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.

4.7 Conclusion

In this chapter, 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 in the source country. 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 in the coinsurance rate in the destination country mitigates the double marginalization effect in the source country; an increase in the coinsurance rate in the source country reinforces the competition effect from

¹⁰Note that $\gamma_F \in (0, 1)$.

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 harmonization 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.

Chapter 5

Pharmaceutical Cost-sharing

Systems and Savings for Health Care

Systems from Parallel Trade

Abstract

This chapter analyzes the consequences of parallel trade on health care systems. In particular, a coinsurance scheme (consumers pay a percentage of the drug price out-of-pocket) and an indemnity insurance scheme (reimbursement is independent of the drug price) are compared with respect to changes of co-payments, i.e. out-of-pocket expenditure for patients, and public pharmaceutical expenditure. Independent of the cost-sharing system, parallel trade generates a price-decreasing competition effect in the destination country and a price-increasing double marginalization effect in the source country. In the destination country, co-payments for patients decrease to a larger extent under indemnity insurance, reductions of public pharmaceutical expenditure occur only under coinsurance. In the source country, co-payments increase less under coinsurance, health expenditure is reduced more under indemnity insurance (the expenditure-decreasing effect of a lower quantity dominates the effect of the drug price increase).

5.1 Introduction

This chapter studies the consequences of parallel trade on health care systems, especially changes of co-payments and changes of public pharmaceutical expenditure, under a coinsurance scheme and an indemnity insurance scheme.

This analysis is motivated by the observation that the institutional setting, in which parallel trade takes place, is relevant for the consequences of parallel trade. More precisely, the design of the cost-sharing system, i.e. rules of co-payment and reimbursement, are an important factor in determining consequences of price changes. Although the price effects induced by parallel trade are independent of the cost-sharing system, the cost-sharing system determines the magnitude of price changes and whether savings accrue and for whom.

Due to the high level of public health expenditure, savings for health care systems are a major concern in health policy. In addition to influencing pharmaceutical prices directly, e.g. via direct price regulation or regulation of wholesale margins, promoting the substitution of higher-priced brand-name drugs by less expensive equivalents is an instrument to contain pharmaceutical expenditure. This might include generic versions of brand-name drugs, which may differ in terms of binders, fillers, preservatives and density of packing. Alternatively, parallel imported drugs are de facto identical, lower-priced versions of (locally sourced) brand-name drugs, which are imported from other countries without the permission of the manufacturer. In the European Economic Area, where parallel trade is legal, price differences between countries are substantial. For instance, a package of the drug Capecitabine (Xeloda[®], Roche), a chemotherapeutic agent costs € 424.20 in Germany, but is available for only € 295.60 in the United Kingdom (Kanavos et al., 2011).

Whereas the exploitation of these arbitrage opportunities is intended to contain (public) pharmaceutical expenditure in the destination countries of parallel imports, empirical evidence on this is ambiguous. Three recent studies, Kanavos et al. (2004), West & Mahon (2003), and Enemark et al. (2006), have presented contradictory results with respect to the ability of parallel trade to generate savings for health insurance funds and patients.

For patients, the cost-sharing system and co-payment rules constitute the direct channel through which they may benefit from purchasing parallel imports (Kanavos et al., 2004). Patients will only chose parallel imports over locally sourced drugs if they benefit financially, i.e. save on out-of-pocket expenditure from doing so. Consequently, the cost-sharing system provides incentives for patients to buy parallel imports (Enemark et al., 2006). For example, a flat fee co-payment, a charge per service, fails to sensitize patients for price differences between locally sourced and parallel imported drugs and there is no incentive to buy lower priced parallel imports. A co-payment in the form of a coinsurance (patients pay a percentage of the total price) however, makes patients benefit from choosing a cheaper drug. Patients are encouraged to buy parallel imports. Similarly, this applies to cost-sharing systems including deductibles (patients pay the first x Euros before insurance coverage begins) or indemnity insurance (a fixed amount independent of the price is reimbursed) (Robinson, 2002). The extent of the co-payment and the price elasticity of demand are important in determining incentives to choose lower-priced drugs. In other words, patients are more likely to purchase parallel imports, the more they are exposed to the price difference between locally sourced drugs and parallel imports.

For health insurances, the cost-sharing design determines the level of savings. In particular, the link between reimbursement and drug prices is relevant, as it allows health insurances to benefit from lower drug prices. If patients pay a percentage of the total price (coinsurance) and the remaining fraction of total expenditure is reimbursed by health insurances, lower drug prices - both lower prices for parallel imports and lower prices of locally sourced versions due to competitive effects - then translate to lower public pharmaceutical expenditure. Furthermore, as co-payment rules provide incentives for patients to buy parallel imports, cost-sharing determines the competitive pressure by parallel trade. Accordingly, also public pharmaceutical expenditure is reduced by more, if competition from parallel trade is strong and the market share of parallel imports is high. Consequently, the cost-sharing system is both a driver of parallel trade, as it determines cross-country price differences, and, more importantly, an important factor in determining savings from parallel trade.

In the European Union, all 27 member states apply some form of cost-sharing in relation to pharmaceuticals, mostly in the form of coinsurance, where patients pay a percentage of the price (Mossialos & Le Grand, 1999). In the United Kingdom a flat rate co-payment per package applies. Denmark, Sweden and Norway¹ use a combination of deductibles and coinsurance (Robinson, 2002; Kanavos et al., 2004). In addition, the reference price system, in which the regulator sets a ceiling for the amount reimbursable (reference price) for a group of pharmaceuticals (cluster), can be found in Belgium, Denmark, Finland, France, Germany, Greece, Italy, the Netherlands, Portugal, Spain, and the United Kingdom (Puig-Junoy, 2010). By making reimbursement independent of drug choice, reference pricing is similar to indemnity insurance, where patients are reimbursed a price-independent amount and pay the difference between the drug price and the reimbursement amount out-of-pocket. But whereas reference pricing can be considered to impose an avoidable co-payment (avoidable if a drug priced at the reference price is chosen), indemnity insurance is a form of mandatory cost-sharing, as reimbursement amounts are lower than drug prices. That is, in the European Union, member states apply different cost-sharing systems, which differ in their impact on savings generated by parallel trade. However, not all textbook examples of co-payments can also be found in reality due to inherent structural weaknesses, as this analysis shows.

The importance of cost-sharing systems for the consequences of parallel trade has been emphasized in the empirical literature, but it has not attracted much attention in the theoretical literature on parallel trade. Only Bordoy & Jelovac (2005) and Köksal (2009) examine the effect of cost-sharing structure on parallel trade. Bordoy & Jelovac (2005) argue that if parallel trade is driven by differences in coinsurance rates, it reduces welfare. Based on the Bordoy & Jelovac (2005)-model, Köksal (2009) compares price effects caused by parallel trade under coinsurance and reference pricing. Under reference pricing, price reductions from parallel trade in the destination country are higher than under coinsurance. Furthermore, reference pricing does not affect the drug price in the exporting country.

¹Norway is not part of the European Union, but of the European Economic Area, for which the principle of free movement of goods applies.

This analysis differs from Köksal (2009) in the object of study. Köksal compares coinsurance as an instrument of cost-sharing to reference pricing, which can rather be characterized as an instrument of pharmaceutical price regulation. Apart from the Netherlands, all European countries applying reference pricing also use coinsurance as a cost-sharing system. That is, reference pricing rather constitutes a supplementary instrument to contain drug prices than to reduce moral hazard. Consequently, the two instruments studied by Köksal (2009) are not policy substitutes, but rather complements. On the contrary, this chapter compares coinsurance and indemnity insurance, which are both (pure) cost-sharing instruments. Coinsurance and indemnity insurance can be considered policy alternatives, among which policymakers may choose from. Indemnity insurance is a textbook example of a cost-sharing instrument, but cannot be found in its pure form in any member state of the European Union. This chapter considers indemnity insurance as a policy alternative to commonly applied coinsurance and may provide an explanation for coinsurance being preferred to indemnity insurance, if minimizing health expenditure is the prevailing health policy objective.

In addition, whereas Bordoy & Jelovac (2005) (and Köksal, 2009) consider parallel trade as retail-level horizontal arbitrage, where parallel traders buy the drugs at market prices in the source country, this chapter explains parallel trade as a by-product of vertical control structures: Indirect sales through an intermediary are the trigger for parallel trade, as an intermediary may resell a drug in other ways than intended by the manufacturer. Accordingly, parallel trade amounts to vertical arbitrage and is mainly determined by the wholesale price set by the manufacturer. Commonly, pharmaceutical manufacturers sell not directly, but through independent wholesalers (Taylor, Mrazek & Mossialos, 2004). In addition, this approach separates the *cause for* from the *consequences of* parallel trade. Horizontal arbitrage is triggered by retail price differences and accordingly, differences in the cost-sharing system or the extent of the co-payment, which contributes to retail price differences. That is, in Bordoy & Jelovac (2005) and Köksal (2009) the design of the cost-sharing system is the determining factor for whether parallel trade occurs (cost-sharing as a trigger), but also for what consequences parallel trade has (impact of

cost-sharing). Vertical arbitrage however, assumes vertical restraints as the driver of parallel trade. Even for identical cost-sharing systems and/or identical co-payments, arbitrage would be profitable and parallel trade would occur. Accordingly, under vertical arbitrage, parallel trade flows are not contingent on assumptions of differences in cost-sharing systems or instruments. Thus, taking vertical arbitrage as a starting point, the design of the cost-sharing system is only the determining factor for the consequences of parallel trade. This allows me to analyze the interaction between cost-sharing systems and parallel trade also for identical cost-sharing systems.

Against this background, this chapter explores the role of cost-sharing for the effects of parallel trade in a two-country model following Maskus & Chen (2002) and Chen & Maskus (2005).

Independent of the cost-sharing scheme, parallel trade generates a competition effect in the destination country, resulting in lower drug prices and a higher quantity sold. Due to the higher wholesale price, as compared to segmented markets, a double-marginalization effect occurs in the source country. In the destination country, savings for patients occur under both systems, with savings being relatively higher under indemnity insurance. However, savings for health insurance occur only under coinsurance. Indemnity insurance fails to link reimbursement to drug prices, and via the increase in the quantity demanded, lower drug prices result in higher expenditure. In the source country, the drug price increase following from the increase of the wholesale price results in additional expenses for consumers under both cost-sharing systems. Under coinsurance, additional expenses are relatively lower. Parallel trade results in lower health expenditure under both cost-sharing systems, as the effect from a lower quantity consumed dominates the effect of a higher drug price on expenditure. Under indemnity insurance, the relative reduction of health expenditure is higher.

The rest of this chapter is organized as follows. The next two sections analyze the equilibrium without parallel trade and the equilibrium with parallel trade. In Section 5.4, the effects of parallel trade with respect to price changes, changes in co-payments, and public pharmaceutical

expenditure are studied. Section 5.5 concludes.

5.2 Equilibrium without Parallel Trade

First consider the case of no parallel trade, when 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_S , which determines the drug price in country S – are independent.

The manufacturer's profit is given as

$$\pi_M = \underbrace{p_{b,D}(1 - c_{b,D})}_{\pi_{b,D}} + \underbrace{w_S(1 - c_{b,S})}_{\pi_{w_b}} + \phi_S, \quad (5.1)$$

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_S the fixed fee, which is used to extract the intermediary's profit.

The wholesaler's total profit is given as

$$\pi_I = \underbrace{(p_{b,S} - w_S)(1 - c_{b,S})}_{\pi_{b,S}} - \phi_S, \quad (5.2)$$

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

In market D , the manufacturer M maximizes (5.1) with respect to $p_{b,D}$. The first order condition to this problem is

$$\underbrace{(1 - c_{b,D})}_I + \underbrace{p_{b,D} \left(-\frac{\partial c_{b,D}}{\partial p_{b,D}} \right)}_{II} = 0, \quad (5.3)$$

yielding the monopoly drug price $p_{b,D}(c_{b,D})$.

In market S , the intermediary I maximizes (5.2) with respect to $p_{b,S}$. The first order condi-

tion to this problem is

$$(1 - c_{b,S}) + (p_{b,S} - w_S) \left(-\frac{\partial c_{b,S}}{\partial p_{b,S}} \right) = 0, \quad (5.4)$$

resulting in the monopoly drug price $p_{b,S}(c_{b,S})$. The first order condition shows that $p_{b,S}$ increases in the wholesale price w_S .

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

$$\phi_S = \pi_{b,S} = (p_{b,S} - w_S)(1 - c_{b,S}) \quad (5.5)$$

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. $w_S = 0^2$. 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.

Drug prices in both countries are given as:

D↓, S→	Coinsurance	Indemnity insurance	
Coinsurance	$p_{b,D}(\gamma) = \frac{1}{2\gamma_D}$, $p_{b,S}(\gamma) = \frac{1}{2\gamma_S}$	$p_{b,D}(\gamma) = \frac{1}{2\gamma_D}$, $p_{b,S}(\delta) = \frac{1+\delta_S}{2}$	(5.6)
Indemnity insurance	$p_{b,D}(\delta) = \frac{1+\delta_D}{2}$, $p_{b,S}(\gamma) = \frac{1}{2\gamma_S}$	$p_{b,D}(\delta) = \frac{1+\delta_D}{2}$, $p_{b,S}(\delta) = \frac{1+\delta_S}{2}$.	

Note that drug prices only depend on the cost-sharing system in the respective country.

Under coinsurance, drug prices decrease in coinsurance rates. Effective prices for consumers ($\gamma_D p_{b,D}(\gamma) = \gamma_S p_{b,S}(\gamma) = \frac{1}{2}$) are equivalent to prices without insurance coverage ($p_{b,D} = p_{b,S} = \frac{1}{2}$). That is, the effect from reimbursement by health insurance is completely appropriated by

²This result can also be obtained by substituting (5.5) and equilibrium prices into (5.1) and maximizing with respect to w_S .

the manufacturer. Price differences result from differences in health care systems (= coinsurance rates) only. Quantities are independent of coinsurance rates, as the effect from reimbursement completely accrues to the manufacturer.

Under indemnity insurance, drug prices increase in the reimbursement amount. Effective prices ($p_{b,D}(\delta) - \delta_D = \frac{1-\delta_D}{2}$ and $p_{b,S}(\delta) - \delta_S = \frac{1-\delta_S}{2}$) are lower than prices without insurance ($p_{b,D} = p_{b,S} = \frac{1}{2}$). The effect from reimbursement benefits both the manufacturer (higher market prices than without insurance) and patients (lower effective prices than without insurance). Price differences occur, when reimbursement differs across countries. Quantities increase in the reimbursement amount, as effective prices decrease in the reimbursement amount and more consumers buy. Differences in quantities sold in country D and S stem from differences in health care system (reimbursement amounts).

5.3 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 the intermediary – 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

$$\pi_M^* = \underbrace{p_{b,D}^* \left(1 - \frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} \right)}_{\pi_b^*} + \underbrace{w_S^* (1 - c_{b,S}^*)}_{\pi_{w_b}^*} + \underbrace{w_S^* \left(\frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} - \frac{c_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{w_\beta}^*} + \phi_S^*, \quad (5.7)$$

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_S^* the fixed fee. An asterisk is used to denote variables associated with parallel trade.

The manufacturer's profit differs from the profit under no parallel trade in three respects: First, it 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 importing. Third, the intermediary's sales as reimports result in additional wholesale profit for the manufacturer.

The intermediary's profit is given as

$$\pi_I^* = \underbrace{(p_{b,S}^* - w_S^*) (1 - c_{b,S}^*)}_{\pi_{b,S}^*} + \underbrace{(p_{\beta,D}^* - w_S^*) \left(\frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} - \frac{c_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{\beta,D}^*} - \phi_S^*, \quad (5.8)$$

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 (5.7) with respect to $p_{b,D}^*$. The first order condition of this problem is

$$\underbrace{\left(1 - \frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} \right)}_I + \underbrace{p_{b,D}^* \left(-\frac{\partial c_{b,D}^*}{\partial p_{b,D}^*} \frac{1}{\tau} \right)}_{II} + \underbrace{w_S^* \left(\frac{\partial c_{b,D}^*}{\partial p_{b,D}^*} \frac{1}{\tau} \right)}_{III} = 0, \quad (5.9)$$

which yields the best response function $p_{b,D}^*(c_{b,D}^*, w_S^*, p_{\beta,D}^*)$. Compared to the first order condition for segmented markets, part I and consequently $p_{b,D}^*$ are higher (lower) under parallel trade, if $c_{b,D}^* < \frac{c_{\beta,D}^*}{(1-\tau)}$ ($c_{b,D}^* > \frac{c_{\beta,D}^*}{(1-\tau)}$). For coinsurance, $c_{b,D}^* < \frac{c_{\beta,D}^*}{(1-\tau)}$ if $p_{b,D}^* < \frac{p_{\beta,D}^*}{(1-\tau)}$, that is, if the parallel import weighted by the quality discount is priced higher than the locally sourced version. For indemnity insurance, $c_{b,D}^* < \frac{c_{\beta,D}^*}{(1-\tau)}$ if $p_{b,D}^* < \frac{p_{\beta,D}^* - \delta_D \tau}{(1-\tau)}$. 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 (5.8) with respect to $p_{\beta,D}^*$, which yields the first order condition

$$\left(\frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} - \frac{c_{\beta,D}^*}{(1-\tau)} \right) + (p_{\beta,D}^* - w_S^*) \left(-\frac{\partial c_{\beta,D}^*}{\partial p_{\beta,D}^*} \frac{1}{\tau} - \frac{\partial c_{\beta,D}^*}{\partial p_{\beta,D}^*} \frac{1}{1-\tau} \right) = 0 \quad (5.10)$$

and the best response function $p_{\beta,D}^*(c_{\beta,D}^*, w_S^*, p_{b,D}^*)$. Solving for equilibrium prices results in $p_{b,D}^*(c_{b,D}^*, w_S^*)$ and $p_{\beta,D}^*(c_{\beta,D}^*, w_S^*)$.

In country S , the intermediary maximizes (5.8) with respect to $p_{b,S}^*$. The first order condition to this maximization problem is

$$(1 - c_{b,S}^*) + (p_{b,S}^* - w_S^*) \left(-\frac{\partial c_{b,S}^*}{\partial p_{b,S}^*} \right) = 0, \quad (5.11)$$

resulting in the price $p_{b,S}^*(c_{b,S}^*, w_S^*)$. Note that as $p_{b,S}^*$ increases in the wholesale price w_S^* , $p_{b,S}^*$ will be higher under parallel trade, if $w_S^* > 0$.

With

$$\begin{aligned} \phi_S^* &= \pi_{b,S}^* + \pi_{\beta,D}^* \\ &= \underbrace{(p_{b,S}^* - w_S^*) (1 - c_{b,S}^*)}_{\pi_{b,S}^*} + \underbrace{(p_{\beta,D}^* - w_S^*) \left(\frac{c_{b,D}^* - c_{\beta,D}^*}{\tau} - \frac{c_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{\beta,D}^*} \end{aligned} \quad (5.12)$$

the manufacturer extracts the intermediary's total profit. Substituting (5.12) and equilibrium prices into (5.7) and maximizing with respect to w_S^* gives the wholesale price $w_S^*(c_{i,j}^*)$.

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. $w_S = 0$. 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_S^* reflects the trade-off between an aggravated double marginalization problem in country S and intensified competition in country D . Note that as markets are integrated by parallel trade and the wholesale price incorporates the effects in both the destination country D and the source country S , the wholesale price depends on cost-sharing systems in both countries. Similarly, prices and quantities also depend on cost-sharing systems in both countries. To account for the interdependence of pricing decisions under parallel trade, variables such as prices, quantities, will be characterized contingent on cost-sharing systems in both countries. The cost-sharing system's parameters in parentheses denote the cost-sharing in country D and S , respectively, e.g. $p_{b,S}^*(\gamma, \delta)$ denotes the drug price in the source country, if country D applies coinsurance and country S indemnity insurance. The specification of drug prices and quantities for the different combinations of cost-sharing systems in the destination and source country can be found in Appendix C.2.

5.4 The Effect of Parallel Trade on Health Care Systems

If parallel trade is allowed, the manufacturer raises the wholesale price to limit competition from parallel trade. This creates a double marginalization effect with a higher drug price and a lower quantity in the source country S . As the manufacturer cannot block parallel trade entirely, parallel trade generates a competition effect with lower drug prices and a higher quantity consumed in the destination country D .

The direct link between these price changes (resulting from parallel trade) in both countries and the consequences for public health care systems is the cost-sharing system: It drives the changes in co-payments for consumers and public pharmaceutical expenditure funded by health insurance.

The co-payment mechanism determines, whether and to what extent consumers in the destination country benefit from price decreases through savings and whether and to what extent consumers in the source country are exposed to price increases by higher co-payments. In addi-

tion, the reimbursement mechanism determines also the consequences of parallel trade for public pharmaceutical expenditure. Ex ante, price and quantity changes generated by parallel trade have an ambiguous impact on public pharmaceutical expenditure: In the destination country, lower prices may contribute to lower health expenditure, but a higher quantity consumed may work towards higher spending. In the source country, a higher price may have an expenditure-increasing effect, but a lower quantity consumed may reduce health expenditure.

5.4.1 Changes in Co-payments and Public Pharmaceutical Expenditure in the Destination Country

This subsection investigates the consequences of parallel trade for public health care systems in the destination country. Proposition 5.1 summarizes the effects of parallel trade on consumer co-payments and public pharmaceutical expenditure in the destination country D .

Proposition 5.1. *In the destination country D , lower drug prices under parallel trade decrease co-payments under both cost-sharing systems. Suppose that drug prices under segmented markets are identical under coinsurance and indemnity insurance. Then savings for consumers are higher under indemnity insurance, independent of the cost-sharing system in the source country S . Parallel trade generates savings for health insurance only under coinsurance; under indemnity insurance, health expenditure is higher than without parallel trade.*

Proof. See Appendix C.3. □

Under both cost-sharing systems, co-payments strictly increase in drug prices. Consequently, under both coinsurance and indemnity insurance, price decreases translate to lower co-payments:

$$\begin{aligned}\Delta c_{b,D}(\gamma, \cdot) &= c_{b,D}^*(\gamma) - c_{b,D}(\gamma) = \gamma_D (p_{b,D}^*(\gamma, \cdot) - p_{b,D}(\gamma)) < 0, \\ \Delta c_{b,D}(\delta, \cdot) &= c_{b,D}^*(\delta) - c_{b,D}(\delta) = p_{b,D}^*(\delta, \cdot) - p_{b,D}(\delta) < 0,\end{aligned}\tag{5.13}$$

since $p_{b,D}^*(\gamma, \cdot) < p_{b,D}(\gamma)$ and $p_{b,D}^*(\delta, \cdot) < p_{b,D}(\delta)$, respectively. That is, consumers benefit from parallel trade independently of the cost-sharing system.

Which cost-sharing system creates higher savings, is determined by two factors: First, there is a direct impact of the cost-sharing system for a given wholesale price: Under indemnity insurance, reimbursement is price-independent and consumers benefit from the full price decrease. Under coinsurance, only fraction γ_D of the price decrease is passed on to consumers. Second, the wholesale price drives the intensity of competition, as it is the lower bound for the price of the parallel import and thus limits the intermediary undercutting the manufacturer's price.

These two factors are interdependent, as the wholesale price also depends on the degree of competition in the destination country. That is, the first factor would result in a higher wholesale price under indemnity insurance c.p. At the same time, the increase of the wholesale price is limited by the double marginalization effect in the source country. The (relative) magnitude of this effect also depends on the cost-sharing system. Under indemnity insurance, the double marginalization effect is higher c.p.: Consumers bear the full price increase under indemnity insurance and the higher price elasticity turns a given price increase into a higher reduction of quantity.

Accordingly, for a comparison of co-payment changes under coinsurance and indemnity insurance, cost-sharing systems in both countries are important: In the destination country, the total effect of the cost-sharing system and the effect of the cost-sharing system itself on the wholesale price is important. In the source country, the cost-sharing system determines by how much the manufacturer can increase the wholesale price in response to the competition effect in the destination country.

Comparing (relative) co-payments under coinsurance and indemnity insurance, I assume identical drug prices for both cost-sharing systems under segmented markets. Taking into account that reimbursement may not exceed the drug price and consumers co-pay a positive amount, identical drug prices under coinsurance and indemnity insurance imply high coinsurance rates of $\gamma_D > 0.6$, see Appendix C.3 for details. Identical drug prices as standard of comparison imply that the reimbursement amount under indemnity insurance can be written in

terms of the coinsurance rate as follows:

$$p_{b,D}(\gamma) = \frac{1}{2\gamma_D} = \frac{1 + \delta_D}{2} = p_{b,D}(\delta) \iff \delta_D = \frac{1}{\gamma_D} - 1. \quad (5.14)$$

Note that identical drug prices under both cost-sharing instruments do not imply identical co-payments under coinsurance and indemnity insurance due to the different insurance effects. Suppose, $\gamma_D = 0.75$. For $\delta_D = \frac{1}{3}$ drug prices are identical under both coinsurance and indemnity insurance ($p_{b,D}(\gamma) = \frac{2}{3} = p_{b,D}(\delta)$). Then, under coinsurance, consumers pay $c_{b,D}(\gamma) = \frac{1}{2}$ ³ and under indemnity insurance $c_{b,D}(\delta) = \frac{1}{3}$. Assuming identical co-payments for both cost-sharing systems under segmented markets would involve no reimbursement under indemnity insurance due to the insurance absorbing effect of coinsurance under segmented markets. See Appendix C.4. for a detailed explanation of this limitation associated with identical co-payments under both cost-sharing systems as a basis of comparison.

The change in co-payments is lower under indemnity insurance than under coinsurance independent of the cost-sharing system in the source country:

$$\frac{c_{b,D}^*(\gamma, \gamma)}{c_{b,D}(\gamma)} > \frac{c_{b,D}^*(\delta, \gamma)}{c_{b,D}(\delta)} \Big|_{\delta_D = \frac{1}{\gamma_D} - 1}, \quad \frac{c_{b,D}^*(\gamma, \delta)}{c_{b,D}(\gamma)} > \frac{c_{b,D}^*(\delta, \delta)}{c_{b,D}(\delta)} \Big|_{\delta_D = \frac{1}{\gamma_D} - 1}. \quad (5.15)$$

As a lower ratio of relative co-payments corresponds to a higher reduction of co-payments, this is equivalent to co-payments being reduced to a larger extent under indemnity insurance. No matter whether coinsurance or indemnity insurance is applied in the source country, the wholesale price is higher under indemnity insurance ($w_S^*(\delta, \gamma) > w_S^*(\gamma, \gamma)$, $w_S^*(\delta, \delta) > w_S^*(\gamma, \delta)$ resp.). That is, the higher intensity of competition under indemnity insurance induces the manufacturer to raise the wholesale price more under indemnity insurance as compared to coinsurance. The cost-sharing system in the source country restricts the increase of the wholesale price with respect to the double marginalization effect, but does not inhibit the higher increase of the

³Note that the co-payment under coinsurance is independent of the coinsurance rate, as the insurance effect is absorbed by the manufacturer completely.

wholesale price under indemnity insurance. In total, the impact of the full price difference accruing to consumers exceeds the effect of the higher wholesale price and consumer co-payments are reduced more under indemnity insurance.

The competition effect generated by parallel trade has an ambiguous impact on total expenditure. Lower prices decrease expenditure, higher quantities increase expenditure. Whether parallel trade results in reductions of public pharmaceutical expenditure, i.e. the part of total expenditure which is reimbursed by health insurance, depends on the cost-sharing system. Only if reimbursement is linked to drug prices, health insurance benefits from lower prices under parallel trade.

Under coinsurance, the change in public pharmaceutical expenditure associated parallel trade is given as:

$$\begin{aligned} \Delta E_D(\gamma) &= E_D^*(\gamma) - E_D(\gamma) \\ &= (1 - \gamma_D) (p_{b,D}^*(\gamma, \cdot) q_{b,D}^*(\gamma, \cdot) + p_{\beta,D}^*(\gamma, \cdot) q_{\beta,D}^*(\gamma, \cdot) - p_{b,D}(\gamma) q_{b,D}(\gamma)), \end{aligned} \quad (5.16)$$

of which the term in parentheses denotes the change in total expenditure. This can be decomposed into the expenditure-decreasing effect of lower drug prices and the expenditure-increasing effect of a higher quantity being reimbursed:

$$\begin{aligned} & (p_{b,D}^*(\gamma, \cdot) q_{b,D}^*(\gamma, \cdot) + p_{\beta,D}^*(\gamma, \cdot) q_{\beta,D}^*(\gamma, \cdot) - p_{b,D}(\gamma) q_{b,D}(\gamma)) \\ = & \underbrace{-(p_{b,D}(\gamma) - p_{b,D}^*(\gamma, \cdot)) q_{b,D}}_I + \underbrace{p_{b,D}^*(\gamma, \cdot) (q_{b,D}^*(\gamma, \cdot) - q_{b,D}(\gamma))}_{II} \\ & \quad + \underbrace{p_{\beta,D}^*(\gamma, \cdot) q_{\beta,D}^*(\gamma, \cdot)}_{III}. \end{aligned} \quad (5.17)$$

The negative part I of the decomposition exhibits the expenditure-decreasing effect from a lower price, while ignoring changes in quantity. The monopoly quantity $q_{b,D}$ sold and reimbursed under segmented markets is reimbursed based on a lower price, when parallel trade is possible. The positive parts II and III of the decomposition indicate the expenditure-increasing effect of

a higher quantity being sold and hence reimbursed under parallel trade. Due to lower prices, more is sold of the locally sourced version (part II) and in addition, also the parallel import (part III) is sold and reimbursed.

The effect of lower prices dominates the effect of a higher quantity and independent of the cost-sharing system in the source country S , parallel trade results in lower public pharmaceutical expenditure under coinsurance:

$$\frac{E_D^*(\gamma, \gamma)}{E_D(\gamma)} < 1, \quad \frac{E_D^*(\gamma, \delta)}{E_D(\gamma)} < 1. \quad (5.18)$$

Under indemnity insurance, reimbursement is independent of the drug price and the decrease of drug prices under from parallel trade has no expenditure-decreasing effect. The higher quantity sold results in an increase in public pharmaceutical expenditure associated with parallel trade:

$$\Delta E_D(\delta) = E_D^*(\delta, \cdot) - E_D(\delta) = \delta_D (q_{b,D}^*(\delta, \cdot) + q_{\beta,D}^*(\delta, \cdot) - q_{b,D}(\delta)) > 0, \quad (5.19)$$

since $q_{b,D}^*(\delta, \cdot) + q_{\beta,D}^*(\delta, \cdot) > q_{b,D}(\delta)$. As reimbursement is not linked to drug prices, health insurance do not benefit from lower prices under parallel trade.

5.4.2 Changes in Co-payments and Public Pharmaceutical Expenditure in the Source Country

Comparing the consequences of parallel trade for consumer co-payments and public pharmaceutical expenditure in the source country yields the following proposition:

Proposition 5.2. *In the source country S , parallel trade increases co-payments under both cost-sharing systems. Suppose that drug prices under segmented markets are identical under coinsurance and indemnity insurance. Then additional expenses for consumers are lower under coinsurance, independent of the cost-sharing system in the destination country D . Savings for health insurance from parallel trade are higher under indemnity insurance, regardless of the*

cost-sharing system in the destination country.

Proof. See Appendix C.3. □

Under both coinsurance and indemnity insurance, co-payments strictly increase in drug prices and consequently, price increases give rise to higher co-payments:

$$\begin{aligned}\Delta c_{b,S}(\gamma, \cdot) &= c_{b,S}^*(\gamma) - c_{b,S}(\gamma) = \gamma_S (p_{b,S}^*(\gamma, \cdot) - p_{b,S}(\gamma)) > 0, \\ \Delta c_{b,S}(\delta, \cdot) &= c_{b,S}^*(\delta) - c_{b,S}(\delta) = p_{b,S}^*(\delta, \cdot) - p_{b,S}(\delta) > 0.\end{aligned}\tag{5.20}$$

since $p_{b,S}^*(\gamma, \cdot) > p_{b,S}(\gamma)$ and $p_{b,S}^*(\delta, \cdot) > p_{b,S}(\delta)$, resp. Thus, consumers lose from parallel trade independent of the cost-sharing system.

Under which cost-sharing system the additional expense is higher is determined by two factors, similar to the link between competition effect and consumer savings in the destination country: First, the direct impact of the cost-sharing system for a given drug price now works in the opposite direction as compared to impact on savings in the destination country: Under indemnity insurance, consumers have to bear the full price increase. Under coinsurance, only the fraction γ_S of the price increase is incurred by consumers, the remaining is borne by health insurance. Second, price changes resulting from the increase of the wholesale price contribute to changes in co-payments. Here, the higher reduction in quantity due to the higher price elasticity under indemnity insurance tends to bring about a lower increase of the wholesale price.

As the change in the wholesale price is determined by both the competition effect and the double marginalization effect, thus by cost-sharing systems in both countries, relative co-payment changes are compared for a given cost-sharing system in the other country.

Again, I assume identical drug prices for both cost-sharing systems under segmented markets. This implies the following relationship between the coinsurance rate and the reimbursement amount under indemnity insurance:

$$p_{b,S}(\gamma) = \frac{1}{2\gamma_S} = \frac{1 + \delta_S}{2} = p_{b,S}(\delta) \iff \delta_S = \frac{1}{\gamma_S} - 1 \iff \gamma_S = \frac{1}{\delta_S + 1}.\tag{5.21}$$

Note that, similar to assuming identical drug prices in the destination country D implying high coinsurance rates, this standard of comparison for the source country requires rather high coinsurance rates due to the positive co-payment condition under indemnity insurance, see Appendix C.3 for details.

The change in co-payments is lower under coinsurance than under indemnity insurance, independent of the cost-sharing system in the destination country:

$$\frac{c_{b,S}^*(\gamma, \gamma)}{c_{b,S}(\gamma)} < \frac{c_{b,S}^*(\gamma, \delta)}{c_{b,S}(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1}, \quad \frac{c_{b,S}^*(\delta, \gamma)}{c_{b,S}(\gamma)} < \frac{c_{b,S}^*(\delta, \delta)}{c_{b,S}(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1}. \quad (5.22)$$

As a higher ratio of relative co-payments implies higher price increases, this corresponds to co-payments being increased less under coinsurance. Although the wholesale price is higher under coinsurance regardless of the cost-sharing system in the destination country ($w_S^*(\gamma, \gamma) > w_S^*(\gamma, \delta)$, $w_S^*(\delta, \gamma) > w_S^*(\delta, \delta)$), the direct effect of the cost-sharing system of consumers being insulated from part of the price increases dominates and co-payments are less increased.

The double marginalization effect induced by parallel trade has an ambiguous impact on total expenditure and public pharmaceutical expenditure. The higher drug price tends to increase expenditure, the lower quantity contributes to a reduction of quantity.

Under coinsurance, the change in public pharmaceutical expenditure associated parallel trade is given as:

$$\Delta E_S(\gamma) = E_S^*(., \gamma) - E_S(\gamma) = (1 - \gamma_S) (p_{b,S}^*(., \gamma) q_{b,S}^*(., \gamma) - p_{b,S}(\gamma) q_{b,S}(\gamma)), \quad (5.23)$$

of which the part in parentheses denotes the change in total expenditure. This can be decomposed into the expenditure-increasing effect of a higher drug price and the expenditure-decreasing

effect from a lower quantity:

$$\begin{aligned}
& (p_{b,S}^*(\cdot, \gamma) q_{b,S}^*(\cdot, \gamma) - p_{b,S}(\gamma) q_{b,S}(\gamma)) \\
= & \underbrace{(p_{b,S}^*(\cdot, \gamma) - p_{b,S}(\gamma)) q_{b,S}(\gamma)}_I - \underbrace{p_{b,S}^*(q_{b,S}(\gamma) - q_{b,S}^*(\cdot, \gamma))}_{II}. \tag{5.24}
\end{aligned}$$

The positive part I of the decomposition exhibits the expenditure-increasing effect of a higher drug price being the basis for reimbursement under parallel trade, while neglecting changes in quantity. The negative part II reflects the reduction of expenditure due to the lower quantity sold and reimbursed under parallel trade.

The effect of a lower quantity exceeds the effect of a higher drug price and independent of the cost-sharing system in the destination country D , parallel trade results in lower public pharmaceutical expenditure under coinsurance:

$$\frac{E_S^*(\gamma, \gamma)}{E_S(\gamma)} < 1, \frac{E_S^*(\delta, \gamma)}{E_S(\gamma)} < 1. \tag{5.25}$$

Under indemnity insurance, the change in public pharmaceutical expenditure associated parallel trade is given as:

$$\Delta E_S(\delta) = E_S^*(\cdot, \delta) - E_S(\delta) = \delta_S (q_{b,S}^* - q_{b,S}) < 0, \tag{5.26}$$

since $q_{b,S}^* < q_{b,S}$. Since reimbursement is not linked to drug prices, the drug price increase does not give rise to an expenditure-increasing effect. Instead, the decrease in quantity unambiguously decreases public pharmaceutical expenditure.

Comparing relative expenditure reductions under coinsurance and indemnity insurance for identical drug prices for both cost-sharing systems under segmented markets, public pharmaceutical expenditure is more reduced under indemnity insurance, independent of the cost-sharing

system in the destination country:

$$\frac{E_S^*(\gamma, \gamma)}{E_D(\gamma)} - \frac{E_S^*(\gamma, \delta)}{E_D(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1} > 0, \quad \frac{E_S^*(\delta, \gamma)}{E_S(\gamma)} - \frac{E_S^*(\delta, \delta)}{E_S(\delta)} \Big|_{\gamma_S = \frac{1}{\delta_S + 1}} > 0 \quad (5.27)$$

Under indemnity insurance, the drug-price increase under parallel trade has no expenditure-increasing effect, as reimbursement is independent of the drug price. Accordingly, expenditure is decreased more, if reimbursement is price-independent.

5.5 Conclusion

In this chapter, I have studied the consequences of parallel trade for health care systems. In particular, I have compared a coinsurance scheme and an indemnity insurance scheme with respect to changes of co-payments, i.e. out-of-pocket expenditure for patients, and public pharmaceutical expenditure.

The model suggests a strong link between the consequences of parallel trade and cost-sharing systems. Although, under both cost-sharing schemes, parallel trade results in price decreases in the destination country and a price increase in the source country, the cost-sharing scheme determines whether savings from parallel trade accrue and for whom. The extent to which price decreases (in the destination country) and price increases (in the source country) are passed on to consumers and health insurance and translate to co-payment and health expenditure reductions or increases depends on the reimbursement mechanism.

In the destination country, savings for patients from parallel trade occur under both systems, but are relatively higher under indemnity insurance. However, savings for health insurance from parallel trade occur only under coinsurance. Indemnity insurance fails to establish a link between the reimbursement and drug prices. The procompetitive effect of parallel trade does not benefit health insurance, but rather results in a growth of expenditure, as consumption increases due to lower drug prices. Consequently, maximizing savings for patients by adopting indemnity insurance occurs at the expense of public pharmaceutical expenditure. The higher incentive for

patients to buy the less expensive parallel import under indemnity insurance rather hurts health insurance than benefiting it. Applying coinsurance instead partially shifts savings from patients to health insurance: Savings for patients are relatively lower than under indemnity insurance, but also health insurance benefits from lower prices. This makes coinsurance a more attractive cost-sharing mechanism for health insurance; the failure to link prices to expenditure and full responsiveness to changes in quantity may constitute a reason why indemnity insurance is considered to be inappropriate as a cost-sharing instrument by regulatory bodies. Pure indemnity insurance is not observed in any member state of the European Union.

In the source country, the drug price increase following from the increase of the wholesale price results in additional expenses for consumers under parallel trade, but the associated reduction in quantity consumed benefits health insurance, as it dominates the effect of higher prices on public pharmaceutical expenditure. Again a conflict between minimizing negative consequences from parallel trade for patients and maximizing positive consequences for health insurance emerges. Under coinsurance, additional expenses for patients are relatively lower; under indemnity insurance, the reduction of health expenditure is relatively higher. Consequently, the distributive effects of parallel trade in the source country, too, are determined by the cost-sharing scheme. Considering that regulatory bodies in source countries should prefer indemnity insurance to minimize public pharmaceutical expenditure, it remains unexplained, why indemnity insurance is not applied in the respective source countries of parallel imports.

Accordingly, if minimizing public pharmaceutical expenditure is the dominant health policy objective, coinsurance should be the cost-sharing instrument of choice in the destination country and indemnity insurance in the source country. If minimizing financial exposure of patients and maximizing access to pharmaceuticals is the prevailing health policy objective, indemnity insurance is the appropriate cost-sharing scheme in the destination country and coinsurance in the source country. Thus, these health policy objectives should be carefully balanced.

In the European Union, health policy, including the general organization of health care systems as well as cost-sharing systems, is in the national competence of member states. However, a

harmonization of health care systems with respect to cost-sharing systems would not make sense, as this chapter illustrates. Under parallel trade reimbursement systems have different effects in source and destination countries of parallel imports. Thus, a harmonization of reimbursement systems would result in welfare losses in source and/or destination countries.

Moreover, the degree of product differentiation between the locally sourced version of the drug and the parallel import in the destination country is a source of conflict between the welfare consequences in the two countries: If both versions of the drug were considered equivalent, the competition effect in the destination country would be maximized. Perfect competition would reduce public expenditure, minimize financial exposure of patients and maximize access to pharmaceuticals. Consequently, the main health policy challenge in the destination country would be to reduce the degree of product differentiation. In addition, information of the public and mandatory substitution as an instrument for patients to gather experience with parallel imports could be an appropriate way of reducing the degree of subjective product differentiation. On the contrary, in the origin country, the perceived equivalence of the locally sourced version and the parallel import would at the same time also maximize the wholesale price and with it the double marginalization effect. Although this would reduce public expenditure, it would also maximize financial exposure of patients and minimize access to pharmaceuticals. Thus, from the perspective of the origin country, there is also some interest in maintaining a certain degree of product differentiation. This could be achieved by imposing country-specific requirements with respect to the packaging, which would result in an increased difference in appearance between the two versions of the drug in the destination country. This corresponds to the manufacturer's incentive: Both the competition effect and the double marginalization effect reduce profit and as raising the degree of product differentiation weakens both effects, it raises profit.

Generally, this model considers stylized, i.e. simplified versions of cost-sharing systems. In the Scandinavian countries, e.g. combinations of different cost-sharing systems (deductibles and coinsurance) are applied. In addition, there is a tendency for cost-sharing systems to become more complex. For instance, coinsurance rates may depend on drug classes, exemptions may be

added etc. The stylized versions of cost-sharing systems in this chapter allow me to compare different systems on a basic level. A specific and precise comparison of cost-sharing systems of different member states would have to take the detailed design of cost-sharing systems into account.

Chapter 6

Pharmaceutical Regulation at the Wholesale Level and Parallel Trade

Abstract

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

6.1 Introduction

This chapter studies the effect of maximum wholesale margins and mandatory rebates on drug prices, quantities, and public pharmaceutical expenditure.

This analysis is motivated by the observation that pharmaceutical regulation and parallel trade are interdependent. Pharmaceutical parallel trade occurs in highly regulated markets. The continuous increase in public health expenditure in many countries over the last decades has induced a considerable number of government interventions (Maynard & Bloor, 2003). Consequently, pharmaceuticals markets are characterized by a variety of regulatory instruments that are partly overlapping and impede each other (see Espin & Rovira 2007 for an overview of regulatory interventions in the European Union).

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

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

information which price level is still profitable for manufacturers and regulatory authorities may adjust maximum prices downwards. At last, in countries with a strong pharmaceutical industry, where pharmaceutical regulation also takes industrial policy goals into account, regulation may respond to parallel trade, as it reduces manufacturers' profits.

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

The effects on drug prices and quantities in destination and source countries of parallel

imports are studied in a two-country model following Maskus & Chen (2002) and Chen & Maskus (2005).

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

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

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

The rest of the chapter is organized as follows. Section 6.2 studies the equilibrium without regulation, section 6.3 examines the equilibrium with maximum wholesale margins. Section 6.4 analyzes mandatory rebates as policy alternative, section 6.5 concludes.

6.2 Equilibrium without Regulation

As a benchmark consider the case of unregulated markets, when the manufacturer and both intermediaries can set prices freely. If parallel trade is allowed, the manufacturer's pricing

decisions – the wholesale price w_D^* charged the intermediary I_D and the wholesale price w_S^* charged the intermediary I_S – are interdependent.

The manufacturer's profit is given as

$$\pi_M^* = \underbrace{w_D^* \left(1 - \frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_{\pi_{w_b,D}^*} + \underbrace{w_S^* (1 - \gamma_S p_{b,S}^*)}_{\pi_{w_b,S}^*} + \underbrace{w_S^* \left(\frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{w_\beta}^*} + \phi_D^* + \phi_S^*, \quad (6.1)$$

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

The intermediaries' profits are given as

$$\pi_{I_D}^* = \underbrace{(p_{b,D}^* - w_D^*) \left(1 - \frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_{\pi_{b,D}^*} - \phi_D^*, \quad (6.2)$$

$$\text{and } \pi_{I_S}^* = \underbrace{(p_{b,S}^* - w_S^*) (1 - \gamma_S p_{b,S}^*)}_{\pi_{b,S}^*} + \underbrace{(p_{\beta,D}^* - w_S^*) \left(\frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{\beta,D}^*} - \phi_S^*, \quad (6.3)$$

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

In country D , the domestic intermediary I_D maximizes (6.2) with respect to $p_{b,D}^*$ which yields the first order condition

$$\left(1 - \frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right) + (p_{b,D}^* - w_D^*) \underbrace{\left(-\frac{\gamma_D}{\tau} \right)}_{\frac{\partial q_{b,D}^*}{\partial p_{b,D}^*}} = 0 \quad (6.4)$$

and the best response function

$$p_{b,D}^* = \frac{1}{2\gamma_D} (\tau + p_{\beta,D}^* \gamma_D + \gamma_D w_D^*). \quad (6.5)$$

The foreign intermediary I_S maximizes (6.3) with respect to $p_{\beta,D}^*$ which yields the first order condition

$$\left(\frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)} \right) + (p_{\beta,D}^* - w_S^*) \underbrace{\left(-\frac{\gamma_D}{\tau} - \frac{\gamma_D}{(1-\tau)} \right)}_{\frac{\partial q_{\beta,D}^*}{\partial p_{\beta,D}^*}} = 0 \quad (6.6)$$

and the best response function

$$p_{\beta,D}^* = \frac{1}{2} (w_S^* + p_{b,D}^* (1-\tau)). \quad (6.7)$$

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

In country S , the intermediary maximizes (6.3) with respect to $p_{b,S}^*$. The first order condition to this maximization problem is

$$(1 - \gamma_S p_{b,S}^*) + (p_{b,S}^* - w_S^*) (-\gamma_S) = 0, \quad (6.8)$$

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

With fixed fees of

$$\phi_D^* = \frac{(2\tau + \gamma_D w_S^* - \gamma_D w_D^* (\tau + 1))^2}{\underbrace{\tau \gamma_D (\tau + 3)^2}_{\pi_{b,D}^*}} \quad (6.9)$$

$$\text{and } \phi_S^* = \underbrace{\frac{(1 - w_S^* \gamma_S)^2}{4\gamma_S}}_{\pi_{b,S}^*} + \underbrace{\frac{(\gamma_D w_D^* (1 - \tau) + (1 - \tau) \tau - \gamma_D w_S^* (1 + \tau))^2}{\tau \gamma_D (1 - \tau) (\tau + 3)^2}}_{\pi_{\beta,D}^*} \quad (6.10)$$

the manufacturer extracts the intermediaries' total profits.

Substituting (6.9), (6.10), and equilibrium prices into (6.1) and maximizing with respect to w_D^* and w_S^* gives the wholesale prices $w_D^* = \frac{(1-\tau)(2\tau+\gamma_D w_S^*)}{\gamma_D(3\tau+1)}$ and $w_S^* = \frac{2(1-\tau)(5\tau-\tau^2+2\gamma_D w_D^*(1-\tau))}{4\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2}$. Wholesale prices mutually reinforce one another; a higher wholesale price in the destination country, w_D^* induces a higher wholesale price in the source country, w_S^* and vice versa.

Equilibrium wholesale prices are given as:

$$w_D^* = \frac{2(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \text{ and } w_S^* = \frac{2(1-\tau)}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)}. \quad (6.11)$$

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

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

Consider Figure 6.1 for a visualization of this effect. Dashed lines are best response functions for $w_D^* = w_S^* = 0$, yielding retail price equilibrium A. Solid lines are best response functions for $w_D^* > 0$, $w_S^* > 0$, yielding retail price equilibrium B. The increase of wholesale prices (from

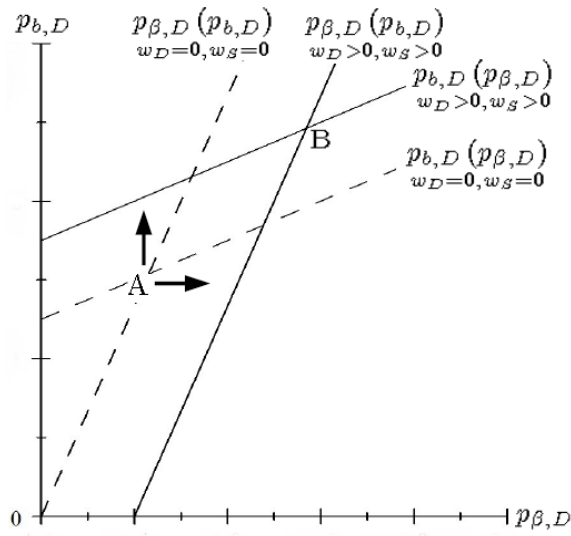


Figure 6.1: Best Response Functions for $w_D^* = w_S^* = 0$ and $w_D^* > 0, w_S^* > 0$.

$w_D^* = w_S^* = 0$ to $w_D^* > 0, w_S^* > 0$) shifts the retail price equilibrium from A to B, inducing higher retail prices. Note that also the increase of only one wholesale price would result in higher retail prices.

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

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

At the same time, an increase of w_S^* also increases the drug price and decreases the quantity

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

The first order conditions illustrate the effects of the choice of the wholesale price on the manufacturer's profit, see Appendix D.1 for details.

Equilibrium drug prices are

$$p_{b,D}^* = \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}, p_{\beta,D}^* = \frac{(1-\tau)(2\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \text{ and} \quad (6.12)$$

$$p_{b,S}^* = \frac{4\gamma_D + 3\gamma_S(1-\tau^2)}{2\gamma_S(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (6.13)$$

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

Equilibrium quantities are

$$q_{b,D}^* = \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)}, q_{\beta,D}^* = \frac{(1-\tau)\gamma_S}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)} \text{ and} \quad (6.14)$$

$$q_{b,S}^* = \frac{4\gamma_D - \gamma_S(1-3\tau)(1-\tau)}{2(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (6.15)$$

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

Under parallel trade, the manufacturer may increase the profit allocated to the destination country. Competition from parallel trade has a profit-decreasing effect, but the strategic effect of

reducing competition by increasing the wholesale prices and inducing higher retail prices works in the opposite direction. If the effect of a higher wholesale profit from sales as parallel imports and a higher fixed fee extracted from intermediary I_S exceeds the effect of competition in the destination country¹, the profit allocated to the destination country is higher under parallel trade. The strategic effect of reducing competition by increasing wholesale prices is crucial; the manufacturer's profit is always lower under direct sales in the destination country. The profit earned in the source country is always lower due to the double marginalization effect with a higher drug price and a lower quantity sold. The total effect of parallel trade on the manufacturer's profit depends on the relative size of these two effects and with it on the price elasticities in both countries (i.e. coinsurance rates) and the substitutability of both products (i.e. the degree of vertical product differentiation).

6.3 Equilibrium with Maximum Wholesale Margins

In this model, both the manufacturer and each intermediary have a monopoly position, allowing them to set prices freely. The vertical separation in an imperfectly competitive market results in inefficient successive markups by both the manufacturer and each intermediary. Under segmented markets, the manufacturer solves this problem by setting wholesale prices equal to marginal cost (and extracting the profits of the intermediaries via the fixed fee). Under parallel trade, the manufacturer exploits the strategic effect of exclusive territories in the destination country and raises wholesale prices to induce higher retail prices and reduce competition. This results in a double marginalization effect in the source country. Consequently, in both countries, drug prices are higher than they would be if the manufacturer sold directly². In addition, in the source country, the drug price is higher than in the absence of parallel trade.

¹That is a lower wholesale profit from sales as locally sourced and a higher fixed fee extracted from intermediary I_D .

²Note that although the price of the locally sourced version of the drug in country D is lower under parallel trade, direct sales would bring about an even lower price, avoiding the impact of double marginalization. The price change associated with the switch from segmented markets to parallel trade is the net of a competition and a double marginalization effect.

This sections studies maximum wholesale margins as a possible regulatory intervention addressing the problem of double marginalization by restricting the markups charged by intermediaries. This instrument is applied in virtually all European countries. Commonly, intermediaries are granted a certain (percentage) markup on the wholesale price³. In a strict specification, this cannot be described in this model⁴. Therefore, I model maximum wholesale margins instead as a restriction of the markup charged under free pricing. In country j , intermediaries may charge a fraction $(1 - \mu_j)$ of the markup under free pricing:

$$\begin{aligned}
p_{b,D}^{*\mu} &= w_D^{*\mu} + (1 - \mu_D) m_{b,D}^*, \text{ with } m_{b,D}^* = p_{b,D}^* - w_D^*, \\
p_{\beta,D}^{*\mu} &= w_S^{*\mu} + (1 - \mu_D) m_{\beta,D}^*, \text{ with } m_{\beta,D}^* = p_{\beta,D}^* - w_S^*, \\
\text{and } p_{b,S}^{*\mu} &= w_S^{*\mu} + (1 - \mu_S) m_{b,S}^*, \text{ with } m_{b,S}^* = p_{b,S}^* - w_S^*.
\end{aligned} \tag{6.16}$$

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

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

6.3.1 Maximum Wholesale Margin Regulation in the Destination Country

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

³In all European countries except for Italy maximum wholesale markups are defined in terms of wholesale prices.

⁴Note that in the symmetric equilibrium regulatory instruments of both countries mutually offset one another. This implies that a binding restriction on markups depends on the restriction in the respective other countries. Therefore, a symmetric equilibrium with both countries effectively restricting markups can only exist, if restriction on markups are scaled in terms of the markup under free pricing.

In the destination country, both intermediaries are forced to price at marginal cost, $p_{b,D}^{*\mu} = w_D^{*\mu}$ and $p_{\beta,D}^{*\mu} = w_S^{*\mu}$. This implies that intermediaries make zero profits in the destination country, the corresponding fixed fees ($\phi_D^{*\mu}$ and the part of $\phi_S^{*\mu}$ associated with parallel trade) are also zero. In the source country, intermediary I_S may set the drug price freely.

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

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

$$w_D^{*\mu} - w_D^* > 0, w_S^{*\mu} - w_S^* > 0. \quad (6.17)$$

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

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

$$p_{b,D}^{*\mu} - p_{b,D}^* < 0. \quad (6.18)$$

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

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

$$p_{\beta,D}^{*\mu} - p_{\beta,D}^* > 0, \text{ if } \gamma_S < \frac{2}{(1-\tau)}\gamma_D. \quad (6.19)$$

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

$$q_{b,D}^{*\mu} - q_{b,D}^* > 0, q_{\beta,D}^{*\mu} - q_{\beta,D}^* < 0. \quad (6.20)$$

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

$$p_{b,S}^{*\mu} - p_{b,S}^* > 0, q_{b,S}^{*\mu} - q_{b,S}^* < 0. \quad (6.21)$$

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

6.3.2 Maximum Wholesale Margin Regulation in the Source Country

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

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

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

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

$$w_D^{*\mu} - w_D^* > 0, \quad w_S^{*\mu} - w_S^* > 0. \quad (6.22)$$

The profit from selling to the sourced country is maximized for a retail price of $p_{b,S}^* = \frac{1}{2\gamma_S}$. Under segmented markets, the manufacturer induces this retail price by setting the wholesale price w_S^* to zero, since the intermediary surcharges a monopoly markup. Thus, if the intermediary's monopoly markup disappears and the intermediary sets the price to marginal cost, a wholesale price of $w_S^{*\mu} = \frac{1}{2\gamma_S}$ would maximize profits from selling in the source country. However, when setting wholesale prices, the manufacturer also considers the strategic effect of higher wholesale prices for the destination country. Consequently, it may be profitable to raise $w_S^{*\mu}$ above this profit-maximizing level of $w_S^{*\mu} = \frac{1}{2\gamma_S}$ and reduce competition from parallel trade, increasing profits from selling in the destination country, while reducing profits from selling in the source country by decreasing the quantity sold⁵. Thus, the manufacturer increases wholesale

⁵Note that an increase of the wholesale price above the profit maximizing $w_S^{*\mu} = \frac{1}{2\gamma_S}$ induces a higher reduction of quantity and thus a higher reduction of profits than an increase of the wholesale price above zero under free pricing.

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

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

$$p_{b,S}^{*\mu} - p_{b,S}^* < 0. \quad (6.23)$$

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

$$q_{b,S}^{*\mu} - q_{b,S}^* > 0. \quad (6.24)$$

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

$$p_{b,D}^{*\mu} - p_{b,D}^* > 0, \quad p_{\beta,D}^{*\mu} - p_{\beta,D}^* > 0. \quad (6.25)$$

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

$$q_{b,D}^{*\mu} - q_{b,D}^* < 0. \quad (6.26)$$

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

$$q_{\beta,D}^{*\mu} - q_{\beta,D}^* < 0, \quad q_{\beta,D}^{*\mu} > 0 \text{ if } \gamma_S > \frac{1}{(1-\tau)}\gamma_D. \quad (6.27)$$

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

marginalization effect is mitigated, the drug price is lower.

6.3.3 Symmetric Equilibrium under Maximum Wholesale Margins

Consider now the symmetric equilibrium with both countries restricting markups of intermediaries.

If the destination country restricts markups, the manufacturer cannot exploit the strategic effect of exclusive territories and increases wholesale prices to reduce competition directly. This effect is limited by the double marginalization effect in the source country. If the source country also restricts markups, the double marginalization effect is mitigated. This increases the incentive to raise wholesale prices to reduce competition. The total effect on drug prices in the destination country is ambiguous, as the further increase of wholesale prices due to the restriction of markups is not offset by stricter markup limits in the destination country. Also in the source country, the total effect on the drug price is ambiguous. The increase of the wholesale price emerging from the restriction of markups in the destination country is not compensated by a tougher restriction of markups in the source country.

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

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

Proposition 6.1 summarizes the effect of maximum wholesale margins on drug prices and quantities:

Proposition 6.1. *A restriction of markups in the destination country i) reduces competition*

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

6.4 Policy Alternative: Mandatory Rebates

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

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

When being forced to grant discounts and subject to free pricing at the same time, manufacturers may simply avoid discounts by increasing prices. Therefore, mandatory rebates are applied in combination with price freezes, which prevent strategic price increases in response to discounts. Here, price freezes apply to wholesale prices only, as the regulatory intervention is intended to be limited with respect to one level only⁶.

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

$$w_D^{*\psi} = (1 - \psi_D) w_D^* \quad (6.28)$$

and in country S to:

$$w_S^{*\psi} = (1 - \psi_S) w_S^*. \quad (6.29)$$

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

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

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

6.4.1 Mandatory Rebates in the Destination Country

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

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

$$w_S^{*\psi} - w_S^* > 0. \tag{6.30}$$

Both wholesale prices are lower than under free pricing and accordingly, also drug prices are

lower in the destination country:

$$p_{b,D}^{*\psi} - p_{b,D}^* < 0, p_{\beta,D}^{*\psi} - p_{\beta,D}^* < 0. \quad (6.31)$$

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

$$q_{b,D}^{*\psi} - q_{b,D}^* > 0. \quad (6.32)$$

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

$$q_{\beta,D}^{*\psi} - q_{\beta,D}^* < 0, q_{\beta,D}^{*\psi} > 0 \text{ if } \gamma_S > \frac{2}{(3 + \tau)} \gamma_D. \quad (6.33)$$

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

$$p_{b,S}^{*\psi} - p_{b,S}^* < 0, q_{b,S}^{*\psi} - q_{b,S}^* > 0. \quad (6.34)$$

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

6.4.2 Mandatory Rebates in the Source Country

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

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

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

$$p_{b,S}^{*\psi} - p_{b,S}^* < 0, q_{b,S}^{*\psi} - q_{b,S}^* > 0. \quad (6.35)$$

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

$$w_D^{*\psi} - w_D^* < 0. \quad (6.36)$$

This reduces both drug prices:

$$p_{b,D}^{*\psi} - p_{b,D}^* < 0, p_{\beta,D}^{*\psi} - p_{\beta,D}^* < 0. \quad (6.37)$$

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

$$q_{b,D}^{*\psi} - q_{b,D}^* < 0, q_{\beta,D}^{*\psi} - q_{\beta,D}^* > 0. \quad (6.38)$$

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

6.4.3 Symmetric Equilibrium under Mandatory Rebates

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

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

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

Proposition 6.2 summarizes the effect of mandatory rebates:

Proposition 6.2. *A mandatory rebate on the wholesale price in the destination country i) reduces competition from parallel trade by reducing the relative price difference between both versions in the destination country, but reduces drug prices and ii) mitigates the double marginalization effect by decreasing the drug price in the source country. A mandatory rebate on the wholesale price in the source country i) mitigates the double marginalization effect by decreasing the drug price in the source country and ii) decreases the drug prices and intensifies competition from parallel trade in the destination country. Regulatory instruments of both countries mutually reinforce each other with respect to drug prices.*

6.5 Conclusion

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

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

Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. Restrictions of pricing by intermediaries in the destination country shift the competition-reducing effect from retail prices to wholesale prices, which improves the effect of price increases. Under maximum wholesale margins, intermediaries do not absorb wholesale price increases partly, but pass them on to retail prices completely. Restrictions of pricing by intermediaries in the source country mit-

igate the double marginalization effect, allowing the manufacturer to focus on exploiting the strategic effect in the destination country. Since the manufacturer cannot offset the effect of markups completely, restrictions of pricing also reduce drug prices in the respective country. But via the increase of wholesale prices, they increase drug prices in the respective other country. In the symmetric equilibrium with both countries applying maximum wholesale margins, regulatory instruments exhibit an offsetting effect. Therefore, this regulatory instrument may be inappropriate in a setting where markets are integrated by parallel trade.

Mandatory rebates may be an alternative, they restrict wholesale prices set by the manufacturer. A restriction of the wholesale price in the destination country allows the manufacturer to reduce competition from parallel trade more easily, since a shift in demand from the parallel import to the locally sourced version requires a lower wholesale price for the foreign manufacturer than under free pricing. At the same time, drug prices in both countries are reduced. A restriction of the wholesale price in the source country intensifies competition from parallel trade. It also reduces drug prices in both countries. In the symmetric equilibrium with both countries applying mandatory rebates, regulatory instruments exhibit a reinforcing effect. This is, the regulation of wholesale prices has a positive externality.

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

Chapter 7

Conclusion

7.1 Summary of Results

In chapter 2, I have compared two instruments of pharmaceutical regulation, a maximum price system and a reference price system with respect to their performance in achieving certain health policy objectives. For identical price reductions of the brand-name drug, the lower reimbursement amount under the reference price system results in lower health expenditure, but higher financial exposure of patients. Access to pharmaceuticals is better under the maximum price system, although the generic quantity is higher under the reference price system. In the aggregate, consumer surplus for brand-name users is higher under the maximum price system, whereas consumer surplus for generic users is higher under the reference price system. Different from the maximum price system, the reference price system reduces the brand-name price premium and brings about a higher generic quantity. Consequently, there is a trade-off between important health policy objectives: The reference price system may be more appropriate to reduce public pharmaceutical expenditure or stimulate competition, but the maximum price system performs better for distributive objectives of limiting financial exposure of patients and guaranteeing access to pharmaceuticals.

In chapter 4, 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 in the source country. Under market integration through parallel trade, national decisions on coinsurance rates affect the trade-off between the double marginalization effect and the competition effect. This makes the manufacturer adjust the wholesale price and externalities occur. An increase in the coinsurance rate in the destination country mitigates the double marginalization effect in the source country; an increase in the coinsurance rate in the source country reinforces the competition effect from parallel trade in the destination country.

In chapter 5, I have analyzed the consequences of parallel trade for health care systems. In particular, I have compared a coinsurance scheme and an indemnity insurance scheme with respect to changes of co-payments and public pharmaceutical expenditure. Under both cost-sharing schemes, parallel trade results in price decreases in the destination country and a price increase in the source country. The cost-sharing scheme determines whether savings from parallel trade accrue and for whom. In the destination country, savings for patients from parallel trade occur under both systems, but are relatively higher under indemnity insurance. However, only under coinsurance, lower drug prices reduce public pharmaceutical expenditure. In the source country, the drug price increase results in additional expenses for consumers under parallel trade, but the associated reduction in quantity consumed benefits health insurance. Under coinsurance, additional expenses for patients are relatively lower; under indemnity insurance the reduction of health expenditure is relatively higher.

In chapter 6 I have studied the effect of pharmaceutical regulation at the wholesale level, namely maximum wholesale margins and mandatory rebates as a policy alternative. Parallel trade provides the manufacturer with the possibility to exploit the strategic effect of exclusive territories in the destination country. By raising wholesale prices in both countries he induces higher retail prices in the destination country and reduces competition from parallel

trade. Maximum wholesale margins enhance the manufacturer's ability to reduce competition from parallel trade in the destination country by increasing wholesale prices. In a symmetric equilibrium, maximum wholesale margins of both countries partly offset each other, mandatory rebates reinforce each other.

7.2 Policy Implications

As stated in the introduction, this thesis addresses two recurring conflicts of pharmaceutical market regulation at the European level. My results have the following implications for health policy:

First, a conflict between different health policy objectives is bound to occur. Chapter 2 illustrates that depending on the specific health policy objective, either the maximum price system or the reference price system is more appropriate. Chapter 4 and chapter 5 emphasize that national decisions on health policy – co-payment changes or the choice of reimbursement systems – may induce a conflict between changes in financial exposure of patients and changes in public pharmaceutical expenditure. Chapter 6 shows that price regulation may reduce drug prices, but also restricts competition at the same time.

These results show that the choice between regulatory instruments implies a choice between different health policy objectives. Vice versa, the decision for a specific health policy objective determines the appropriate regulatory instrument. Furthermore, this implies that a careful weighting of different health policy objectives is necessary. Given that neither expenditure reduction nor distributive objectives should be neglected, regulatory decisions are not isolated, but might induce further interventions. For example, a government intervention with the aim of reducing public pharmaceutical expenditure might raise the issue of financial exposure of patients and access to pharmaceuticals. A potential follow-up intervention could be to subsidize pharmaceuticals for low-income groups, as through co-payment exemptions. On this account, several EU member states such as Belgium, Finland, France, Germany, and the Netherlands have introduced a ceiling for all co-payments (Mossialos & Le Grand, 1999). In Germany, patients

are exempt from co-payments beyond the limit of 2 % of gross income (§ 62 German Social Security Code V). On the other side, guaranteeing a broad access to pharmaceuticals while limiting financial exposure of patients might bring about higher expenditure. This raises the question of financing the health care system.

In general, a decreasing degree of (perceived) product differentiation between brand-names and generics or locally sourced versions and parallel imports may alleviate this conflict. Closer substitutes (from a patient's perspective) intensify the competition between different suppliers (interbrand and intrabrand), reducing drug prices. This reduces public pharmaceutical expenditure, limits financial exposure of patients, and improves access to pharmaceuticals. Consequently, the "invisible hand of competition" is preferable to the "visible hand of regulators", as stimulating competition in pharmaceutical markets mitigates the conflict between reducing public pharmaceutical expenditure and limiting financial exposure of patients. In practice, a major health policy challenge is to reduce the degree of product differentiation. With respect to objective product differentiation between brand-name drugs and generics, this corresponds to reducing permitted bandwidths of equivalence of additives and the degree of bioavailability. As to perceived quality differences, either between brand-names and generics or locally sourced versions and parallel imports, information of the public and mandatory substitution as means for patients to gather experience with generics or respectively parallel imports could help to reduce subjective product differentiation.

Second, under market integration national decisions generate externalities. Chapter 4 illustrates that nationally determined changes in coinsurance rates result in externalities and affect drug prices and public pharmaceutical expenditure in other countries. Chapter 5 shows that national competence in health policy allows member states to design their cost-sharing systems according to their status as a source or destination country of parallel trade and thereby maximize benefits from parallel trade. Chapter 6 emphasizes that the choice of regulatory instruments determines the type of externalities under market integration; depending on what market stage is restricted in pricing, pricing restrictions of two countries may be strategic complements or

substitutes.

This suggests that a coordination of coinsurance rates makes sense, but a harmonization of cost-sharing systems may not be desirable. The decision on cost-sharing systems should remain in the member states' competence. This allows them to design their cost-sharing systems according to their specific characteristics. But member states should coordinate decisions that might affect other countries, such as changes in coinsurance rates. Originally, coinsurance rates are intended to restrict moral hazard in utilization of health services. This does not require frequent changes in coinsurance rates. Also, in order to incentivize patients to purchase lower-priced alternatives, such as generics or parallel imports, a price-dependent design might be adequate.

On the contrary, a harmonization of pricing regulation at the wholesale level makes sense, as addressing the manufacturer is clearly preferable. Here, as the Treaty on the European Union and the Treaty on the Functioning of the European Union provide for national competence of member states in determining health policy, member states may apply the so called Open Method of Coordination. For wholesale regulation, externalities of national decisions of member states are unavoidable. But by the harmonization of pricing restrictions, positive externalities can be exploited.

Another type of externality may arise from efforts to reduce the perceived degree of product differentiation between locally sourced versions and parallel imports. This would maximize the competition effect of parallel trade and with it, as discussed above, reduce public expenditure, minimize financial exposure of patients and maximize access to pharmaceuticals in the destination country. But on the contrary, in the source country, the perceived equivalence of the locally sourced version and the parallel import would also maximize the wholesale price and with it the double marginalization effect. Although this would reduce public expenditure, it would also maximize financial exposure of patients and minimize access to pharmaceuticals. Thus, from the perspective of the source country, there is also some interest in maintaining a certain degree of product differentiation. This could be achieved by imposing country-specific requirements with

respect to the packaging, which would result in an increased difference in appearance between the two versions of the drug in the destination country. Thus, the perceived degree of product differentiation poses a problem with respect to the welfare consequences of parallel trade for destination and source countries.

7.3 Future Research

Based on this thesis, two avenues of future research come to mind:

First, the effect of the manufacturer's contract choice can be included in the analysis based on the framework presented in chapter 3. Whether the manufacturer may employ competition-reducing vertical restraints or is limited in his contract freedom, determines the consequences of parallel trade. On the one hand, in this model, the manufacturer optimizes the contract with the intermediary by adopting a two-part tariff and then parallel trade occurs as a distortion. As a consequence, the previously optimized contract is no longer profit-maximizing and the manufacturer loses from parallel trade. Given that the manufacturer may not solve the double marginalization problem by choosing an appropriate contract, parallel trade may even benefit the manufacturer. On the other hand, vertical restraints adopted by the manufacturer may determine the extent of parallel trade as they might help the manufacturer to block parallel trade.

Second, the dynamic effect of pharmaceutical regulation and parallel trade is subject to further research. Both reduce the manufacturer's profits and might lower investment or inhibit entry. Here, the static gains of regulation or competition by parallel trade have to be traded off against the potential dynamic losses of lower quality, lower innovation or reduced entry.

Bibliography

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.

Brekke, K. R., Königbauer, I. & Straume, O. R. (2006): Reference Pricing of Pharmaceuticals, CESifo Working Paper No. 1825.

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.

Cabrales, A. (2003): Pharmaceutical Generics, Vertical Product Differentiation, and Public Policy, UPF Working Paper No. 662.

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

Danzon, P. M. (1997): *Pharmaceutical price regulation: national policies versus global interests*. Washington, D.C..

Danzon, P. M. (2001): Reference Pricing: Theory and Evidence, in: López-Casanova, G. & Jönsson, B. (eds.): *Reference Pricing and Pharmaceutical Policy: Perspectives on Economics and Innovation*, Barcelona, 86 - 126.

Danzon P. M., Wang, Y. R. & Wang, L. (2005): The Impact of Price Regulation on the Launch Delay of New Drugs: A Study of Twenty-Five Major Markets in 1990s, *Health Economics* 14, 269 - 292.

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

EFPIA (2011): The pharmaceutical industry in figures, <http://www.efpia.eu/content/default.asp?PageID=559&DocID=11586>.

Ellison, S. F., Cockburn, I., Griliches, Z. & Hausman, J. (1998): Characteristics of demand for pharmaceutical products: an examination of four cephalosporins, *RAND Journal of Economics* 28, 426 - 446.

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

Espin, J. & Rovira, J. (2007): Analysis of differences and commonalities in pricing and reimbursement systems in Europe, study funded by DG Enterprise and Industry of the European Commission; Final Report, <http://bit.ly/JcNCBy>.

Ess, S. M., Schneeweiss, S. & Szucs, T. D. (2003): European Healthcare Policies for Controlling Drug Expenditure, *Pharmacoeconomics* 21, 89-103.

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

European Commission (2009): Pharmaceutical Sector Inquiry Preliminary Report, http://ec.europa.eu/competition/sectors/pharmaceuticals/inquiry/communication_en.pdf.

Frank, R. G. & Salkever, D. S. (1992): Pricing Patent Loss and the Market for Pharmaceuticals, *Southern Economic Journal* 59, 195 - 175.

- Frank, R. G. & Salkever, D. S. (1997): Generic entry and the pricing of pharmaceuticals, *Journal of Economics & Management Strategy* 6, 75 - 90.
- Gaither, C. A., Kirking, D. M.; Ascione F. J. & Welage, L. S. (2001): Consumers' views on generic medications, *Journal of the American Pharmaceutical Association* 41, 729-736.
- Ganslandt, M. & Maskus, K. E. (2007): Vertical distribution, parallel trade, and price divergence in integrated markets. *European Economic Review* 51, 943-970.
- Glynn, D. (2009): The effects of parallel trade on affordable access to medicines. *Eurohealth* 15, 1-5.
- Grabowski, H. & Vernon J. (1992): Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act, *Journal of Law and Economics* 35, 331 - 350.
- 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.
- Grootendorst, P. (2006): Prescription drug insurance and reimbursement, in: Jones, A. (ed.): *The Elgar Companion to Health Economics*, Cheltenham, 114-125.
- Grossman, G. M. & Lai, E. L.-C. (2008): Parallel imports and price controls. *The Rand journal of economics* 39, 378-402.
- 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.
- Hopf, G. (2002): Arzneimittel: „Aut idem“ in Kraft – aber nicht so schnell, in: *Deutsches Ärzteblatt* 99(11), A-680 / B-556 / C-524.
- Hurley, J. (2001): An Overview of the Normative Economics of the Health Sector, in: Culyer, A. & Newhouse, J. (eds.): *Handbook of Health Economics*, Amsterdam, 55-118.

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.

Kanavos, P., Vandonos, S., Irwin, R., Nicod, E. & Casson, M. (2011): Differences in costs of and access to pharmaceutical products in the EU. Study requested by the European Parliament's Committee on Environment, public pharmaceutical and Food Safety, Brussels.

Kanavos, P. & Costa-Font, J. (2005): Pharmaceutical parallel trade in Europe: stakeholders and competition effects. *Economic Policy* 20, 751-798.

Köksal, M. Y. (2009): Reference pricing: making parallel trade pharmaceuticals work. University of Gothenburg Economic Working Paper No. 376, http://gupea.ub.gu.se/bitstream/2077/20431/4/gupea_2077_20431_4.pdf.

Königbauer, I. (2004): Die Auswirkung von Parallelimporten auf die optimale Regulierung von Arzneimittelpreisen. *Vierteljahreshefte zur Wirtschaftsforschung* 73, 592 - 604.

Kong, Y. (2008): Competition between brand-name and generics-analysis on pricing of brand-name pharmaceutical, *Health economics* 18, 591-606.

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

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.

Linnosmaa I., Karhunen T. & Vohlonen I. (2003): Parallel importation of pharmaceuticals in Finland: effects on markets and expenditures. *Pharmaceutical Development and Regulation*, 67-74.

- López-Casanova, G. & Puig-Junoy, J. (2000): Review of the Literature on Reference Pricing, *Health Policy* 54, 87 - 123.
- Magazzini, L., Pammolli, F. & Riccaboni, M. (2004): Dynamic Competition in Pharmaceuticals. Patent Expiry, Generic Penetration, and Industry Structure, *European Journal of Health Economics* 5, 175 - 82.
- Maskus, K. E. (2000a): Intellectual property rights in the global economy, *Institute for International Economics*.
- Maskus, K. E. (2000b): Parallel imports. *The world economy*, 23, 1269-1284.
- Maskus, K. E. & Chen, Y. (2002): Parallel imports in a model of vertical distribution: theory, evidence, and policy. *Pacific economic review*, 7, 319-334.
- Maynard, A. & Bloor, K. (2003): Dilemmas In Regulation Of The Market For Pharmaceuticals, *Health Affairs* 22, 31-41.
- Merino-Castello, A. (2003): Impact of the Reference Price System on the Pharmaceutical Market: a Theoretical Approach, *Universidad Pompeu Fabra Working Paper No. 524*. Available at <http://www.econ.upf.edu/docs/papers/downloads/524.pdf>.
- Mestre-Ferrándiz, J. (2001): Reference Prices: The Spanish Way. *Investigaciones económicas* 27, 125-149.
- Miraldo, M. (2007): Reference pricing versus co-payment in the pharmaceutical industry: Firms' pricing strategies. *CHE Working Paper No. 27*.
- Mossialos, E. & Le Grand, J. (1999): Cost containment in the EU: an overview, in: Mossialos, E. & Le Grand, J. (eds.): *Health care and cost containment in the European Union*. Ashgate.
- 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.

- OECD Health Data (2010): Selected Data, <http://stats.oecd.org/Index.aspx?DataSetCode=HEALTH>.
- Pavcnik, N. (2002): Do Pharmaceutical Prices Respond to Potential Patient Out-of-the-Pocket Expenses?, *RAND Journal of Economics* 33, 469 - 487.
- Pecorino, P. (2002): "Should the US allow prescription drug reimports from Canada?. *Journal of Health Economics*, Elsevier 21, 699-708.
- Persson U., Anell A, Persson M. (2001) Parallellhandel med läkemedel i Sverige - en ekonomisk analys. 2001. Lund, Sverige, IHE - Institutet för Hälso- ochsjukvårdsekonomi.
- Puig-Junoy, J. (2010): Impact of European Price Regulation on Generic Price Competition, *Pharmacoeconomics* 28 (8), 649-663.
- Raimondos-Møller, P. & Schmitt, N. (2010): Commodity taxation and parallel imports. *Journal of public economics* 94, 153-162.
- Rey, P. (2003): The Impact of Parallel Imports on Prescription Medicines. mimeo, University of Toulouse.
- Rey, P. & Verge, T. (2008): Economics of Vertical Restraints. in: Buccirosi, P. (ed.): *Handbook of Antitrust Economics*, 353-390.
- Robinson, R. (2002): User charges for health care, in: Mossialos, E.; Dixon, A.; Figueras, J. & Kutzin, J. (eds.): *Funding health care: options for Europe*, Buckingham, 161-183.
- Scherer, F. M. (1996): *Industry structure, strategy, and public policy*, New York.
- Scherer, F. M. (2001): The Pharmaceutical Industry, in: Culyer, A. & Newhouse, J. (eds.): *Handbook of Health Economics*, Amsterdam, 1297- 1338.
- Schmalensee, R. (1982): Product Differentiation Advantages of Pioneering Brands, *American Economic Review* 72, 349 - 365.

German Social Security Code V ("Fünftes Buch Sozialgesetzbuch - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477), das zuletzt durch Artikel 8 des Gesetzes vom 12. April 2012 (BGBl. I S. 579) geändert worden ist").

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

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>.

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

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.

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

Wiggins, S. N. & Maness, R. (1994): Price Competition in Pharmaceutical Markets, Texas A&M University, Department of Economics Working Paper.

Zweifel, P. & Manning, W. (2001): Moral Hazard and Consumer Incentives in Health Care, in: Culyer, A. & Newhouse, J. (eds.): *Handbook of Health Economics*, Amsterdam, 409-459.

Appendix A

Appendix to Chapter 2

A.1 Generic Competition Paradox

Frank & Salkever (1992) explain the generic competition paradox with brand loyalty in a horizontal differentiation model (brand loyal customers would never purchase the generic, no matter, how high the price difference is). In Cabrales (2003) vertical product differentiation accounts for the generic competition paradox, but only for some distribution of tastes. Kong (2008) explains it with different health insurance coverage across the population.

This model is able to explain the emergence of the generic competition paradox with the help of vertical differentiation resulting from consumer switching cost. These are determined by degree of uncertainty about the quality of the drug, thus in this model the generic competition paradox does not arise dependent on demand side characteristics such as brand loyalty or distribution of tastes, but rather emerges in connection with certain product characteristics.

Under patent protection the brand-name producer enjoys monopoly pricing power, but there is also uncertainty about whether the drug is well-working, i.e. effective and well-tolerated. This translates into a quality discount of $(1 - \tau)$ now applicable to the brand-name drug's quality.

A consumer who is indifferent between purchasing and trying the brand-name drug b or not

has a gross valuation θ^* , given by

$$\theta^* s (1 - \tau) - c_b = 0 \quad (\text{A.1})$$

yielding

$$\theta^* = \frac{c_b}{(1 - \tau) s}. \quad (\text{A.2})$$

Hence, with $c_b = \kappa p_b^{mon}$ being the patient co-payment demand for brand-name drug b is given by

$$q_b = 1 - \frac{\kappa p_b^{mon}}{(1 - \tau) s}. \quad (\text{A.3})$$

Thus, under patent protection profit for the brand-name drug producer is given as

$$\Pi_b^{mon} = p_b^{mon} \left(1 - \frac{\kappa p_b^{mon}}{(1 - \tau) s} \right) \quad (\text{A.4})$$

with the first order condition

$$\left(1 - \frac{\kappa p_b^{mon}}{(1 - \tau) s} \right) + p_b^{mon} \left(-\frac{\kappa}{(1 - \tau) s} \right) = 0. \quad (\text{A.5})$$

The resulting monopoly price is

$$p_b^{mon} = \frac{s(1 - \tau)}{2\kappa}. \quad (\text{A.6})$$

The generic competition paradox would imply that the second-period Stackelberg price ($p_b = \frac{s\tau}{k(1+\tau)}$) is higher than the monopoly price that is

$$\frac{s\tau}{k(1 + \tau)} > \frac{s(1 - \tau)}{2\kappa}, \quad (\text{A.7})$$

which is the case when $\tau > \sqrt{2} - 1 \approx 0.4142$. Thus, the emergence of the generic competition solely depends on the parameter τ , which is determined by subjective probability that a drug

does not work ϕ , disutility from non-working drug χ and the discount rate r :

$$\tau = \varphi r \frac{1 + \varkappa}{1 + r - \varphi}. \quad (\text{A.8})$$

The parameter τ positively depends on all of these parameters. The higher the subjective probability that the drug does not work i.e. the greater the uncertainty is, the higher is the quality discount factor:

$$\frac{\partial \tau}{\partial \phi} = \frac{r(1 + \varkappa)(1 + r)}{(1 + r - \varphi)^2} > 0. \quad (\text{A.9})$$

The higher the possible damage from a non-working drug is, the higher is the quality discount:

$$\frac{\partial \tau}{\partial \chi} = \frac{r\phi}{1 + r - \varphi} > 0. \quad (\text{A.10})$$

The higher the discount rate is, i.e. the less frequent the purchases take place (and the longer the treatment time), the more important is the individual purchase and the higher is the quality discount:

$$\frac{\partial \tau}{\partial r} = \frac{\phi(1 + \varkappa)(1 + \phi)}{(1 + r - \varphi)^2} > 0. \quad (\text{A.11})$$

Thus, it depends on product characteristics of a single drug, whether the generic competition paradox can be observed. The subjective probability of non-working, the disutility from non-working drug χ and the frequency of intake (the discount rate r) probably depend on indication for which the drug is intended, e.g. drug against headache is probably associated with a lower potential quality discount τ than drug against more serious conditions. This is in line with empirical observations, that sometimes confirm the generic competition paradox and sometimes not, depending on which drugs were included in the data. According to the models of Frank & Salkever (1992), Cabrales (2003), and Kong (2008) the generic competition paradox would be a persistent phenomenon.

A.2 Equivalence of Exogenous and Endogenous Quality Difference

An exogenous quality difference results from uncertainty with respect to the characteristics and the equivalence of the generic. This uncertainty is captured by the subjective probability φ that a drug does not "work", i.e. is not therapeutically effective and well-tolerated. By assumption, both the brand-name drug and the generic version "work" for all patients. But at the point of time considered in this model, the brand-name drug is already established on the market, whereas the generic has just been introduced¹. Hence, consumers know the brand-name drug well, while for the generic uncertainty regarding the characteristics of this drug and especially the equivalence to the brand-name drug prevails for all buyers.

A consumer who buys a drug i of quality s obtains a net utility of

$$U(\theta, s, c_i) = \begin{cases} \theta s - c_i & \text{if the drug "works"} \\ -\varkappa \theta s - c_i & \text{if the drug does not "work"} \end{cases} \quad (\text{A.12})$$

where c_i is the patient co-payment for drug i and $-\varkappa \theta s_i$, $\varkappa \geq 0$ is the value of a "non-working" drug. The utility derived from no drug purchase is zero. Consumers are risk neutral, have infinite time horizons and behave rationally. They always have enough money to buy a unit, if they consider it to be optimal. If they are indifferent between purchase and non-purchase, they buy; if they are indifferent between both drugs, they buy the brand-name drug. Patients have static expectations, they don't assume prices to change².

Following Schmalensee (1982), a consumer purchases and tries an untested drug during its introductory period (while he/she is not sure, whether it "works"), if the expected surplus from

¹This part follows Schmalensee (1982).

²Schmalensee (1982) shows that the reversal of this assumption into the opposite extreme - perfect foresight - does not change results.

the purchase is at least as high as the surplus from the purchase of an alternative:

$$\varphi\left(-\varkappa\theta s - c_i + \frac{z}{r}\right) + (1 - \varphi)\left((\theta s - c_i)\frac{1+r}{r}\right) \geq z\frac{1+r}{r}, \quad (\text{A.13})$$

where r is the one-period discount rates and z is the surplus from the purchase of a substitute. The term $\varphi\left(-\varkappa\theta s - c_i + \frac{z}{r}\right)$ denotes discounted surplus, when the drug is bought and turns out not to "work", forcing the consumer to switch back to the substitute. The term $(1 - \varphi)\left((\theta s - c_i)\frac{1+r}{r}\right)$ gives discounted surplus, if the drug is bought and "works" and the consumer continues to buy it forever and the term $z\frac{1+r}{r}$ denotes the discounted surplus from buying the substitute.

The subjective probability φ that a drug does not "work" can be transformed into a parameter $(1 - \tau)$ having the same effect as a quality differential compared to a "known", i.e. tested, drug by rewriting the expression above as:

$$c_i \leq \theta s(1 - \tau) - z \text{ with } \tau = \varphi r \frac{1 + \varkappa}{1 + r - \varphi}, \quad (\text{A.14})$$

simply stating that a drug is only bought, if the co-payment c_i does not exceed the utility (reduced by uncertainty) from consumption less the utility that would be obtained from a substitute (Schmalensee (1982)).

Reduction of the pre-purchase uncertainty regarding the generic version to the parameter τ implies less utility from the drug under uncertainty $(\theta s(1 - \tau))$ than under certainty (θs) . This impact of uncertainty on utility increases with the subjective probability φ , the cost of a non-working drug \varkappa and the discount rate r . Thus, the more harm a "non-working" drug causes and the more often a drug has to be taken³, the greater the impact of pre-purchase uncertainty. It is assumed that $0 \leq \tau < 1$ for the following considerations, as at a $\tau \geq 1$ a new drug would never be purchased at a positive c_i . If consumers are sure that the drug "works", $\varphi = 0$ and thus also $\tau = 0$. Vice versa, $\tau = 0$ only, if $\varphi = 0$.

³Inter-purchase time is constant, amounts to one period and can be interpreted as treatment duration, as a drug has to be taken by patients the whole time between two purchases. More frequent purchases and thus shorter treatment periods result c.p. in a smaller r .

In the duopolistic period considered here, as consumers have purchased the brand-name drug before, but don't know the generic, uncertainty only applies to the generic. Hence, for the brand-name drug, φ and τ are 0 and the alternative is switching to the generic yielding a surplus of $z = \theta s(1 - \tau) - c_g$. For the generic, $\varphi > 0$ and therefore $\tau > 0$ and the alternative is purchasing the brand-name drug with a surplus of $z = \theta s - c_b$.

Consumers purchase the brand-name drug, if the co-payment for the brand-name drug is less than the utility from purchase less the surplus from the generic

$$c_b \leq \theta s - (\theta s(1 - \tau) - c_g) \iff \theta \geq \frac{c_b - c_g}{s\tau}. \quad (\text{A.15})$$

Consumers switch to the generic, if the co-payment for the generic is less than the utility from purchase (while uncertainty results in switching cost) net of the surplus from buying the brand-name drug and if the net surplus from purchasing the generic is nonnegative

$$c_g < \theta s(1 - \tau) - (\theta s - c_b) \iff \theta < \frac{c_b - c_g}{s\tau} \quad (\text{A.16})$$

and

$$c_g \leq \theta s(1 - \tau) \iff \theta \leq \frac{c_g}{s(1 - \tau)}. \quad (\text{A.17})$$

Hence, consumers with a valuation higher than or equal to $\frac{c_b - c_g}{s\tau}$ purchase the brand-name drug, while consumers with a valuation lower than $\frac{c_b - c_g}{s\tau}$ and higher than or equal to $\frac{c_g}{s(1 - \tau)}$ buy the generic. Thus, demand for brand-name drug b and for the generic g is given by

$$q_b = 1 - \frac{c_b - c_g}{s\tau} \text{ and } q_g = \frac{c_b - c_g}{s\tau} - \frac{c_g}{s(1 - \tau)}. \quad (\text{A.18})$$

If, however, the quality difference between both drugs is considered to be of endogenous nature, stemming from the firms' decisions to invest in quality, the brand-name drug has quality s_b and the generic quality s_g .

A consumer who is indifferent between purchasing the brand-name drug b or the generic

version g has a gross valuation θ^* , given by

$$\theta^* s_b - c_b = \theta^* s_g - c_g, \text{ yielding } \theta^* = \frac{c_b - c_g}{s_b - s_g} \quad (\text{A.19})$$

while a consumer who is indifferent between buying the generic and not buying at all has a gross valuation $\underline{\theta}$, given by

$$\underline{\theta} s_g - c_g = 0, \text{ yielding } \underline{\theta} = \frac{c_g}{s_g}. \quad (\text{A.20})$$

Hence, demand for brand-name drug b and for the generic g is given by

$$q_b = 1 - \frac{c_b - c_g}{s_b - s_g} \text{ and } q_g = \frac{c_b - c_g}{s_b - s_g} - \frac{c_g}{s_g}. \quad (\text{A.21})$$

If s_g is considered to be a constant proportion of s_b , endogenous and exogenous quality difference can be matched

$$s_b \equiv s, \quad s_g \equiv (1 - \tau) s \quad (\text{A.22})$$

and hence

$$\theta^* = \frac{c_b - c_g}{s_b - s_g} = \frac{c_b - c_g}{s\tau}, \quad \underline{\theta} = \frac{c_g}{s_g} = \frac{c_g}{s(1 - \tau)}. \quad (\text{A.23})$$

That is, without loss of generality, the objective difference between b and g can also be captured by the parameter τ . Accordingly, the utility function is given as

$$U(\theta, \tau, s, c_i) = \begin{cases} \theta s - c_i & \text{if } i = b \\ \theta (1 - \tau) s - c_i & \text{if } i = g. \end{cases} \quad (\text{A.24})$$

A.3 Comparative Statics with respect to Quality, Coinsurance Rate, Product Differentiation, and Regulatory Parameter

Change in Quality Willingness to pay increases with quality. This is capitalized by both firms, which increase prices. Regulation reduces the firms' ability to do so:

$$\begin{aligned}\frac{\partial p_b}{\partial s} &> \frac{\partial p_b^\mu}{\partial s} > 0, & \frac{\partial p_b}{\partial s} &> \frac{\partial p_b^o}{\partial s} > 0, \\ \frac{\partial p_g}{\partial s} &> \frac{\partial p_g^\mu}{\partial s} > 0, & \frac{\partial p_g}{\partial s} &> \frac{\partial p_g^o}{\partial s} > 0.\end{aligned}\tag{A.25}$$

Quantities are independent of s :

$$\begin{aligned}\frac{\partial q_b}{\partial s} &= \frac{\partial q_b^\mu}{\partial s} = \frac{\partial q_b^o}{\partial s} = 0, \\ \frac{\partial q_g}{\partial s} &= \frac{\partial q_g^\mu}{\partial s} = \frac{\partial q_g^o}{\partial s} = 0.\end{aligned}\tag{A.26}$$

Profits increase due to higher prices (and unchanged sales volumes):

$$\begin{aligned}\frac{\partial \Pi_b}{\partial s} &> \frac{\partial \Pi_b^\mu}{\partial s} > 0, & \frac{\partial \Pi_b}{\partial s} &> \frac{\partial \Pi_b^o}{\partial s} > 0, \\ \frac{\partial \Pi_g}{\partial s} &> \frac{\partial \Pi_g^\mu}{\partial s} > 0, & \frac{\partial \Pi_g}{\partial s} &> \frac{\partial \Pi_g^o}{\partial s} > 0.\end{aligned}\tag{A.27}$$

Change in Coinsurance Rate A higher coinsurance rate κ increases the price elasticity of demand. Firms absorb this through prices. Consequently, prices decrease in κ :

$$\begin{aligned}\frac{\partial p_b}{\partial \kappa} &< 0, & \frac{\partial p_b^\mu}{\partial \kappa} &< 0, & \frac{\partial p_b^o}{\partial \kappa} &< 0, \\ \frac{\partial p_g}{\partial \kappa} &< 0, & \frac{\partial p_g^\mu}{\partial \kappa} &< 0, & \frac{\partial p_g^o}{\partial \kappa} &< 0.\end{aligned}\tag{A.28}$$

Regulation reduces the responsiveness of prices:

$$\begin{aligned} \left| \frac{\partial p_b}{\partial \kappa} \right| &> \left| \frac{\partial p_b^\mu}{\partial \kappa} \right|, & \left| \frac{\partial p_b}{\partial \kappa} \right| &> \left| \frac{\partial p_b^g}{\partial \kappa} \right| \\ \left| \frac{\partial p_g}{\partial \kappa} \right| &> \left| \frac{\partial p_g^\mu}{\partial \kappa} \right|, & \left| \frac{\partial p_g}{\partial \kappa} \right| &> \left| \frac{\partial p_g^g}{\partial \kappa} \right|. \end{aligned} \quad (\text{A.29})$$

Under no regulation and the maximum price system, quantities are independent of the coinsurance rate:

$$\begin{aligned} \frac{\partial q_b}{\partial \kappa} &= \frac{\partial q_b^\mu}{\partial \kappa} = 0, \\ \frac{\partial q_g}{\partial \kappa} &= \frac{\partial q_g^\mu}{\partial \kappa} = 0. \end{aligned} \quad (\text{A.30})$$

Under the reference price system, increased price elasticity of demand results in lower quantities:

$$\frac{\partial q_b^g}{\partial \kappa} < 0, \quad \frac{\partial q_g^g}{\partial \kappa} < 0. \quad (\text{A.31})$$

Profits decrease due to higher prices:

$$\begin{aligned} \frac{\partial \Pi_b}{\partial \kappa} &< 0, & \frac{\partial \Pi_b^\mu}{\partial \kappa} &< 0, & \frac{\partial \Pi_b^g}{\partial \kappa} &< 0, \\ \frac{\partial \Pi_g}{\partial \kappa} &< 0, & \frac{\partial \Pi_g^\mu}{\partial \kappa} &< 0, & \frac{\partial \Pi_g^g}{\partial \kappa} &< 0. \end{aligned} \quad (\text{A.32})$$

Change in Degree of Product Differentiation The parameter τ does not only denote the degree of product differentiation, more precisely, the size of the quality difference between the brand-name drug and the generic. It also relates to the degree of competition between the two firms.

If $\tau = 0$, both drugs are considered to be of the same quality and competition between both firms amounts to Bertrand competition with homogenous products. Both firms price at marginal cost, i.e. drug prices are zero. The market will be covered, with a brand-name market share of $\frac{2}{3}$ and a generic market share of $\frac{1}{3}$. The other extreme is $\tau = 1$. In this case, no

quality is associated with the generic and the gross utility from it is zero. Consequently, there is a monopoly for brand-name producer, who charges the monopoly price of $p_b = \frac{s}{2\kappa}$ and sells a quantity of $q_b = \frac{1}{2}$.

Under no regulation, the brand-name drug price increases with τ :

$$\frac{\partial p_b}{\partial \tau} > 0. \quad (\text{A.33})$$

Under no regulation, there are two effects of an increasing τ on the generic price, a direct negative effect from the best-response function and an positive strategic effect from the brand-name drug price. For $\tau < 2\sqrt{3} - 3$ ($\tau > 2\sqrt{3} - 3$) the positive (negative) effect dominates and the generic price increases (decreases) with τ :

$$\frac{\partial p_g}{\partial \tau} > 0 \text{ if } \tau < 2\sqrt{3} - 3, \quad \frac{\partial p_g}{\partial \tau} < 0 \text{ if } \tau > 2\sqrt{3} - 3. \quad (\text{A.34})$$

Under the maximum price system, the impact of increased product differentiation on prices depends on the degree of regulation. If the degree of regulation is low ($\mu < \frac{6}{3+6\tau+\tau^2}$ for the brand-name price and $\mu < \frac{4}{(5+\tau)(1+\tau)}$ for the generic price), prices increase with τ . If regulation is strict ($\mu > \frac{6}{3+6\tau+\tau^2}$ and $\mu > \frac{4}{(5+\tau)(1+\tau)}$, respectively), prices decrease with τ . That is, strict regulation may prevent the price increases due to increased product heterogeneity.

$$\begin{aligned} \frac{\partial p_b^\mu}{\partial \tau} &> 0 \text{ if } \mu < \frac{6}{3+6\tau+\tau^2}, \quad \frac{\partial p_b^\mu}{\partial \tau} < 0 \text{ if } \mu > \frac{6}{3+6\tau+\tau^2} \\ \frac{\partial p_g^\mu}{\partial \tau} &> 0 \text{ if } \mu < \frac{4}{(5+\tau)(1+\tau)}, \quad \frac{\partial p_g^\mu}{\partial \tau} < 0 \text{ if } \mu > \frac{4}{(5+\tau)(1+\tau)}. \end{aligned} \quad (\text{A.35})$$

Under the reference price system, the impact of increased product differentiation on prices does not depend on the degree of regulation, but on the degree of τ . If τ is low ($\tau < \sqrt{15} - 3$ for the brand-name drug and $\tau < 2\sqrt{3} - 3$ for the generic), prices increase with τ , if τ is high ($\tau > \sqrt{15} - 3$ and $\tau > 2\sqrt{3} - 3$ respectively), prices decrease with τ . That is, for more heterogenous products, the mere introduction of a reference price system, independent of its

specific design prevents price increases due to increased product differentiation. On the other hand, if products are closer substitutes, increased product differentiation will always result in price increases, no matter how strictly regulated the market is:

$$\begin{aligned}\frac{\partial p_b^e}{\partial \tau} &> 0 \text{ if } \tau < \sqrt{15} - 3, \quad \frac{\partial p_b^e}{\partial \tau} < 0 \text{ if } \tau > \sqrt{15} - 3 \\ \frac{\partial p_g^e}{\partial \tau} &> 0 \text{ if } \tau < 2\sqrt{3} - 3, \quad \frac{\partial p_g^e}{\partial \tau} < 0 \text{ if } \tau > 2\sqrt{3} - 3.\end{aligned}\tag{A.36}$$

Under no regulation and the reference price system, both quantities decrease with τ .

$$\begin{aligned}\frac{\partial q_b}{\partial \tau} &< 0, \quad \frac{\partial q_b^e}{\partial \tau} < 0, \\ \frac{\partial q_g}{\partial \tau} &< 0, \quad \frac{\partial q_g^e}{\partial \tau} < 0.\end{aligned}\tag{A.37}$$

If the market is not regulated, a higher brand-name drug price induces consumers who bought the brand-name drug before to switch to the generic whose price either decreases (if τ is high) or increases (if τ is low), but by less than the brand-name price ($\frac{\partial p_b}{\partial \tau} > \frac{\partial p_g}{\partial \tau}$). This is also the case under the reference price system, if τ is low. But why do some consumers switch to the generic, when τ is high and the brand-name price decreases? The generic price also decreases and even though it decreases by less than the brand-name price it induces consumers to switch to the generic. The explanation is the design of the co-payment for the brand-name drug. As both drug prices decrease, also the reference price (which is linear function of both drug prices) decreases and with it the subsidy for the brand-name drug decreases. Consequently, the relative co-payment for the brand-name drug is higher than for the generic drug and some consumers switch. The quantity of the generic decreases due to two factors: First, if the price of the generic increases with τ – and this is the case under no regulation and under the reference price, if τ is low – some consumers quit buying the generic. Second, when τ increases, the gross valuation of the generic ($\theta(1 - \tau)$) decreases so that consumers with low valuation θ do no longer buy the generic, even if the price of the generic decreases. That is, as both quantities decrease with τ ,

the more differentiated the two drugs are, the larger is the part of the market that is uncovered. Under the maximum price system, the impact of increased product differentiation on the brand-name price depends on the degree of regulation. The quantity of the brand-name drug decreases (increases) in τ , if μ is small ($\mu < \frac{4}{(5+\tau)(1+\tau)}$) (if μ is large ($\mu > \frac{4}{(5+\tau)(1+\tau)}$)). The quantity of the generic decreases with τ :

$$\begin{aligned} \frac{\partial q_b^\mu}{\partial \tau} &> 0 \text{ if } \mu < \frac{4}{(5+\tau)(1+\tau)}, \quad \frac{\partial q_b^\mu}{\partial \tau} < 0 \text{ if } \mu > \frac{4}{(5+\tau)(1+\tau)} \\ \frac{\partial q_b^\mu}{\partial \tau} &< 0. \end{aligned} \tag{A.38}$$

The impact of increased product differentiation on profits will be discussed in Section 4.3.

Change in Regulatory Parameter An increase in the regulatory parameters implies stricter regulation, *ceteris paribus*. Under the maximum price system, a lower mark-up is granted for the brand-name drug and under the reference price system, the generic is given a higher weight in the reference price and thus, the reimbursement amount is lower. An increase in the regulatory parameters magnifies the effects of maximum price and reference price regulation on prices, quantities and profits:

Both drug prices decrease in μ under the maximum price system and in ϱ under the reference price system:

$$\begin{aligned} \frac{\partial p_b^\mu}{\partial \mu} &< 0, \quad \frac{\partial p_g^\mu}{\partial \mu} < 0, \\ \frac{\partial p_b^\varrho}{\partial \varrho} &< 0, \quad \frac{\partial p_g^\varrho}{\partial \varrho} < 0. \end{aligned} \tag{A.39}$$

Under the maximum price system, the brand-name quantity increases in μ and the generic

quantity decreases in μ . Under the reference price system, both quantities increase in ϱ :

$$\begin{aligned}\frac{\partial q_b^\mu}{\partial \mu} &> 0, & \frac{\partial q_g^\mu}{\partial \mu} &< 0, \\ \frac{\partial q_b^\varrho}{\partial \varrho} &> 0, & \frac{\partial q_g^\varrho}{\partial \varrho} &> 0.\end{aligned}\tag{A.40}$$

Profits decrease in μ and ϱ :

$$\begin{aligned}\frac{\partial \Pi_b^\mu}{\partial \mu} &< 0, & \frac{\partial \Pi_g^\mu}{\partial \mu} &< 0, \\ \frac{\partial \Pi_b^\varrho}{\partial \varrho} &< 0, & \frac{\partial \Pi_g^\varrho}{\partial \varrho} &< 0.\end{aligned}\tag{A.41}$$

A.4 Health Policy Objectives

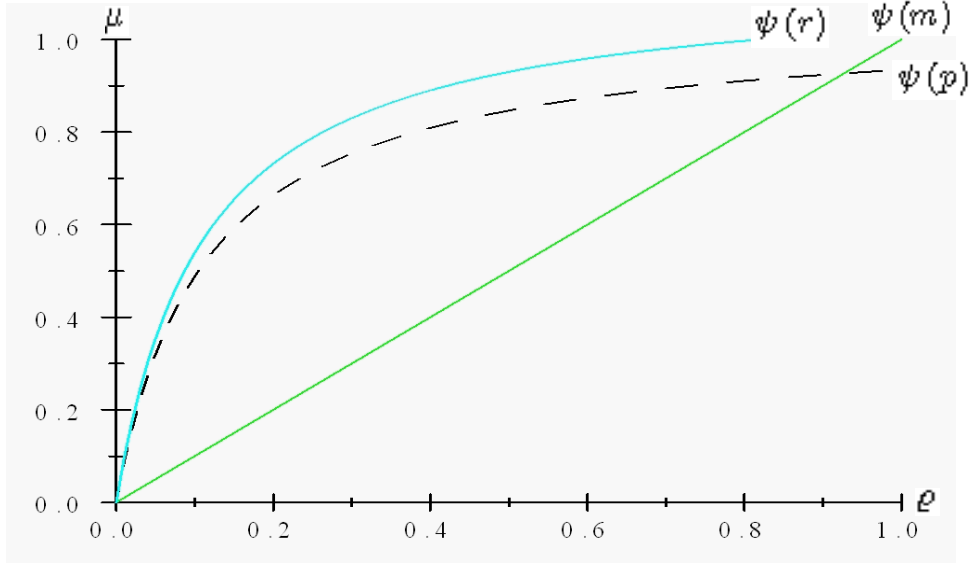
A.4.1 Price Reductions

Regulatory price-equivalence $\psi(r)$

$$\begin{aligned}p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{s\tau((2 - \varrho)(\kappa + \varrho(1 - \kappa)) - (2 - (\kappa + \varrho(1 - \kappa)))\tau\varrho)}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_r^\varrho \\ \mu &= \varrho \frac{6(\kappa + \varrho(1 - \tau)) - \kappa(\varrho(9 - 10\tau) - \tau^2(2 - \varrho)) - \kappa^2(1 - \tau)(3 - \tau)(1 - \varrho)}{(\tau + 1)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}\end{aligned}\tag{A.42}$$

Price reductions-equivalence $\psi(p)$

$$\begin{aligned}p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\ \mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}\end{aligned}\tag{A.43}$$



Brand-name Price, $\tau = 0.9$

A.4.2 Expenditure Reduction

If the pharmaceutical market remains unregulated, public pharmaceutical expenditure is given as the reimbursed fraction $(1 - \kappa)$ of total expenditure (unregulated price \times quantity):

$$E = (1 - \kappa) (p_b q_b + p_g q_g) = \frac{s\tau (1 - \kappa) (5 - \tau)}{\kappa (3 + \tau)^2}. \quad (\text{A.44})$$

Under the maximum price system, the lower prices of both drugs reduce expenditure to

$$\begin{aligned} E^\mu &= (1 - \kappa) (p_b^\mu q_b^\mu + p_g^\mu q_g^\mu) \\ &= \frac{s\tau (1 - \kappa) [4(5 - \tau) - 8\mu(1 - \tau^2) - \mu^2(3\tau + 1)(\tau + 1)^2]}{4\kappa(3 + \tau)^2} \end{aligned} \quad (\text{A.45})$$

and under the reference price system, both lower prices and a lower basis for reimbursement (not the market price, but the reference price is the basis for reimbursement) contribute to lower

expenditure of:

$$\begin{aligned}
E^\varrho &= (1 - \kappa) (p_r^\varrho q_b^\varrho + p_g^\varrho q_g^\varrho) \\
&= \frac{s\tau (1 - \kappa) [\kappa + \varrho (1 - \tau) (1 - \kappa)] \left[\begin{array}{l} \kappa (5 - \tau) - \kappa \varrho (7 - 3\tau) \\ + \varrho (1 - \tau) (5 - 2\varrho (1 - \kappa)) \end{array} \right]}{[\kappa + \varrho (1 - \kappa)] [\kappa (3 + \tau) + 3\varrho (1 - \tau) (1 - \kappa)]^2}. \quad (\text{A.46})
\end{aligned}$$

Thus, both regulatory instruments succeed in reducing pharmaceutical expenditure:

$$\begin{aligned}
E - E^\mu &= \frac{s\tau \mu (\tau + 1) (1 - \kappa) (8 (1 - \tau) + \mu (3\tau + 1) (\tau + 1))}{4\kappa (3 + \tau)^2} > 0, \\
E - E^\varrho &= s\tau \frac{(1 - \kappa)}{\kappa} \frac{\left[\begin{array}{l} \kappa \varrho^2 (1 - \tau) (1 - \kappa) (90 - 27\tau + 28\tau^2 + 5\tau^3) \\ + \kappa^2 \varrho (\tau + 3) (15 - 11\tau + 13\tau^2 - \tau^3) \\ - \kappa^3 \varrho (\tau + 3) (9 - \tau) (1 - \tau)^2 \\ - 3\kappa^2 \varrho^2 (1 - \tau)^2 (1 - \kappa) (18 - 5\tau - \tau^2) \\ + \varrho^3 (1 - \tau)^2 (1 - \kappa)^2 9 (5 - \tau) \\ - \kappa \varrho^3 (1 - \tau)^2 (1 - \kappa)^2 (27 - 21\tau - 2\tau^2) \end{array} \right]}{[3 + \tau]^2 [\kappa + \varrho (1 - \kappa)] [\kappa (3 + \tau) + 3\varrho (1 - \tau) (1 - \kappa)]^2} > 0. \quad (\text{A.47})
\end{aligned}$$

Expenditure-equivalence E^*

$$\begin{aligned}
E^\mu &= \frac{s\tau (1 - \kappa) \left[4 (5 - \tau) - 8\mu (1 - \tau^2) - \mu^2 (3\tau + 1) (\tau + 1)^2 \right]}{4\kappa (3 + \tau)^2} \\
&= \frac{s\tau (1 - \kappa) [\kappa + \varrho (1 - \tau) (1 - \kappa)] \left[\begin{array}{l} \kappa (5 - \tau) - \kappa \varrho (7 - 3\tau) \\ + \varrho (1 - \tau) (5 - 2\varrho (1 - \kappa)) \end{array} \right]}{[\kappa + \varrho (1 - \kappa)] [\kappa (3 + \tau) + 3\varrho (1 - \tau) (1 - \kappa)]^2} = E^\varrho \quad (\text{A.48})
\end{aligned}$$

$$\tau = 0.1, \kappa = 0.1$$

$$E^\mu = -3.6829 \times 10^{-2} \mu^2 - 0.18543 \mu + 0.45890 = \frac{0.09(0.81\rho + 0.1)(-1.62\rho^2 + 3.83\rho + 0.49)}{(0.9\rho + 0.1)(2.43\rho + 0.31)^2} = E^\rho$$

$$\mu = - \left[\frac{\frac{2.1987 \times 10^9 \rho + 1.8046 \times 10^{10} \rho^2 + 4.9273 \times 10^{10} \rho^3 + 8.9099 \times 10^7}{1.9572 \times 10^{10} \rho^3 + 7.1685 \times 10^9 \rho^2 + 8.7340 \times 10^8 \rho + 3.5393 \times 10^7}}{50000 \sqrt{\frac{2.9729 \times 10^{10} \rho^6 + 1.9366 \times 10^{10} \rho^5 + 5.1676 \times 10^9 \rho^4 + 7.1914 \times 10^8 \rho^3 + 5.4598 \times 10^7 \rho^2 + 2.1142 \times 10^6 \rho + 31755}{1.9572 \times 10^{10} \rho^3 + 7.1685 \times 10^9 \rho^2 + 8.7340 \times 10^8 \rho + 3.5393 \times 10^7}}} \right]$$

$$\tau = 0.5, \kappa = 0.1$$

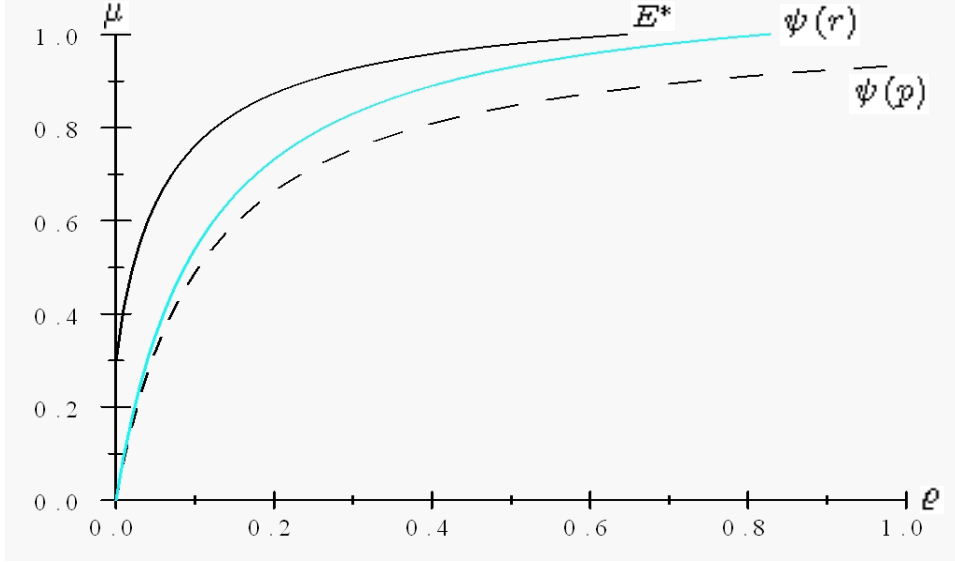
$$E^\mu = -0.51658 \mu^2 - 0.55102 \mu + 1.6531 = \frac{0.45(0.45\rho + 0.1)(-0.9\rho^2 + 1.95\rho + 0.45)}{(0.9\rho + 0.1)(1.35\rho + 0.35)^2} = E^\rho$$

$$\mu = - \left[\frac{\frac{2.2564 \times 10^7 \rho + 1.1381 \times 10^8 \rho^2 + 1.8076 \times 10^8 \rho^3 + 1.3500 \times 10^6}{3.3893 \times 10^8 \rho^3 + 2.1340 \times 10^8 \rho^2 + 4.2308 \times 10^7 \rho + 2.5312 \times 10^6}}{50000 \sqrt{\frac{1.6999 \times 10^8 \rho^6 + 1.8863 \times 10^8 \rho^5 + 8.2880 \times 10^7 \rho^4 + 1.8029 \times 10^7 \rho^3 + 1.9549 \times 10^6 \rho^2 + 89009. \rho + 729.19}{3.3893 \times 10^8 \rho^3 + 2.1340 \times 10^8 \rho^2 + 4.2308 \times 10^7 \rho + 2.5312 \times 10^6}}} \right]$$

$$\tau = 0.9, \kappa = 0.1$$

$$E^\mu = -1.7783 \mu^2 - 0.20237 \mu + 2.1834 = \frac{0.81(0.09\rho + 0.1)(-0.18\rho^2 + 0.07\rho + 0.41)}{(0.9\rho + 0.1)(0.27\rho + 0.39)^2} = E^\rho$$

$$\mu = - \left[\frac{\frac{3.5516 \times 10^7 \rho + 4.4258 \times 10^7 \rho^2 + 1.4753 \times 10^7 \rho^3 + 3.4201 \times 10^6}{2.5928 \times 10^8 \rho^3 + 7.7783 \times 10^8 \rho^2 + 6.2418 \times 10^8 \rho + 6.0107 \times 10^7}}{1.0 \times 10^5 \sqrt{\frac{9.0316 \times 10^6 \rho^6 + 5.2468 \times 10^7 \rho^5 + 1.1573 \times 10^8 \rho^4 + 1.168 \times 10^8 \rho^3 + 4.8927 \times 10^7 \rho^2 + 4.1557 \times 10^6 \rho + 1163.2}{2.5928 \times 10^8 \rho^3 + 7.7783 \times 10^8 \rho^2 + 6.2418 \times 10^8 \rho + 6.0107 \times 10^7}}} \right]$$



Expenditure, $\tau = 0.9$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ϱ , see also price reductions-equivalence $\psi(p)$:

$$\begin{aligned}
 p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\
 \mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}. \tag{A.49}
 \end{aligned}$$

Then public expenditure under the maximum price system can be written in terms of the regulatory parameter ϱ as:

$$E^\mu(\varrho) = s\tau \frac{(1 - \kappa) \left[\begin{aligned} &\kappa^3(5 - \tau) + 4\kappa^2\varrho(4 - 3\tau + \tau^2)(1 - \kappa) \\ &+ \kappa\varrho^2(1 - \tau)(17 - 6\tau + 3\tau^2)(1 - \kappa)^2 + 6\varrho^3(1 - \kappa)^3(1 - \tau)^2 \end{aligned} \right]}{[\kappa + \varrho(1 - \kappa)]^2 [\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2}. \tag{A.50}$$

For identical price reductions, public expenditure is lower under the reference price system:

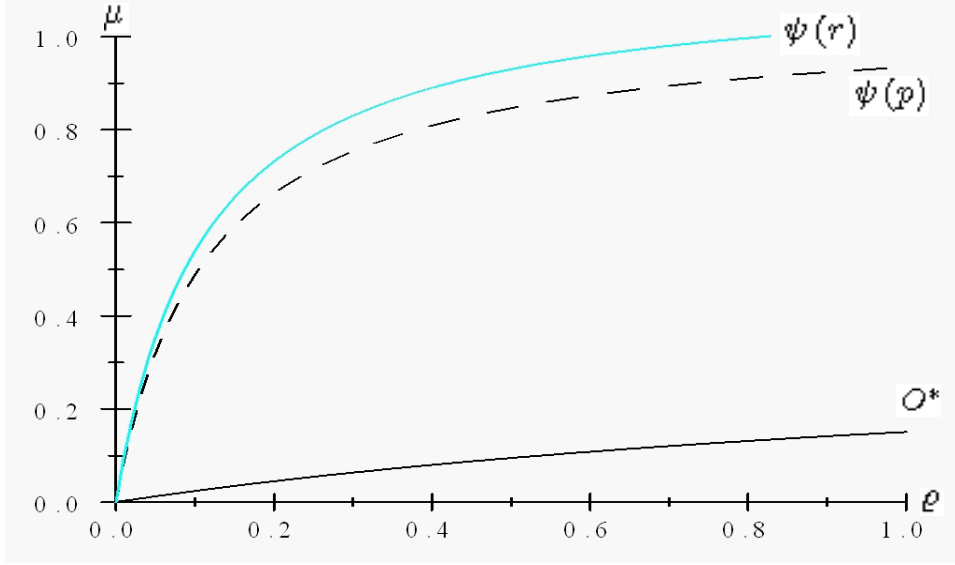
$$\begin{aligned}
& E^\mu - E^\varrho \Big|_{p_b^\mu = p_b^\varrho} \\
& s\tau\varrho(1-\kappa) \left[\begin{array}{l} \kappa^2(1+3\tau^2+\kappa(3\tau+1)(1-\tau)) \\ +\kappa\varrho(1-\kappa)(1-\tau)(2+3\tau^2) \\ +\kappa^2\varrho(1-\kappa)(4+2\tau-5\tau^2+3\tau^3) \\ +\varrho^2(1-\kappa)^2(1-\tau)(1-\tau+\kappa(\tau+5)) \\ +2\varrho^3(1-\kappa)^3(1-\tau)^2 \end{array} \right] \\
= & \frac{\quad}{[\kappa+\varrho(1-\kappa)]^2[\kappa(3+\tau)+3\varrho(1-\tau)(1-\kappa)]^2} > 0 \tag{A.51}
\end{aligned}$$

A.4.3 Equity and Access

Out-of-pocket Expenditure/Co-payments

Co-payment-equivalence for the brand-name

$$\begin{aligned}
c_b^\mu &= \frac{s\tau(2-\mu(1+\tau))}{(3+\tau)} = \frac{s\tau(2\kappa+\varrho(1-\tau)(1-\kappa))}{\kappa(3+\tau)+3\varrho(1-\tau)(1-\kappa)} = c_b^\varrho \\
\mu &= \frac{\varrho(1-\kappa)(-4\tau+\tau^2+3)}{\kappa(4\tau+\tau^2+3)+3\varrho(1-\kappa)(1-\tau^2)} \tag{A.52}
\end{aligned}$$



Brand-name Co-payment, $\tau = 0.9$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ϱ , see also price reductions-equivalence $\psi(p)$:

$$\begin{aligned}
 p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\
 \mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}.
 \end{aligned} \tag{A.53}$$

Then the copayment for the brand-name drug under the maximum price system can be written in terms of the regulatory parameter ϱ as:

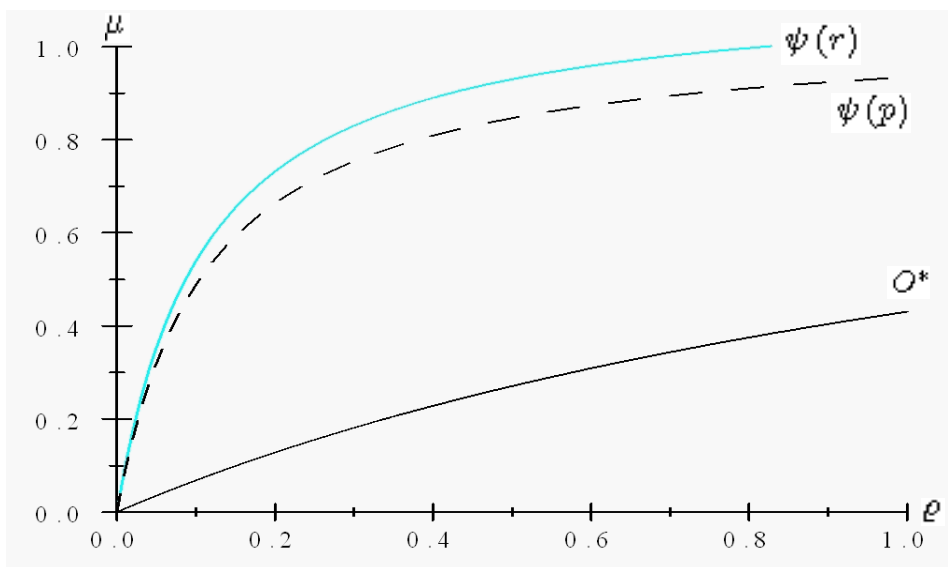
$$c_b^\mu(\varrho) = \frac{2s\kappa\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}. \tag{A.54}$$

Taking identical price reductions as a basis of comparison, the copayment for the brand-name drug is lower under the maximum price system:

$$\begin{aligned}
 & c_b^\mu - c_b^\varrho \big|_{p_b^\mu = p_b^\varrho} \\
 = & -s\tau\varrho \frac{(1-\kappa)(\kappa(1+\tau) + \varrho(1-\tau)(1-\kappa))}{[\kappa + \varrho(1-\kappa)][\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]} < 0.
 \end{aligned} \tag{A.55}$$

Co-payment-equivalence for the generic

$$\begin{aligned}
 c_g^\mu &= \frac{s\tau(2-\mu(1+\tau))(1-\tau)}{2(3+\tau)} = \frac{\kappa s\tau(1-\tau)}{\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)} = c_g^\varrho \\
 \mu &= \frac{6\varrho(1-\tau)(1-\kappa)}{(\tau+1)[\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]}
 \end{aligned} \tag{A.56}$$



Generic Co-payment, $\tau = 0.9$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ϱ , see also

price reductions-equivalence $\psi(p)$:

$$\begin{aligned}
p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\
\mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}.
\end{aligned} \tag{A.57}$$

Then the copayment for the generic drug under the maximum price system can be written in terms of the regulatory parameter ϱ as:

$$c_g^\mu(\varrho) = \frac{2s\kappa\tau(1 - \tau)(\tau + 3)(\kappa + \varrho(1 - \tau)(1 - \kappa))}{2(3 + \tau)(\kappa + \varrho(1 - \kappa))(\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa))}. \tag{A.58}$$

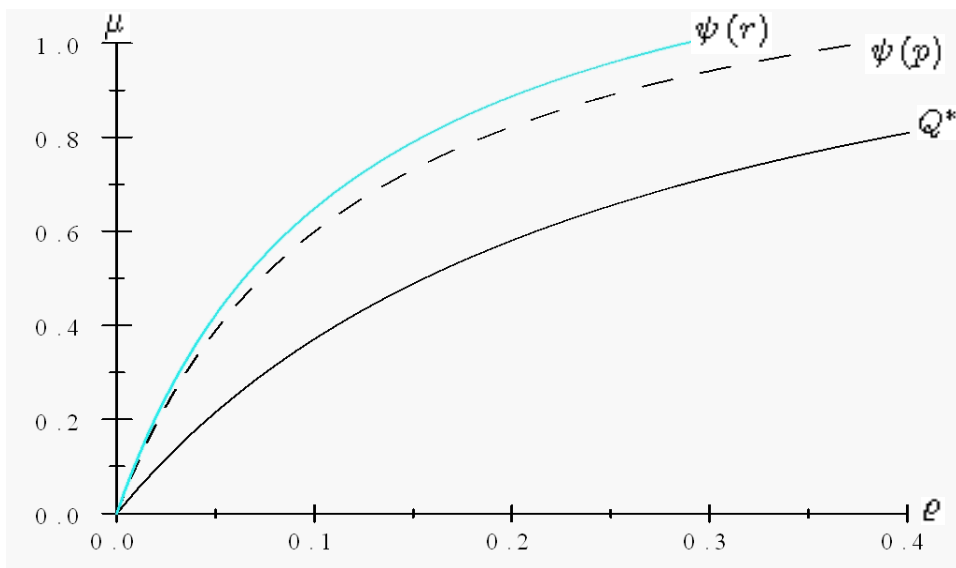
Taking identical price reductions as a basis of comparison, the copayment for the generic drug is lower under the maximum price system:

$$\begin{aligned}
&c_g^\mu - c_g^\varrho \big|_{p_b^\mu = p_b^\varrho} \\
&= -\frac{s\kappa\tau^2\varrho(1 - \kappa)(1 - \tau)}{(\kappa + \varrho(1 - \kappa))(\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa))} < 0.
\end{aligned} \tag{A.59}$$

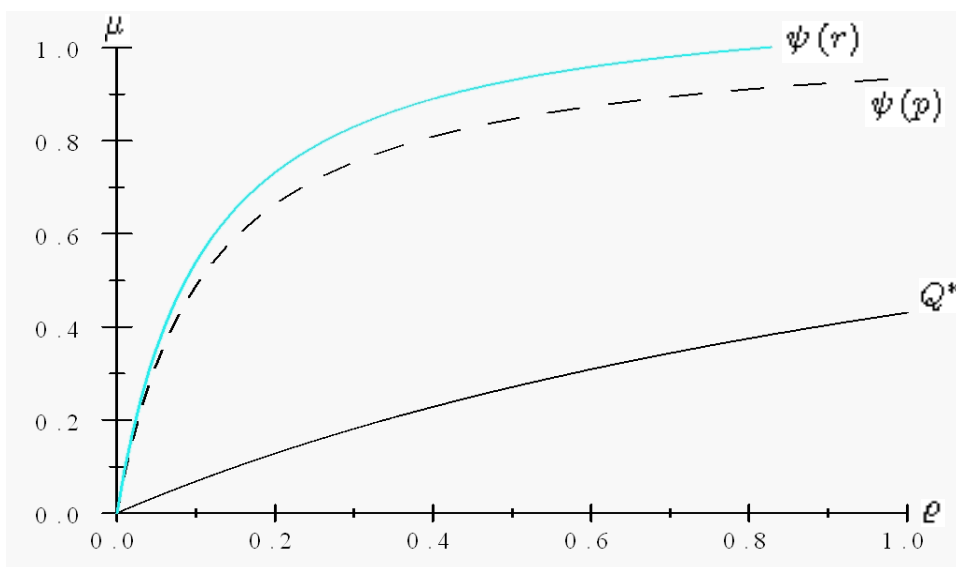
Total Quantity

Quantity-equivalence

$$\begin{aligned}
Q^\mu &= \frac{6 + \tau\mu(1 + \tau)}{2(3 + \tau)} = \frac{3(\kappa + \varrho(1 - \tau)(1 - \kappa))}{\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)} = Q^\varrho \\
\mu &= \frac{6\varrho(1 - \tau)(1 - \kappa)}{(\tau + 1)[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}
\end{aligned} \tag{A.60}$$



Total Quantity, $\tau = 0.5$



Total Quantity, $\tau = 0.9$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ρ , see also

price reductions-equivalence $\psi(p)$:

$$\begin{aligned} p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\ \mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}. \end{aligned} \quad (\text{A.61})$$

Then total quantity under the maximum price system can be written in terms of the regulatory parameter ϱ as:

$$Q^\mu(\varrho) = \frac{\kappa\tau^2\varrho(1 - \kappa)}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} \quad (\text{A.62})$$

Taking identical price reductions as a basis of comparison, total quantity is higher under the maximum price system:

$$\begin{aligned} &Q^\mu - Q^\varrho \Big|_{p_b^\mu = p_b^\varrho} \\ &= \frac{\kappa\tau^2\varrho(1 - \kappa)}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} > 0. \end{aligned} \quad (\text{A.63})$$

Consumer Surplus

If the market is not regulated, consumer surplus for the brand-name users is given as

$$CS_b = \int_{\theta^*}^1 (\theta s - \kappa p_b) d\theta = \frac{2s(2 - \tau)}{(\tau + 3)^2} \quad (\text{A.64})$$

and for generic user as

$$CS_g = \int_{\underline{\theta}}^{\theta^*} (\theta(1 - \tau)s - \kappa p_g) d\theta = \frac{s(1 - \tau)}{2(\tau + 3)^2}. \quad (\text{A.65})$$

Under the maximum price system, consumer surplus for the brand-name users is given as

$$CS_b^\mu = \int_{\theta^*}^1 (\theta s - \kappa p_b^\mu) d\theta = \frac{s[4(2 - \tau) - \mu(\tau + 1)(1 - 3\tau)][4 + \mu(\tau + 1)^2]}{8(\tau + 3)^2}, \quad (\text{A.66})$$

which is higher than in the benchmark case, as a larger quantity is consumed at a lower price:

$$CS_b - CS_b^\mu = -\frac{s\mu(\tau+1)\left[4(1+4\tau-\tau^2) - \mu(1-3\tau)(\tau+1)^2\right]}{8(\tau+3)^2} < 0. \quad (\text{A.67})$$

Consumer surplus for generic users is given as

$$CS_g^\mu = \int_{\underline{\theta}}^{\theta^*} (\theta(1-\tau)s - \kappa p_g^\mu) d\theta = \frac{s(1-\tau)(2-\mu(1+\tau))^2}{8(\tau+3)^2}, \quad (\text{A.68})$$

which is lower than in the case of no regulation, as the effect of a lower quantity dominates the effect of a lower price on consumer surplus:

$$CS_g - CS_g^\mu = \frac{s\mu(1-\tau^2)(4-\mu(1+\tau))}{8(\tau+3)^2} > 0. \quad (\text{A.69})$$

Under the reference price system, consumer surplus for both brand-name users and generic users is higher than under no regulation, since drug prices are lower and higher quantities are consumed:

$$CS_b^{\varrho} = \int_{\theta^*}^1 (\theta s - \kappa p_R^{\varrho} - (p_b^{\varrho} - p_R^{\varrho})) d\theta = \frac{2s(2-\tau)(\kappa + \varrho(1-\tau)(1-\kappa))^2}{[\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2},$$

$$CS_b - CS_b^{\varrho} = -\frac{2s\tau\varrho(1-\kappa)(1-\tau)(2-\tau) \left[\begin{array}{c} 2\kappa(\tau+3) \\ +\varrho(1-\kappa)(6+\tau)(1-\tau) \end{array} \right]}{(\tau+3)^2 [\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2} < 0 \quad (\text{A.70})$$

$$CS_g^{RP} = \int_{\underline{\theta}}^{\theta^*} (\theta(1-\tau)s - \kappa p_g) d\theta = \frac{1}{2}s(1-\tau) \frac{(\kappa + \varrho(1-\tau)(1-\kappa))^2}{[\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2},$$

$$CS_g - CS_g^{\varrho} = -\frac{s\tau\varrho(1-\kappa)(1-\tau)^2 [2\kappa(\tau+3) + \varrho(1-\kappa)(6+\tau)(1-\tau)]}{2(3+\tau)^2 [\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2} < 0. \quad (\text{A.71})$$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ϱ , see also

price reductions-equivalence $\psi(p)$:

$$\begin{aligned} p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \varrho(1 - \tau)(1 - \kappa))}{[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]} = p_b^\varrho \\ \mu &= \frac{2\kappa\varrho(1 - \kappa)(\tau^2 + 3) + 6\varrho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \varrho(1 - \kappa)][\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]}. \end{aligned} \quad (\text{A.72})$$

Then consumer surplus for brand-name users under the maximum price system can be written in terms of the regulatory parameter ϱ as:

$$CS_b^\mu(\varrho) = \frac{s \left[\begin{aligned} &4\kappa^4(2 - \tau) + 2\kappa^3\varrho(17 - 15\tau + 7\tau^2 - \tau^3)(1 - \kappa) \\ &+ \kappa^2\varrho^2(53 - 68\tau + 40\tau^2 - 12\tau^3 + 3\tau^4)(1 - \kappa)^2 \\ &+ 12\kappa\varrho^3(1 - \tau)(3 - 2\tau + \tau^2)(1 - \kappa)^3 + 9\varrho^4(1 - \kappa)^4(1 - \tau)^2 \end{aligned} \right]}{2[\kappa + \varrho(1 - \kappa)]^2[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2}. \quad (\text{A.73})$$

Taking identical price reductions as a basis of comparison, consumer surplus for brand-name users is higher under the maximum price system:

$$\begin{aligned} &CS_b^\mu - CS_b^\varrho \Big|_{p_b^\mu = p_b^\varrho} \\ &= \frac{s\varrho(1 - \kappa) \left[\begin{aligned} &2\kappa^3(1 + \tau + 3\tau^2 - \tau^3) \\ &+ \kappa^2\varrho(5 + 4\tau + 8\tau^2 - 8\tau^3 + 3\tau^4)(1 - \kappa) \\ &+ 4\kappa\varrho^2(1 - \tau)(\tau + 1)^2(1 - \kappa)^2 \\ &+ \varrho^3(4\tau + 1)(1 - \tau)^2(1 - \kappa)^3 \end{aligned} \right]}{2[\kappa + \varrho(1 - \kappa)]^2[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2} > 0. \end{aligned} \quad (\text{A.74})$$

A.4.4 Competition in Pharmaceutical Markets

Incentive for Product Differentiation

If the market is unregulated, brand-name and generic profit are given as

$$\Pi_b = \frac{4s\tau}{\kappa(\tau + 3)^2} \text{ and } \Pi_g = \frac{s\tau(1 - \tau)}{\kappa(\tau + 3)^2}. \quad (\text{A.75})$$

Brand-name profit increases with the degree of product differentiation

$$\frac{\partial \Pi_b}{\partial \tau} = \frac{4s(3-\tau)}{\kappa(\tau+3)^3} > 0. \quad (\text{A.76})$$

Hence, there is an incentive for the brand-name producer to raise his profit by raising the degree of product differentiation. Generic profit increases in τ , if $\tau < \frac{3}{7}$ and decreases otherwise:

$$\frac{\partial \Pi_g}{\partial \tau} = \frac{s(3-7\tau)}{\kappa(\tau+3)^3} > 0 \text{ if } \tau < \frac{3}{7}. \quad (\text{A.77})$$

This is, to some extent there is also incentive for the generic producer to raise τ . For a low degree of product differentiation, the positive strategic effect exceeds the negative direct effect. A certain degree of product differentiation allows the generic producer to attract additional consumers with a low valuation, but also forces it to lower his price.

Under the maximum price system, brand-name and generic profit are given as

$$\Pi_b^\mu = \frac{s\tau(2-\mu(1+\tau)) [4+\mu(\tau+1)^2]}{2\kappa(\tau+3)^2} \text{ and } \Pi_g^\mu = \frac{s\tau(1-\tau) [2-\mu(1+\tau)]^2}{4\kappa(\tau+3)^2}. \quad (\text{A.78})$$

Both profits increase in τ for a low degree of regulation and decrease in τ for a high degree of regulation:

$$\frac{\partial \Pi_b^\mu}{\partial \tau} = \frac{s \left[8(3-\tau) - 2\mu(3-\tau-9\tau^2-\tau^3) - \mu^2(11\tau+2\tau^2+3)(\tau+1)^2 \right]}{2\kappa(\tau+3)^3} < 0, \quad (\text{A.79})$$

if $\mu > \mu_b^* = \frac{\tau+9\tau^2+\tau^3+\tau\sqrt{36\tau+26\tau^2-4\tau^3+\tau^4+9}+3\sqrt{36\tau+26\tau^2-4\tau^3+\tau^4+9}-3}{(\tau+1)^2(11\tau+2\tau^2+3)}.$

$$\frac{\partial \Pi_g^\mu}{\partial \tau} = \frac{s \left[\begin{array}{l} 4(3-7\tau) - 4\mu(3-\tau-9\tau^2-\tau^3) \\ +\mu^2(\tau+1)(2\tau-11\tau^2-2\tau^3+3) \end{array} \right]}{4\kappa(\tau+3)^3} < 0 \quad (\text{A.80})$$

if $\mu > \mu_g^* = \frac{6-14\tau}{2\tau-11\tau^2-2\tau^3+3}$.

That is, for sufficiently strict regulation, there is no incentive to raise τ for both the brand-name producer and generic producer.

Under the reference price system, brand-name and generic profit are given as

$$\Pi_b^e = \frac{4s\tau(\kappa + \varrho(1-\tau)(1-\kappa))^2}{[\kappa + \varrho(1-\kappa)][\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2}$$

and $\Pi_g^e = \frac{s\tau(1-\tau)(\kappa + \varrho(1-\tau)(1-\kappa))}{[\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2}$. (A.81)

The relationship between profits and the degree of product differentiation is determined by the degree of regulation and the degree of product differentiation. The brand-name profit decreases in τ for a high degree of regulation and a high degree of product differentiation:

$$\frac{\partial \Pi_b^e}{\partial \tau} = \frac{4s[\kappa + \varrho(1-\tau)(1-\kappa)] \left[\begin{array}{l} 3\varrho^2(1-\tau)^2(1-\kappa)^2 + \kappa^2(3-\tau) \\ +\kappa\varrho(6-7\tau-\tau^2)(1-\kappa) \end{array} \right]}{[\kappa + \varrho(1-\kappa)][\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^3} < 0, \quad (\text{A.82})$$

if $\varrho > \varrho_b^* = \frac{1}{6}\kappa \frac{7\tau + \tau^2 - \tau\sqrt{26\tau + \tau^2 - 23} - 6}{(1-\tau)^2(1-\kappa)} \wedge \tau > 8\sqrt{3} - 13$.

The generic profit decreases in τ for a high degree of regulation:

$$\frac{\partial \Pi_g^e}{\partial \tau} = \frac{s \left[\begin{array}{l} \kappa^2(3-7\tau) + 3\varrho^2(1-\tau)^3(1-\kappa)^2 \\ +\kappa\varrho(1-\tau)(6-10\tau-\tau^2)(1-\kappa) \end{array} \right]}{[\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^3} < 0, \quad (\text{A.83})$$

if $\varrho > \varrho_g^* = \kappa \frac{10\tau + \tau^2 + \tau\sqrt{20\tau + \tau^2 + 4} - 6}{6(1-\tau)^2(1-\kappa)}$.

The reference price system can only reduce the incentive to raise τ for the brand-name producer, if τ is already very high and if regulation is sufficiently strict. In this case, there is also no incentive to raise τ for the generic producer.

A.5 Welfare Analysis

Welfare under no regulation:

$$\begin{aligned} W &= CS_b + CS_g + \Pi_b + \Pi_g - E \\ &= \frac{s(9 + 5\tau - 2\tau^2)}{(\tau + 3)^2}. \end{aligned} \quad (\text{A.84})$$

Welfare under the maximum price system:

$$\begin{aligned} W^\mu &= CS_b^\mu + CS_g^\mu + \Pi_b^\mu + \Pi_g^\mu - E^\mu \\ &= \frac{s \left[4(9 + 5\tau - 2\tau^2) + 4\tau\mu(\tau + 1)(3\tau + 1) - \tau\mu^2(3\tau + 1)(\tau + 1)^2 \right]}{8(\tau + 3)^2}. \end{aligned} \quad (\text{A.85})$$

Welfare under the reference price system:

$$\begin{aligned} W^e &= CS_b^e + CS_g^e + \Pi_b^e + \Pi_g^e - E^e \\ &= \frac{s \left[\begin{aligned} &\kappa^2(5\tau - 2\tau^2 + 9) + \varrho^2(9 - \tau)(\tau - 1)^2(\kappa - 1)^2 \\ &+ 2\kappa\varrho(1 - \tau)(9 + 2\tau - \tau^2)(1 - \kappa) \end{aligned} \right]}{2[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2}. \end{aligned} \quad (\text{A.86})$$

Both regulatory instruments increase welfare:

$$\begin{aligned}
W - W^\mu &= -\frac{s[4(9 - 5\tau) + \tau\mu(\tau + 1)(3\tau + 1)(4 - \mu - \tau\mu)]}{8(\tau + 3)^2} < 0, \\
W - W^e &= -\frac{s \left[\begin{aligned} &\kappa^2(9 - 5\tau)(\tau + 3)^2 \\ &+ 2\kappa\varrho(1 - \tau)(\tau + 3)(3 - \tau)(9 - 2\tau + \tau^2)(1 - \kappa) \\ &+ \varrho^2(81 - 45\tau + 21\tau^2 - \tau^3)(\tau - 1)^2(\kappa - 1)^2 \end{aligned} \right]}{2(\tau + 3)^2[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2} < 0 \quad (\text{A.87})
\end{aligned}$$

Welfare-equivalence W^*

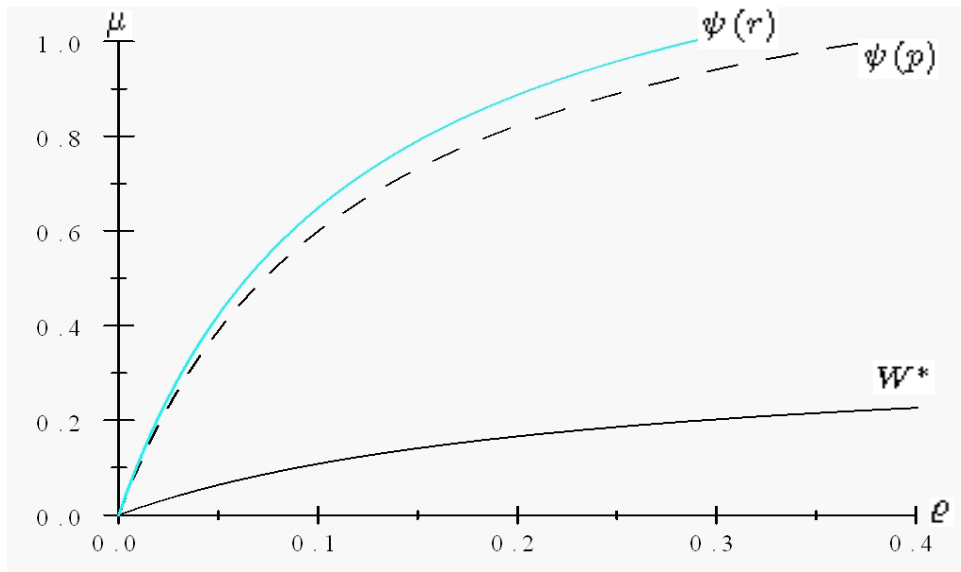
$$\begin{aligned}
W^\mu &= \frac{s \left[4(9 + 5\tau - 2\tau^2) + 4\tau\mu(\tau + 1)(3\tau + 1) - \tau\mu^2(3\tau + 1)(\tau + 1)^2 \right]}{8(\tau + 3)^2} \\
&= \frac{s \left[\begin{aligned} &\kappa^2(5\tau - 2\tau^2 + 9) + \varrho^2(9 - \tau)(\tau - 1)^2(\kappa - 1)^2 \\ &+ 2\kappa\varrho(1 - \tau)(9 + 2\tau - \tau^2)(1 - \kappa) \end{aligned} \right]}{2[\kappa(3 + \tau) + 3\varrho(1 - \tau)(1 - \kappa)]^2} = W^e \quad (\text{A.88})
\end{aligned}$$

$\tau = 0.1, \kappa = 0.1$

$$\begin{aligned}
W^\mu &= -2.046 \times 10^{-3}\mu^2 + 7.4402 \times 10^{-3}\mu + 0.49324 \\
&= \frac{7.209\kappa^2\varrho^2 - 16.542\kappa^2\varrho + 9.48\kappa^2 - 14.418\kappa\varrho^2 + 16.542\kappa\varrho + 7.209\varrho^2}{2(3.1\kappa + 2.7\varrho - 2.7\kappa\varrho)^2} = W^e \\
\mu &= \frac{6.318\varrho - 6.2\sqrt{0.85293\varrho^2 + 0.23166\varrho + 0.0169} + 0.806}{3.4749\varrho + 0.4433}
\end{aligned}$$

$\tau = 0.5, \kappa = 0.1$

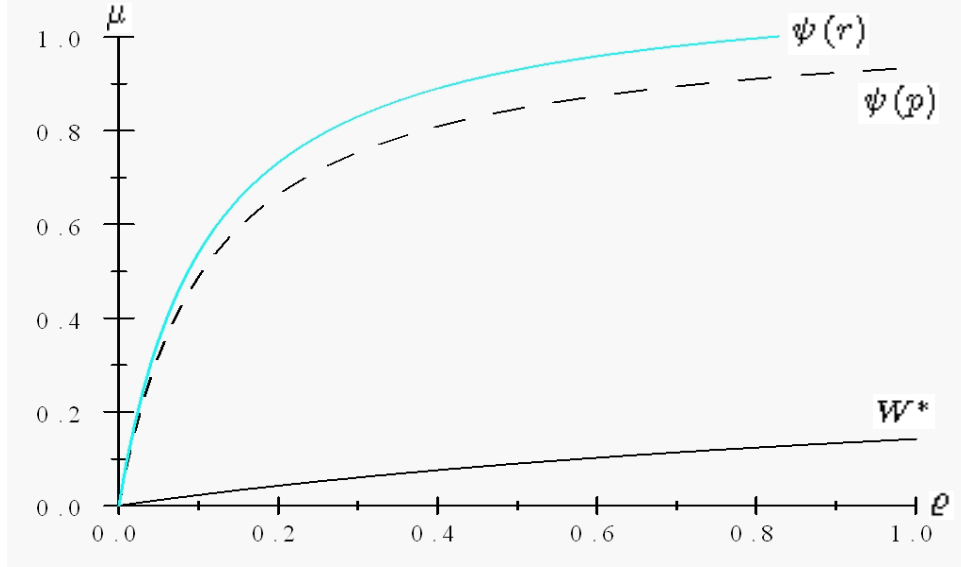
$$\begin{aligned}
W^\mu &= -2.8699 \times 10^{-2}\mu^2 + 7.6531 \times 10^{-2}\mu + 0.44898 \\
&= \frac{2.125\kappa^2\varrho^2 - 9.75\kappa^2\varrho + 11.0\kappa^2 - 4.25\kappa\varrho^2 + 9.75\kappa\varrho + 2.125\varrho^2}{2(35\kappa + 1.5\varrho - 1.5\kappa\varrho)^2} = W^e \\
\mu &= \frac{6.75\varrho - 7.0\sqrt{0.50625\varrho^2 + 0.3375\varrho + 0.0625} + 1.75}{5.0625\varrho + 1.3125}
\end{aligned}$$



Welfare, $\tau = 0.5$

$\tau = 0.9, \kappa = 0.1$

$$\begin{aligned}
 W^\mu &= -9.8794 \times 10^{-2} \mu^2 + 0.20799 \mu + 0.39053 \\
 &= \frac{0.081 \kappa^2 \rho^2 - 1.998 \kappa^2 \rho + 11.88 \kappa^2 - 0.162 \kappa \rho^2 + 1.998 \kappa \rho + 0.081 \rho^2}{2(3.9 \kappa + 0.3 \rho - 0.3 \kappa \rho)^2} = W^e \\
 \mu &= \frac{1.998 \rho - 7.8 \sqrt{0.02997 \rho^2 + 0.12654 \rho + 0.1369} + 2.886}{1.8981 \rho + 2.7417}
 \end{aligned}$$



Welfare, $\tau = 0.9$

Assuming identical price reductions under the maximum price system and the reference price system implies the following relationship between the regulatory parameters μ and ρ , see also price reductions-equivalence $\psi(p)$:

$$\begin{aligned}
 p_b^\mu &= \frac{s\tau(2 - \mu(1 + \tau))}{\kappa(3 + \tau)} = \frac{2s\tau(\kappa + \rho(1 - \tau)(1 - \kappa))}{(\kappa + \rho(1 - \kappa))(\kappa(3 + \tau) + 3\rho(1 - \tau)(1 - \kappa))} = p_b^\rho \\
 \mu &= \frac{2\kappa\rho(1 - \kappa)(\tau^2 + 3) + 6\rho^2(1 - \kappa)^2(1 - \tau)}{(1 + \tau)[\kappa + \rho(1 - \kappa)][\kappa(3 + \tau) + 3\rho(1 - \tau)(1 - \kappa)]}. \tag{A.89}
 \end{aligned}$$

Then total welfare under the maximum price system can be written in terms of the regulatory parameter ρ as:

$$W^\mu(\rho) = \frac{s \left[\begin{aligned} &\kappa^4(9 + 5\tau - 2\tau^2) + 2\kappa^3\rho(18 - \tau - 4\tau^2 + 3\tau^3)(1 - \kappa) \\ &+ \kappa^2\rho^2(54 - 37\tau - 3\tau^2 + 5\tau^3 - 3\tau^4)(1 - \kappa)^2 \\ &+ 12\kappa\rho^3(1 - \kappa)^3(1 - \tau)(3 - \tau) + 9\rho^4(1 - \kappa)^4(1 - \tau)^2 \end{aligned} \right]}{2(\kappa + \rho(1 - \kappa))^2(\kappa(3 + \tau) + 3\rho(1 - \tau)(1 - \kappa))^2}. \tag{A.90}$$

Taking identical price reductions as a basis of comparison, total welfare is higher under the maximum price system:

$$\begin{aligned}
& W^\mu - W^\varrho \Big|_{p_b^\mu = p_b^\varrho} \\
&= \frac{s\tau\varrho(1-\kappa) \left[\begin{array}{c} 2\kappa^3(1+\tau+2\tau^2) \\ +\kappa^2\varrho(5+2\tau^2-3\tau^3)(1-\kappa) \\ +4\kappa\varrho^2(1-\kappa)^2(1-\tau) + \varrho^3(1-\kappa)^3(1-\tau)^2 \end{array} \right]}{2[\kappa + \varrho(1-\kappa)]^2 [\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2} > 0. \quad (\text{A.91})
\end{aligned}$$

Total profits under the maximum price system:

$$\Pi_b^\mu + \Pi_g^\mu = \frac{s\tau[4(5-\tau) - \mu(\tau+1)[8(1-\tau) + \mu(\tau+1)(3\tau+1)]]}{4\kappa(\tau+3)^2} \quad (\text{A.92})$$

Total profits under the reference price system:

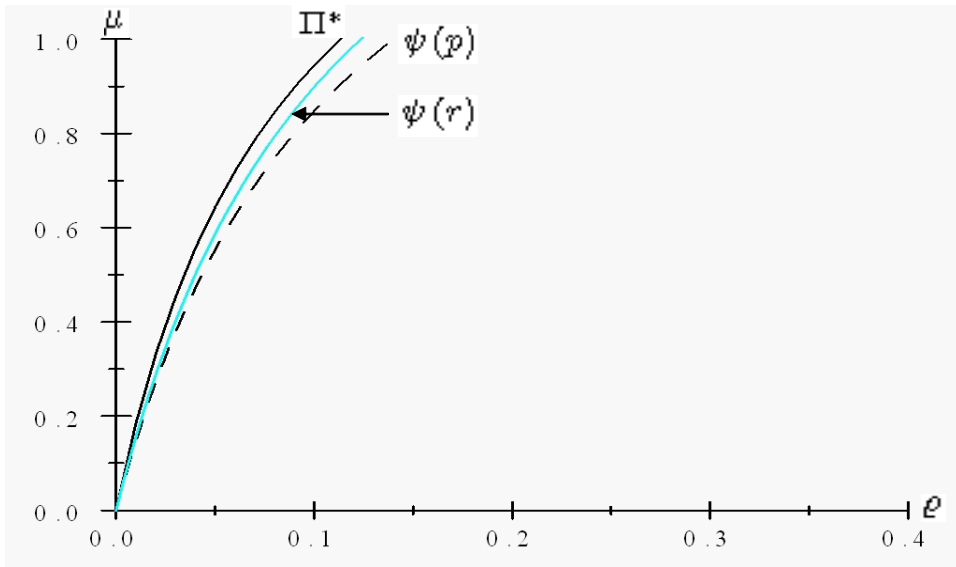
$$\Pi_b^\varrho + \Pi_g^\varrho = \frac{s\tau[\kappa + \varrho(1-\tau)(1-\kappa)][\kappa(5-\tau) + 5\varrho(1-\tau)(1-\kappa)]}{[\kappa + \varrho(1-\kappa)][\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2} \quad (\text{A.93})$$

Profit-equivalence Π^*

$$\begin{aligned}
& \Pi_b^\mu + \Pi_g^\mu = \frac{s\tau[4(5-\tau) - \mu(\tau+1)[8(1-\tau) + \mu(\tau+1)(3\tau+1)]]}{4\kappa(\tau+3)^2} \\
&= \frac{s\tau[\kappa + \varrho(1-\tau)(1-\kappa)][\kappa(5-\tau) + 5\varrho(1-\tau)(1-\kappa)]}{[\kappa + \varrho(1-\kappa)][\kappa(3+\tau) + 3\varrho(1-\tau)(1-\kappa)]^2} = \Pi_b^\varrho + \Pi_g^\varrho \quad (\text{A.94})
\end{aligned}$$

$\tau = 0.1, \kappa = 0.1$

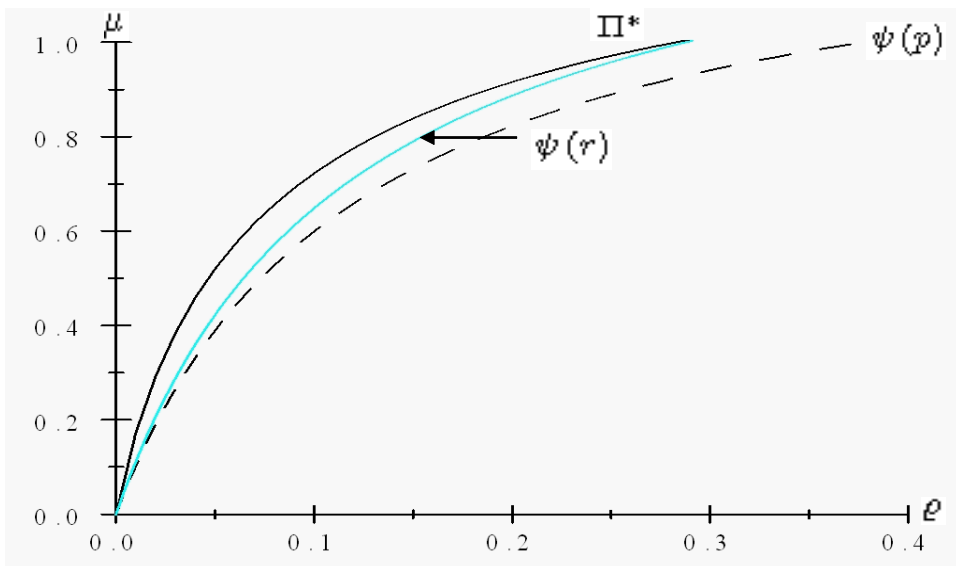
$$\mu = - \frac{8.748\varrho - (5.58\varrho + 0.62) \sqrt{\frac{5.3144\varrho^3 + 1.5200\varrho^2 + 0.1329\varrho + 0.00324}{(0.9\varrho + 0.1)^3}} + 1.116}{3.4749\varrho + 0.4433}$$



Profits, $\tau = 0.1$

$\tau = 0.5, \kappa = 0.1$

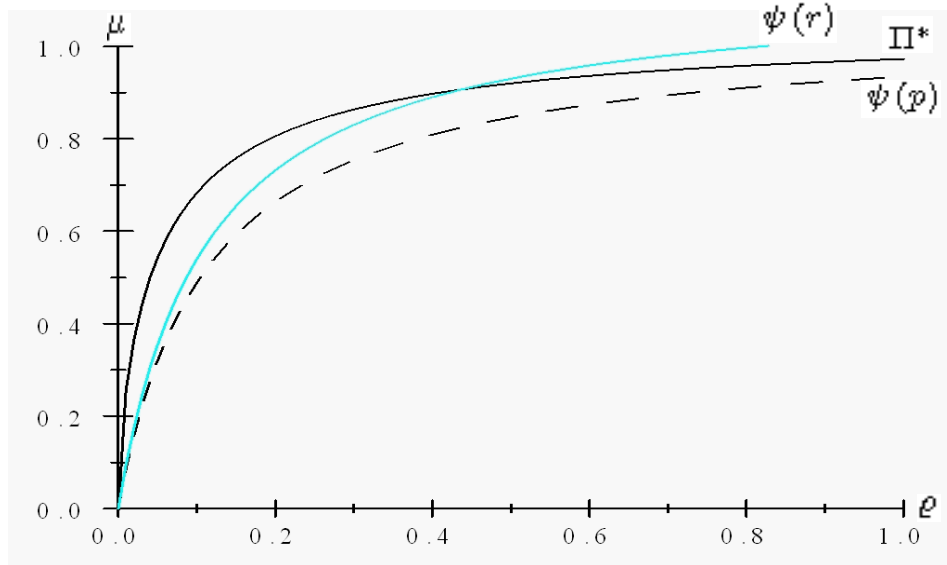
$$\mu = \frac{2.7\rho - (6.3\rho + 0.7) \sqrt{\frac{(1.6403\rho^3 + 0.77963\rho^2 + 9.7875 \times 10^{-2}\rho + 0.001)}{(0.9\rho + 0.1)^3}} + 0.7}{5.0625\rho + 1.3125}$$



Profits, $\tau = 0.5$

$\tau = 0.9, \kappa = 0.1$

$$\mu = \frac{(7.02\varrho + 0.78) \sqrt{\frac{(0.06561\varrho^3 + 0.18185\varrho^2 + 0.12765\varrho + 0.00004)}{(0.9\varrho + 0.1)^3}} - 0.156 - 0.108\varrho}{1.8981\varrho + 2.7417}$$



Profits, $\tau = 0.9$

Appendix B

Appendix to Chapter 4

B.1 The Effect of Parallel Trade

B.1.1 Competition Effect in the Destination Country

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

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

$$\frac{p_{b,D}^*}{p_{b,D}} = \frac{4\gamma_D(9-5\tau) + 4\tau\gamma_S(1-\tau)(3+\tau)}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2} < 1, \quad (\text{B.1})$$

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

$$\frac{p_{\beta,D}^*}{p_{b,D}^*} = (1-\tau) \frac{2\gamma_D(9-5\tau) + \tau\gamma_S(3+\tau)(1-\tau)}{2\gamma_D(9-5\tau) + 2\tau\gamma_S(3+\tau)(1-\tau)} < 1. \quad (\text{B.2})$$

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 version¹.

¹Note that the intermediary's best response function is $p_{\beta,H}^* = \frac{1}{2}(w + p_{b,H}^*(1-\tau))$.

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:

$$\frac{q_{b,D}^*}{q_{b,D}} = \frac{4\gamma_D(9-5\tau) + 4\gamma_S(1-\tau)(3+\tau)}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2} > 1. \quad (\text{B.3})$$

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.

B.1.2 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.

Compared to segmented markets, the wholesale price w_S^* 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:

$$w_S^* > w_S = 0. \quad (\text{B.4})$$

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

$$\frac{p_{b,S}^*}{p_{b,S}} = \frac{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(27-4\tau+\tau^2)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]} > 1 \quad (\text{B.5})$$

and the higher price reduces the quantity sold:

$$\frac{q_{b,S}^*}{q_{b,S}} = \frac{4\gamma_D(9-5\tau) - \gamma_S(1-\tau)(9-16\tau-\tau^2)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]} < 1. \quad (\text{B.6})$$

B.1.3 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^2):

$$\begin{aligned} & \frac{p_{b,D}^* - p_{b,S}^*}{p_{b,D} - p_{b,S}} \\ = & 1 - \frac{\gamma_S (1 - \tau) [2\gamma_D (9 - 5\tau) + 3\gamma_S (\tau + 3) (1 - \tau)]}{(\gamma_S - \gamma_D) [4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]} < 1 \end{aligned} \quad (\text{B.7})$$

if $\gamma_S > \gamma_D$.

B.2 Change of Coinsurance Rates under Segmented Markets

B.2.1 Change of the Coinsurance Rate in the Destination Country

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

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

$$\frac{\partial q_{b,D}}{\partial \gamma_D} = -p_{b,D} < 0. \quad (\text{B.8})$$

As a result, the manufacturer lowers the drug price:

$$\frac{\partial p_{b,D}}{\partial \gamma_D} = -\frac{1}{2\gamma_D^2} < 0, \quad (\text{B.9})$$

leaving the effective consumer price unchanged

$$\frac{\partial \gamma_D p_{b,D}}{\partial \gamma_D} = 0^3. \quad (\text{B.10})$$

²Note 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 $\gamma_F > \gamma_H$.

Consequently, the quantity consumed is unchanged

$$\frac{\partial q_{b,D}}{\partial \gamma_D} = 0. \quad (\text{B.11})$$

Figure B.2.1 illustrates the effect of an increase in 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 rate⁴. In other words, as the effective consumer price is independent of the coinsurance rate, so is the quantity consumed.

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

$$\frac{\partial E_D}{\partial \gamma_D} = -\frac{1}{4\gamma_D^2} < 0. \quad (\text{B.12})$$

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 :

$$\frac{\partial p_{b,S}}{\partial \gamma_D} = 0, \quad \frac{\partial q_{b,S}}{\partial \gamma_D} = 0. \quad (\text{B.13})$$

³Note that the effective consumer price $\gamma_H p_{b,H} = \frac{1}{2}$ is independent of the coinsurance rate.

⁴As 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.

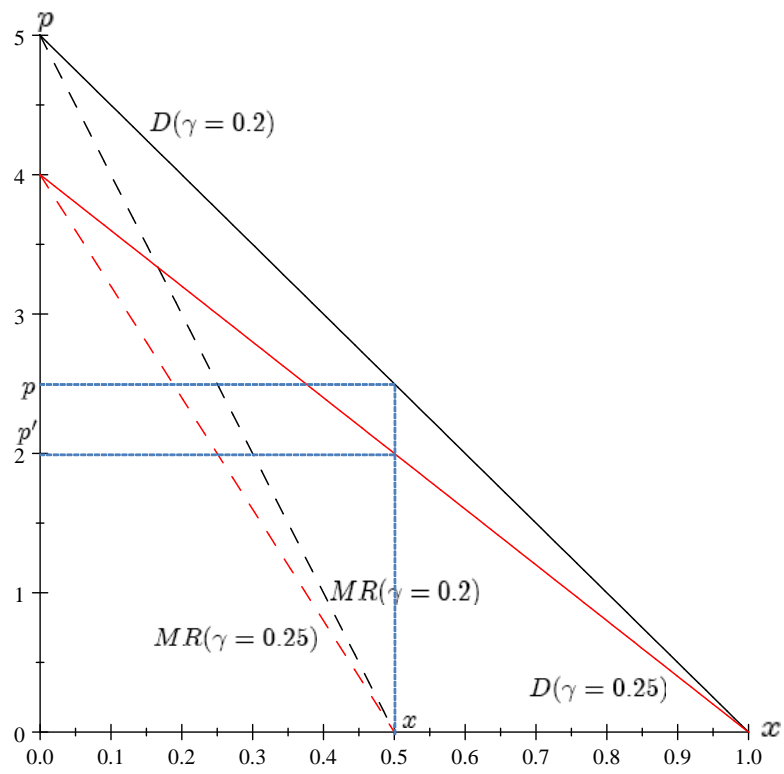


Figure B.2.1: Increase of the Coinsurance Rate, $c = 0$.

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

B.2.2 Change of the Coinsurance Rate 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 in the coinsurance rate in the destination country S increases price elasticity of demand and thus decreases demand for the drug:

$$\frac{\partial q_{b,S}}{\partial \gamma_S} = -p_{b,S} < 0. \quad (\text{B.14})$$

Consequently, the intermediary lowers the drug price:

$$\frac{\partial p_{b,S}}{\partial \gamma_S} = -\frac{1}{2\gamma_S^2} < 0, \quad (\text{B.15})$$

leaving the effective consumer price unchanged:

$$\frac{\partial \gamma_S p_{b,S}}{\partial \gamma_S} = 0^5. \quad (\text{B.16})$$

Also the quantity consumed is unchanged:

$$\frac{\partial q_{b,S}}{\partial \gamma_S} = 0. \quad (\text{B.17})$$

A lower drug price reduces public pharmaceutical expenditure:

$$\frac{\partial E_S}{\partial \gamma_S} = -\frac{1}{4\gamma_S^2} < 0. \quad (\text{B.18})$$

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)

⁵Note that the effective consumer price $\gamma_H p_{b,H} = \frac{1}{2}$ is independent of the coinsurance rate.

the coinsurance rate in the source country S :

$$\frac{\partial p_{b,S}}{\partial \gamma_D} = 0, \quad \frac{\partial q_{b,S}}{\partial \gamma_D} = 0. \quad (\text{B.19})$$

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.

B.3 Change of Coinsurance Rates under Parallel Trade

B.3.1 Change of the Coinsurance Rate 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 lowers effective consumer prices, increases the quantity consumed, and raises health expenditure in the source country S .

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

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

$$\frac{\partial q_{b,D}^*}{\partial \gamma_D} = -\frac{(p_{b,D}^* - p_{\beta,D}^*)}{\tau} < 0. \quad (\text{B.20})$$

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

$$\begin{aligned} \frac{\partial q_{\beta,D}^*}{\partial \gamma_D} &= \frac{(1 - \tau) p_{b,D}^* - p_{\beta,D}^*}{\tau(1 - \tau)} > 0, \\ \text{if } p_{\beta,D}^* &< (1 - \tau) p_{b,D}^*. \end{aligned} \quad (\text{B.21})$$

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 well⁶. This is demonstrated by the best response function: $p_{\beta,D}^* = \frac{1}{2} (w_S^* + p_{b,D}^* (1 - \tau))$.

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

$$\begin{aligned} \frac{\partial p_{b,D}^*}{\partial \gamma_D} &= -\frac{2[4\gamma_D^2 (9 - 5\tau)^2 + 8\tau\gamma_S\gamma_D (1 - \tau) (\tau + 3) (9 - 5\tau) + \tau\gamma_S^2 (1 - \tau)^2 (\tau + 3)^3]}{\gamma_D^2 [4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} < 0, \\ \frac{\partial p_{\beta,D}^*}{\partial \gamma_D} &= -\frac{(1 - \tau) \left[8\gamma_D^2 (9 - 5\tau)^2 + \tau\gamma_S^2 (1 - \tau)^2 (\tau + 3)^3 + 8\tau\gamma_S\gamma_D (1 - \tau) (\tau + 3) (9 - 5\tau) \right]}{\gamma_D^2 [4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} < 0. \end{aligned} \quad (\text{B.22})$$

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

$$\begin{aligned} \frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_D} &= \frac{6\gamma_S (1 - \tau)^2 (9 - 5\tau) (3 + \tau)}{[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} > 0, \\ \frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_D} &= \frac{2\gamma_S (1 - \tau)^2 (3 - \tau) (\tau + 3) (9 - 5\tau)}{[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} > 0. \end{aligned} \quad (\text{B.23})$$

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

$$\begin{aligned} \frac{\partial q_{b,D}^*}{\partial \gamma_D} &= -\frac{2\gamma_S (1 - \tau)^2 (27 - 6\tau - 5\tau^2)}{[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} < 0, \\ \frac{\partial q_{\beta,D}^*}{\partial \gamma_D} &= -\frac{\gamma_S (1 - \tau) (\tau + 3) 4 (9 - 5\tau)}{[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2]^2} < 0. \end{aligned} \quad (\text{B.24})$$

⁶In addition, B.21 implies that the intermediary has to lower $p_{\beta,H}^*$ in order to prevent a decrease of demand.

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

$$\frac{\partial E_D^*}{\partial \gamma_D} = - \frac{2\gamma_S \gamma_D (1-\tau)(9-5\tau)(\tau+3)^2 \left[\begin{array}{c} 6\gamma_D(9-5\tau) - 6\tau\gamma_S(1-\tau)(5-\tau) \\ + \gamma_S \gamma_D (9-5\tau)(1-\tau)^2 \end{array} \right]}{\gamma_D^2 \left[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2 \right]^3} - \frac{16\gamma_D^3(9-5\tau)^3 + \tau\gamma_S^3(5-\tau)(1-\tau)^3(\tau+3)^4}{\gamma_D^2 \left[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2 \right]^3} < 0 \quad (\text{B.25})$$

Spillovers of co-payment 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:

$$\frac{\partial w_S^*}{\partial \gamma_D} = - \frac{8(9-5\tau)^2(1-\tau)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2} < 0. \quad (\text{B.26})$$

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

$$\frac{\partial p_{b,S}^*}{\partial \gamma_D} = - \frac{4(1-\tau)(9-5\tau)^2}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2} < 0. \quad (\text{B.27})$$

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

$$\frac{\partial \gamma_S p_{b,S}^*}{\partial \gamma_D} = - \frac{4\gamma_S(1-\tau)(9-5\tau)^2}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2} < 0, \quad (\text{B.28})$$

which increases the quantity consumed:

$$\frac{\partial q_{b,S}^*}{\partial \gamma_D} = \frac{4\gamma_S(1-\tau)(9-5\tau)^2}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2} > 0. \quad (\text{B.29})$$

Health expenditure increases, as the effect from a higher quantity consumed exceeds the

effect of a lower drug price (see Appendix B.3):

$$\frac{\partial E_S^*}{\partial \gamma_D} = \frac{8\gamma_S (1-\tau)^2 (1-\gamma_S) (9-5\tau)^3}{\left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (\tau+3)^2\right]^3} > 0. \quad (\text{B.30})$$

B.3.2 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 D .

In the source country, the increase in co-payments, 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.:

$$\frac{\partial (1 - \gamma_S p_{b,S}^*)}{\partial \gamma_S} = -\gamma_S p_{b,S}^* < 0. \quad (\text{B.31})$$

Accordingly, the drug price decreases in γ_S :

$$\frac{\partial p_{b,S}^*}{\partial \gamma_S} = - \frac{16\gamma_D^2 (9-5\tau)^2 + \gamma_S (1-\tau) (\tau+3)^2 \left[\begin{array}{c} 8\gamma_D (9-5\tau) \\ + \gamma_S (1-\tau) (27-4\tau+\tau^2) \end{array} \right]}{2\gamma_S^2 [4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (\tau+3)^2]^2} < 0. \quad (\text{B.32})$$

The effective drug price increases, as marginal cost is no longer zero⁷:

$$\frac{\partial \gamma_S p_{b,S}^*}{\partial \gamma_S} = \frac{4\gamma_D (1-\tau) (9-5\tau)^2}{\left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (3+\tau)^2\right]^2} > 0 \quad (\text{B.33})$$

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

$$\frac{\partial q_{b,S}^*}{\partial \gamma_S} = -\frac{4\gamma_D (1-\tau) (5\tau-9)^2}{\left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (3+\tau)^2\right]^2} < 0. \quad (\text{B.34})$$

Similar to the effect of an increase of the co-payment in country D on drug prices and quantities in country D , the increase of the co-payment in country S results in a lower drug price and a lower quantity sold, which decreases health expenditure:

$$\begin{aligned} \frac{\partial E_S^*}{\partial \gamma_S} = & -\frac{4\gamma_S \gamma_D (1-\tau) (9-5\tau) \left[\begin{array}{l} \gamma_S (1-\tau) (567 - 36\tau + 262\tau^2 + 36\tau^3 + 3\tau^4) \\ -8\gamma_S^2 (1-\tau) (9-5\tau)^2 + 12\gamma_D (9-5\tau) (\tau+3)^2 \end{array} \right]}{4\gamma_S^2 \left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (\tau+3)^2\right]^3} \\ & - \frac{64\gamma_D^3 (9-5\tau)^3 - \gamma_S^3 (9-16\tau-\tau^2) (27-4\tau+\tau^2) (\tau+3)^2 (1-\tau)^3}{4\gamma_S^2 \left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (\tau+3)^2\right]^3} < 0. \quad (\text{B.35}) \end{aligned}$$

Spillovers of co-payment 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:

$$\frac{\partial w_S^*}{\partial \gamma_S} = -\frac{2(1-\tau)^2 (3+\tau)^2 (9-5\tau)}{\left[4\gamma_D (9-5\tau) + \gamma_S (1-\tau) (3+\tau)^2\right]^2} < 0. \quad (\text{B.36})$$

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

⁷Note that $\frac{\partial \left(\gamma_F \frac{1+w^* \gamma_E}{2\gamma_F}\right)}{\partial \gamma_F} = \frac{1}{2}w^*$. 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.

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:

$$\begin{aligned}\frac{\partial p_{b,D}^*}{\partial \gamma_S} &= -\frac{6(1-\tau)^2(27-6\tau-5\tau^2)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} < 0, \\ \frac{\partial p_{\beta,D}^*}{\partial \gamma_S} &= -\frac{2(1-\tau)^2(3-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} < 0.\end{aligned}\quad (\text{B.37})$$

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

$$\begin{aligned}\frac{\partial \gamma_D p_{b,D}^*}{\partial \gamma_S} &= -\frac{6\gamma_D(1-\tau)^2(3+\tau)(9-5\tau)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} < 0, \\ \frac{\partial \gamma_D p_{\beta,D}^*}{\partial \gamma_S} &= -\frac{2\gamma_D(1-\tau)^2(3-\tau)(\tau+3)(9-5\tau)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} < 0.\end{aligned}\quad (\text{B.38})$$

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

$$\begin{aligned}\frac{\partial q_{b,D}^*}{\partial \gamma_S} &= \frac{2\gamma_D(1-\tau)^2(27-6\tau-5\tau^2)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} > 0, \\ \frac{\partial q_{\beta,D}^*}{\partial \gamma_S} &= \frac{4\gamma_D(1-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(3+\tau)^2]^2} > 0.\end{aligned}\quad (\text{B.39})$$

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

$$\frac{\partial E_D^*}{\partial \gamma_S} = -\frac{2\gamma_S(1-\tau)^3(1-\gamma_D)(27-6\tau-5\tau^2)^2}{[4\gamma_D(9-5\tau)+\gamma_S(1-\tau)(\tau+3)^2]^3}.\quad (\text{B.40})$$

B.4 Change in Health Expenditure

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

$$E_D(\gamma) = (1-\gamma_D)p_{b,D}q_{b,D} = \frac{(1-\gamma_D)}{4\gamma_D}.\quad (\text{B.41})$$

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

$$E_S(\gamma) = (1 - \gamma_S) p_{b,S} q_{b,S} = \frac{(1 - \gamma_S)}{4\gamma_S}. \quad (\text{B.42})$$

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

$$\begin{aligned} E_D^* &= (1 - \gamma_D) (p_{b,D}^* q_{b,D}^* + p_{\beta,D}^* q_{\beta,D}^*) \\ &= (1 - \gamma_D) \left[\frac{4\gamma_D^2(9-5\tau)^2 + 2\gamma_S\gamma_D(1-\tau)(9-5\tau)(\tau+3)^2}{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2} \right. \\ &\quad \left. + \frac{\tau\gamma_S^2(5-\tau)(3+\tau)^2(1-\tau)^2}{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2} \right]. \end{aligned} \quad (\text{B.43})$$

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

$$\begin{aligned} E_S^* &= (1 - \gamma_S) p_{b,S}^* q_{b,S}^* \\ &= (1 - \gamma_S) \left[\frac{16\gamma_D^2(9-5\tau)^2 + 8\gamma_S\gamma_D(1-\tau)(9-5\tau)(3+\tau)^2}{4\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2} \right. \\ &\quad \left. - \frac{\gamma_S^2(1-\tau)^2(27-4\tau+\tau^2)(9-16\tau-\tau^2)}{4\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2} \right]. \end{aligned} \quad (\text{B.44})$$

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

$$\begin{aligned} \frac{\partial(E_S^*(\gamma))}{\partial\gamma_D} &= \frac{\partial(1 - \gamma_S) p_{b,S}^* q_{b,S}^*}{\partial\gamma_D} = (1 - \gamma_S) \left[\frac{\partial p_{b,S}^*}{\partial\gamma_D} q_{b,S}^* + p_{b,S}^* \frac{\partial q_{b,S}^*}{\partial\gamma_D} \right] > 0, \\ \text{since } &\underbrace{\frac{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(27-4\tau+\tau^2)}{2\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}}_{p_{b,S}^*} \underbrace{\frac{4\gamma_S(1-\tau)(9-5\tau)^2}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{\frac{\partial q_{b,S}^*}{\partial\gamma_D}} \\ &> \underbrace{\frac{4\gamma_D(9-5\tau) - \gamma_S(1-\tau)(9-16\tau-\tau^2)}{2[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}}_{q_{b,S}^*} \underbrace{\frac{4(1-\tau)(9-5\tau)^2}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{\frac{\partial p_{b,S}^*}{\partial\gamma_D}}. \end{aligned} \quad (\text{B.45})$$

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

coinsurance rate in the source country

$$\begin{aligned}
& \frac{\partial (E_D^*(\gamma))}{\partial \gamma_S} = \frac{\partial \left((1 - \gamma_D) \left(p_{b,D}^* q_{b,D}^* + p_{\beta,D}^* q_{\beta,D}^* \right) \right)}{\partial \gamma_S} \\
= & (1 - \gamma_D) \left[\frac{\partial p_{b,D}^*}{\partial \gamma_S} q_{b,D}^* + p_{b,D}^* \frac{\partial q_{b,D}^*}{\partial \gamma_S} + \frac{\partial p_{\beta,D}^*}{\partial \gamma_S} q_{\beta,D}^* + p_{\beta,D}^* \frac{\partial q_{\beta,D}^*}{\partial \gamma_S} \right] < 0, \\
\text{since} & \underbrace{\frac{2\gamma_D(9-5\tau) + 2\tau\gamma_S(3+\tau)(1-\tau)}{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}}_{p_{b,D}^*} \underbrace{\frac{2\gamma_D(1-\tau)^2(27-6\tau-5\tau^2)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{\frac{\partial q_{b,D}^*}{\partial \gamma_S}} \\
& + \underbrace{\frac{(1-\tau)[2\gamma_D(9-5\tau) + \tau\gamma_S(3+\tau)(1-\tau)]}{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}}_{p_{\beta,D}^*} \underbrace{\frac{4\gamma_D(1-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{\frac{\partial q_{\beta,D}^*}{\partial \gamma_S}} \\
< & \underbrace{\frac{6(1-\tau)^2(27-6\tau-5\tau^2)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{q_{b,D}^*} \underbrace{\frac{2[\gamma_D(9-5\tau) + \gamma_S(3+\tau)(1-\tau)]}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2}}_{\frac{\partial p_{b,D}^*}{\partial \gamma_S}} \\
& + \underbrace{\frac{2(1-\tau)^2(3-\tau)(3+\tau)(9-5\tau)}{[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}}_{q_{\beta,D}^*} \underbrace{\frac{\gamma_S(3+\tau)(1-\tau)}{4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2}}_{\frac{\partial p_{\beta,D}^*}{\partial \gamma_S}}. \tag{B.46}
\end{aligned}$$

B.5 Implications for Health Policy

Total welfare in the destination country is given as:

$$W_D^* = CS_D^* + \pi_M^* - E_D^*, \tag{B.47}$$

with

$$\begin{aligned}
CS_D^* &= \int_{\theta_D^{b,\beta}}^1 (\theta - \gamma_D p_{b,D}) d\theta + \int_{\theta_D^{\beta,0}}^{\theta_D^{b,\beta}} (\theta(1-\tau) - \gamma_D p_{\beta,D}) d\theta \\
&= \frac{(9-5\tau)[4\gamma_D^2(9-5\tau) + 4\gamma_D\gamma_S(1-\tau)(9-\tau^2) + \gamma_S^2(3+\tau)^2(1-\tau)^2]}{2[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}, \tag{B.48}
\end{aligned}$$

$$\pi_M^* = \frac{4\gamma_D^2(9-5\tau) + \gamma_S[\gamma_D[(1-\tau)(3+\tau)^2 + 4(9-5\tau)] + 4\tau\gamma_S(1-\tau)(5-\tau)]}{4\gamma_D\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]}, \quad (\text{B.49})$$

and

$$\begin{aligned} E_D^* &= (1-\gamma_D)(p_{b,D}^*q_{b,D}^* + p_{\beta,D}^*q_{\beta,D}^*) \\ &\quad (1-\gamma_D)[4\gamma_D^2(5\tau-9)^2 \\ &\quad + 2\gamma_S\gamma_D(1-\tau)(9-5\tau)(\tau+3)^2 \\ &\quad + \tau\gamma_S^2(5-\tau)(\tau+3)^2(1-\tau)^2] \\ &= \frac{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2}{\gamma_D[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2}. \end{aligned} \quad (\text{B.50})$$

Total welfare in the source country is given as:

$$W_S^* = CS_S^* - E_S^*. \quad (\text{B.51})$$

$$\begin{aligned} CS_S^* &= \int_{\theta_S^{2,0}}^1 (\theta - \gamma_S p_{b,S}) d\theta \\ &= \frac{[4\gamma_D(9-5\tau) - \gamma_S(1-\tau)(9-16\tau-\tau^2)]^2}{8[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2]^2}, \end{aligned} \quad (\text{B.52})$$

and

$$\begin{aligned} E_S^* &= (1-\gamma_S)p_{b,S}^*q_{b,S}^* \\ &\quad (1-\gamma_S)[16\gamma_D^2(9-5\tau)^2 \\ &\quad + 8\gamma_S\gamma_D(1-\tau)(9-5\tau)(3+\tau)^2 \\ &\quad - \gamma_S^2(1-\tau)^2(27-4\tau+\tau^2)(9-16\tau-\tau^2)] \\ &= \frac{4\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2}{4\gamma_S[4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(\tau+3)^2]^2}. \end{aligned} \quad (\text{B.53})$$

Total welfare for both countries is given as:

$$W = \frac{\left[\begin{array}{l} 96\gamma_D^2 (9 - 5\tau)^2 + 8\gamma_S\gamma_D (1 - \tau) (9 - 5\tau) (40\tau + 3\tau^2 + 45) \\ -\gamma_S^2 (81 - 1044\tau - 230\tau^2 - 28\tau^3 + 5\tau^4) (1 - \tau)^2 \end{array} \right]}{8 \left[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (3 + \tau)^2 \right]^2}. \quad (\text{B.54})$$

Total welfare for both countries increases in γ_D :

$$\begin{aligned} & \frac{\partial W}{\partial \gamma_D} \quad (\text{B.55}) \\ &= \frac{2\gamma_S (1 - \tau) (9 - 5\tau) \left[\begin{array}{l} 2\gamma_D (9 - 5\tau) (9 - 4\tau + 3\tau^2) \\ +\gamma_S (1 - \tau) (243 - 207\tau + 41\tau^2 + 15\tau^3 + 4\tau^4) \end{array} \right]}{\left[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (\tau + 3)^2 \right]^3} > 0. \quad (\text{B.56}) \end{aligned}$$

Total welfare for both countries decreases in γ_S :

$$\begin{aligned} & \frac{\partial W}{\partial \gamma_S} \quad (\text{B.57}) \\ &= \frac{-2\gamma_D (1 - \tau) (9 - 5\tau) \left[\begin{array}{l} 2\gamma_D (9 - 5\tau) (9 - 4\tau + 3\tau^2) \\ +\gamma_S (1 - \tau) (243 - 207\tau + 41\tau^2 + 15\tau^3 + 4\tau^4) \end{array} \right]}{\left[4\gamma_D (9 - 5\tau) + \gamma_S (1 - \tau) (\tau + 3)^2 \right]^3} < 0. \quad (\text{B.58}) \end{aligned}$$

Appendix C

Appendix to Chapter 5

C.1 Drug Quantities under Segmented Markets

Equilibrium quantities are given as:

D↓, S→	Coinsurance	Indemnity insurance	
Coinsurance	$q_{b,D}(\gamma) = \frac{1}{2},$ $q_{b,S}(\gamma) = \frac{1}{2}$	$q_{b,D}(\gamma) = \frac{1}{2},$ $q_{b,S}(\delta) = \frac{1+\delta_S}{2}$	(C.1)
Indemnity insurance	$q_{b,D}(\delta) = \frac{1+\delta_D}{2},$ $q_{b,S}(\gamma) = \frac{1}{2}$	$q_{b,D}(\delta) = \frac{1+\delta_D}{2},$ $q_{b,S}(\delta) = \frac{1+\delta_S}{2}.$	

C.2 Drug Prices and Quantities under Parallel Trade

The wholesale price is given as:

D↓, S→	Coinsurance	Indemnity insurance
Coinsurance	$w_S^*(\gamma, \gamma) = \frac{2(1-\tau)(9-5\tau)}{\Omega(\gamma, \gamma)}$	$w_S^*(\gamma, \delta) = \frac{2(1-\tau)(9-5\tau)}{\Omega(\gamma, \delta)}$
Indemnity insurance	$w_S^*(\delta, \gamma) = \frac{2(1-\tau)[(9-5\tau)+\delta_D(9-\tau)]}{\Omega(\delta, \gamma)}$	$w_S^*(\delta, \delta) = \frac{2(1-\tau)[(9-5\tau)+\delta_D(9-\tau)]}{\Omega(\delta, \delta)}$,

(C.2)

with $\Omega(\gamma, \gamma) = 4\gamma_D(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2$, $\Omega(\gamma, \delta) = 4\gamma_D(9-5\tau) + (1-\tau)(\tau+3)^2$, $\Omega(\delta, \gamma) = 4(9-5\tau) + \gamma_S(1-\tau)(3+\tau)^2$, and $\Omega(\delta, \delta) = 4(9-5\tau) + (1-\tau)(3+\tau)^2$.

Drug prices in both countries are given as:

D↓, S→	Coinsurance
Coinsurance	$p_{b,D}^*(\gamma, \gamma) = \frac{2\gamma_D(9-5\tau)+2\tau\gamma_S(3+\tau)(1-\tau)}{\gamma_D\Omega(\gamma, \gamma)},$ $p_{\beta,D}^*(\gamma, \gamma) = \frac{(1-\tau)[2\gamma_D(9-5\tau)+\tau\gamma_S(3+\tau)(1-\tau)]}{\gamma_D\Omega(\gamma, \gamma)},$ $p_{b,S}^*(\gamma, \gamma) = \frac{\Omega(\gamma, \gamma)+2\gamma_S(1-\tau)(9-5\tau)}{2\gamma_S\Omega(\gamma, \gamma)}$
Indemnity insurance	$p_{b,D}^*(\delta, \gamma) = \frac{2\delta_D(9-7\tau)+2(9-5\tau)+(\delta_D+2)\tau\gamma_S(\tau+3)(1-\tau)}{\Omega(\delta, \gamma)},$ $p_{\beta,D}^*(\delta, \gamma) = \frac{2(1-\tau)(9-5\tau)+2\delta_D(9-4\tau-\tau^2)+(\delta_D+2)\tau\gamma_S(\tau+3)(1-\tau)}{\Omega(\delta, \gamma)},$ $p_{b,S}^*(\delta, \gamma) = \frac{\Omega(\delta, \gamma)+2\gamma_S(1-\tau)[(9-5\tau)+(9-\tau)]}{2\gamma_S\Omega(\delta, \gamma)}$

D↓, S→	Indemnity insurance
Coinsurance	$p_{b,D}^*(\gamma, \delta) = \frac{2(\gamma_D(9-5\tau)+\tau(1-\tau)(\tau+3))}{\gamma_D\Omega(\gamma, \delta)},$ $p_{\beta,D}^*(\gamma, \delta) = \frac{(1-\tau)(2\gamma_D(9-5\tau)+\tau(1-\tau)(\tau+3))}{\gamma_D\Omega(\gamma, \delta)},$ $p_{b,S}^*(\gamma, \delta) = \frac{2(1-\tau)(9-5\tau)+(\delta_S+1)\Omega(\gamma, \delta)}{2\Omega(\gamma, \delta)}$
Indemnity insurance	$p_{b,D}^*(\delta, \delta) = \frac{2(9-2\tau-2\tau^2-\tau^3)+\delta_D(18-11\tau-2\tau^2-\tau^3)}{\Omega(\delta, \delta)},$ $p_{\beta,D}^*(\delta, \delta) = \frac{(1-\tau)(18-7\tau-2\tau^2-\tau^3)+\delta_D(9-\tau-3\tau^2-\tau^3)}{\Omega(\delta, \delta)},$ $p_{b,S}^*(\delta, \delta) = \frac{2(1-\tau)[(18+\tau+\tau^2)+\delta_D(9-\tau)]+(\delta_S+1)\Omega(\delta, \delta)}{2\Omega(\delta, \delta)}$

(C.3)

Equilibrium quantities are given as:

H, F	Coinsurance	
Coinsurance		$q_{b,D}^*(\gamma, \gamma) = \frac{2[\gamma_D(9-5\tau) + \gamma_S(3+\tau)(1-\tau)]}{\Omega(\gamma, \gamma)},$ $q_{\beta,D}^*(\gamma, \gamma) = \frac{\gamma_S(3+\tau)(1-\tau)}{\Omega(\gamma, \gamma)},$ $q_{b,S}^*(\gamma, \gamma) = \frac{\Omega(\gamma, \gamma) - 2\gamma_S(9-5\tau)(1-\tau)}{2\Omega(\gamma, \gamma)}$
Indemnity insurance		$q_{b,D}^*(\delta, \gamma) = \frac{2(9-5\tau) + 2\delta_D(3-\tau) + \gamma_S(\tau+3)(1-\tau)(\delta_D+2)}{\Omega(\delta, \gamma)},$ $q_{\beta,D}^*(\delta, \gamma) = \frac{4\delta_D(3-\tau) + (2\delta_D + (1-\tau))\gamma_S(\tau+3)(1-\tau)}{(1-\tau)\Omega(\delta, \gamma)},$ $q_{b,S}^*(\delta, \gamma) = \frac{\Omega(\delta, \gamma) - 2\gamma_S(1-\tau)[(9-5\tau) + \delta_D(9-\tau)]}{2\Omega(\delta, \gamma)}$
D↓, S→	Indemnity insurance	
Coinsurance		$q_{b,D}^*(\gamma, \delta) = \frac{2(\gamma_D(9-5\tau) + (\tau+3)(1-\tau))}{\Omega(\gamma, \delta)},$ $q_{\beta,D}^*(\gamma, \delta) = \frac{(\tau+3)(1-\tau)}{\Omega(\gamma, \delta)},$ $q_{b,S}^*(\gamma, \delta) = \frac{(1+\delta_S)\Omega(\gamma, \delta) - 2(9-5\tau)(1-\tau)}{2\Omega(\gamma, \delta)}$
Indemnity insurance		$q_{b,D}^*(\delta, \delta) = \frac{2(12-7\tau-\tau^2) + \delta_D(9-4\tau-\tau^2)}{\Omega(\delta, \delta)},$ $q_{\beta,D}^*(\delta, \delta) = \frac{(3+\tau)(1-\tau)^2 + 2\delta_D(9-4\tau-\tau^2)}{(1-\tau)\Omega(\delta, \delta)},$ $q_{b,S}^*(\delta, \delta) = \frac{(1+\delta_S)\Omega(\delta, \delta) - 2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)]}{2\Omega(\delta, \delta)}.$

(C.4)

C.3 The Effect of Parallel Trade

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

Compared to segmented markets, parallel trade reduces the price of the drug sold directly by the manufacturer under both cost-sharing systems and independent of the cost-sharing system

in the source country:

$$\begin{aligned}
\frac{p_{b,D}^*(\gamma, \gamma)}{p_{b,D}(\gamma)} &= \frac{4\gamma_D(9-5\tau) + 4\tau\gamma_S(1-\tau)(3+\tau)}{\Omega(\gamma, \gamma)} < 1, \\
\frac{p_{b,D}^*(\gamma, \delta)}{p_{b,D}(\gamma)} &= \frac{4\gamma_D(9-5\tau) + 4\tau(\tau+3)(1-\tau)}{\Omega(\gamma, \delta)} < 1, \\
\frac{p_{b,D}^*(\delta, \gamma)}{p_{b,D}(\delta)} &= \frac{2(2\delta_D(9-7\tau) + 2(9-5\tau) + (\delta_D+2)\tau\gamma_S(\tau+3)(1-\tau))}{(\delta_D+1)\Omega(\delta, \gamma)} < 1, \\
\frac{p_{b,D}^*(\delta, \delta)}{p_{b,D}(\delta)} &= \frac{4(9-2\tau-2\tau^2-\tau^3) + 2\delta_D(18-11\tau-2\tau^2-\tau^3)}{(1+\delta_D)\Omega(\delta, \delta)} < 1. \tag{C.5}
\end{aligned}$$

Under both coinsurance and indemnity insurance, the price of the parallel import is lower than the price of the locally sourced version:

$$\begin{aligned}
\frac{p_{\beta,D}^*(\gamma, \gamma)}{p_{b,D}^*(\gamma, \gamma)} &= (1-\tau) \frac{2\gamma_D(9-5\tau) + \tau\gamma_S(3+\tau)(1-\tau)}{2\gamma_D(9-5\tau) + 2\tau\gamma_S(3+\tau)(1-\tau)} < 1, \\
\frac{p_{\beta,D}^*(\gamma, \delta)}{p_{b,D}^*(\gamma, \delta)} &= (1-\tau) \frac{(2\gamma_D(9-5\tau) + \tau(1-\tau)(\tau+3))}{(2\gamma_D(9-5\tau) + 2\tau(1-\tau)(\tau+3))} < 1, \\
\frac{p_{\beta,D}^*(\delta, \gamma)}{p_{b,D}^*(\delta, \gamma)} &= \frac{2(1-\tau)(9-5\tau) + 2\delta_D(9-4\tau-\tau^2) + (2\delta_D + (1-\tau))\gamma_S\tau(\tau+3)(1-\tau)}{2(9-5\tau) + 2\delta_D(9-7\tau) + (\delta_D+2)\gamma_S\tau(\tau+3)(1-\tau)} < 1, \\
\frac{p_{\beta,D}^*(\delta, \delta)}{p_{b,D}^*(\delta, \delta)} &= \frac{(1-\tau)(18-7\tau-2\tau^2-\tau^3) + \delta_D(18-2\tau-6\tau^2-2\tau^3)}{(18-4\tau-4\tau^2-2\tau^3) + \delta_D(18-11\tau-2\tau^2-\tau^3)} < 1. \tag{C.6}
\end{aligned}$$

Under coinsurance, the quantity of the locally sourced version is higher under parallel trade:

$$\begin{aligned}
\frac{q_{b,D}^*(\gamma, \gamma)}{q_{b,D}(\gamma)} &= \frac{4\gamma_D(9-5\tau) + 4\gamma_S(1-\tau)(3+\tau)}{\Omega(\gamma, \gamma)} > 1, \\
\frac{q_{b,D}^*(\gamma, \delta)}{q_{b,D}(\gamma)} &= \frac{4\gamma_D(9-5\tau) + 4(1-\tau)(\tau+3)}{\Omega(\gamma, \delta)} > 1, \tag{C.7}
\end{aligned}$$

Under indemnity insurance, the quantity of the locally sourced version is higher under parallel

trade, if the reimbursement amount is low:

$$\begin{aligned}
\frac{q_{b,D}^*(\delta, \gamma)}{q_{b,D}(\delta)} &= \frac{2(2(9-5\tau) + 2\delta_D(3-\tau) + \gamma_S(\tau+3)(1-\tau)(\delta_D+2))}{(\delta_D+1)\Omega(\delta, \gamma)} > 1, \\
\text{if } \delta_D &< \frac{\gamma_S(\tau+3)(1-\tau)^2}{8(3-2\tau) + \gamma_S(\tau+3)(1-\tau)(\tau+1)}, \\
\frac{q_{b,D}^*(\delta, \delta)}{q_{b,D}(\delta)} &= \frac{4(12-7\tau-\tau^2) + 2\delta_D(9-4\tau-\tau^2)}{(1+\delta_D)\Omega(\delta, \delta)} > 1, \\
\text{if } \delta_D &< \frac{(3+\tau)(1-\tau)^2}{(27-15\tau-3\tau^2-\tau^3)}. \tag{C.8}
\end{aligned}$$

However, the total quantity sold in country D , i.e. locally sourced version plus parallel import, is (always) higher than the sales volume under no parallel trade:

$$\begin{aligned}
&\frac{q_{b,D}^*(\delta, \gamma) + q_{\beta,D}^*(\delta, \gamma)}{q_{b,D}(\delta)} \\
= &\frac{4(1-\tau)(9-5\tau) + 4\delta_D(3-\tau)^2 + 6\gamma_S(3+\tau)(1-\tau)^2 + 2\gamma_S\delta_D(1-\tau)(3+\tau)(3-\tau)}{(1-\tau)(\delta_D+1)\Omega(\delta, \gamma)} > 1, \\
&\frac{q_{b,D}^*(\delta, \delta) + q_{\beta,D}^*(\delta, \delta)}{q_{b,D}(\delta)} \\
= &\frac{2(1-\tau)(27-16\tau-3\tau^2) + 2\delta_D(3-\tau)(9-4\tau-\tau^2)}{(1+\delta_D)(1-\tau)\Omega(\delta, \delta)} > 1. \tag{C.9}
\end{aligned}$$

In country S , the introduction of parallel trade induces an increase of the wholesale price w . This translates to an increase of the drug price (double marginalization effect) under both cost-sharing systems:

$$\begin{aligned}
\frac{p_{b,S}^*(\gamma, \gamma)}{p_{b,S}(\gamma)} &= \frac{\Omega(\gamma, \gamma) + 2\gamma_S(1-\tau)(9-5\tau)}{\Omega(\gamma, \gamma)} > 1, \\
\frac{p_{b,S}^*(\delta, \gamma)}{p_{b,S}(\gamma)} &= \frac{\Omega(\delta, \gamma) + 2\gamma_S(1-\tau)((9-5\tau) + (9-\tau))}{\Omega(\delta, \gamma)} > 1, \\
\frac{p_{b,S}^*(\gamma, \delta)}{p_{b,S}(\delta)} &= \frac{2(1-\tau)(9-5\tau) + (\delta_S+1)\Omega(\gamma, \delta)}{(\delta_S+1)\Omega(\gamma, \delta)} > 1, \\
\frac{p_{b,S}^*(\delta, \delta)}{p_{b,S}(\delta)} &= \frac{2(1-\tau)[(18+\tau+\tau^2) + \delta_D(9-\tau)] + (\delta_S+1)\Omega(\delta, \delta)}{(\delta_S+1)\Omega(\delta, \delta)} > 1. \tag{C.10}
\end{aligned}$$

Due to a higher price, parallel trade reduces the quantity sold:

$$\begin{aligned}
\frac{q_{b,S}^*(\gamma, \gamma)}{q_{b,S}(\gamma)} &= \frac{\Omega(\gamma, \gamma) - 2\gamma_S(9 - 5\tau)(1 - \tau)}{\Omega(\gamma, \gamma)} < 1, \\
\frac{q_{b,S}^*(\delta, \gamma)}{q_{b,S}(\gamma)} &= \frac{\Omega(\delta, \gamma) - 2\gamma_S(1 - \tau)[(9 - 5\tau) + \delta_D(9 - \tau)]}{\Omega(\delta, \gamma)} < 1, \\
\frac{q_{b,S}^*(\gamma, \delta)}{q_{b,S}(\delta)} &= \frac{(1 + \delta_S)\Omega(\gamma, \delta) - 2(9 - 5\tau)(1 - \tau)}{(\delta_S + 1)\Omega(\gamma, \delta)} < 1, \\
\frac{q_{b,S}^*(\delta, \delta)}{q_{b,S}(\delta)} &= \frac{(1 + \delta_S)\Omega(\delta, \delta) - 2(1 - \tau)[(9 - 5\tau) + \delta_D(9 - \tau)]}{(1 + \delta_S)\Omega(\delta, \delta)} < 1. \tag{C.11}
\end{aligned}$$

C.3.1 Changes in Co-payments and Public Pharmaceutical Expenditure in the Destination Country

Given that coinsurance is applied in the source country S , the relative co-payment change in the destination country D under coinsurance is given as:

$$\frac{c_{b,D}^*(\gamma, \gamma)}{c_{b,D}(\gamma)} = \frac{4\gamma_D(9 - 5\tau) + 4\tau\gamma_S(1 - \tau)(3 + \tau)}{\Omega(\gamma, \gamma)}, \tag{C.12}$$

which is equivalent to the price change under coinsurance, since $\frac{c_{b,D}^*(\gamma, \gamma)}{c_{b,D}(\gamma, \gamma)} = \frac{\gamma_D p_{b,D}^*(\gamma, \gamma)}{\gamma_D p_{b,D}(\gamma, \gamma)}$. The relative co-payment change in the destination country D under indemnity insurance is given as:

$$\frac{c_{b,D}^*(\delta, \gamma)}{c_{b,D}(\delta)} = \frac{2 \left[\begin{array}{l} 2(9 - 5\tau) + 2\tau\gamma_S(\tau + 3)(1 - \tau) \\ -3\gamma_S\delta_D(\tau + 3)(1 - \tau) - 6\delta_D(3 - \tau) \end{array} \right]}{(1 - \delta_D)\Omega(\delta, \gamma)}, \tag{C.13}$$

which does not coincide with the price change, since $\frac{c_{b,D}^*(\delta,\gamma)}{c_{b,D}(\delta,\gamma)} = \frac{p_{b,D}^*(\delta,\gamma)-\delta_D}{p_{b,D}(\delta,\gamma)-\delta_D}$. The difference between relative co-payment changes then is given as:

$$\begin{aligned}
& \frac{c_{b,D}^*(\gamma,\gamma)}{c_{b,D}(\gamma)} - \frac{c_{b,D}^*(\delta,\gamma)}{c_{b,D}(\delta)} \\
= & \frac{32\tau\gamma_D\delta_D(9-5\tau) - 12\gamma_S(\tau+3)(9-5\tau)(\tau-1)^2(1-\gamma_D)}{[1-\delta_D]\Omega(\gamma,\gamma)\Omega(\delta,\gamma)} \\
& + \frac{2\gamma_S\delta_D(1-\tau)(\tau+3) \begin{bmatrix} \gamma_S(3-2\tau)(1-\tau)(\tau+3)^2 \\ +2(27-36\tau+17\tau^2) \\ +2\gamma_D(3-\tau)(9-5\tau) \end{bmatrix}}{[1-\delta_D]\Omega(\gamma,\gamma)\Omega(\delta,\gamma)} \tag{C.14}
\end{aligned}$$

Identical drug prices as standard of comparison imply that the reimbursement amount under indemnity insurance can be written in terms of the coinsurance rate as follows:

$$p_{b,D}(\gamma) = \frac{1}{2\gamma_D} = \frac{1+\delta_D}{2} = p_{b,D}(\delta) \iff \delta_D = \frac{1}{\gamma_D} - 1. \tag{C.15}$$

Identical drug prices under segmented markets as standard of comparison imply coinsurance rates of $\gamma_D > 0.5$ or $\gamma_D > 0.6$, depending on the cost-sharing system in the source country.

Identical drug prices without parallel trade implies that the reimbursement amount under indemnity insurance can be written in terms of the coinsurance rate as follows: $p_{b,D}(\gamma) = \frac{1}{2\gamma_D} =$

$\frac{1+\delta_D}{2} = p_{b,D}(\delta) \iff \delta_D = \frac{1}{\gamma_D} - 1$. Taking into account that consumers are required to co-

pay a positive amount, the reimbursement amount δ_D is restricted to $\delta_D < \widetilde{\delta}_D$, with $\widetilde{\delta}_D(\delta, \gamma) =$

$\frac{2(1-\tau)(9-5\tau)+\tau\gamma_S(\tau+3)(1-\tau)^2}{2(3-\tau)^2+\gamma_S(1-\tau)(\tau+3)(3-\tau)}$ and $\widetilde{\delta}_D(\delta, \delta) = \frac{(1-\tau)(18-7\tau-2\tau^2-\tau^3)}{(27-21\tau+\tau^2+\tau^3)}$. In combination with $\delta_D = \frac{1}{\gamma_D} -$

1, this implies $\gamma_D > \frac{(3-\tau)(2(3-\tau)+\gamma_S(1-\tau)(\tau+3))}{4(9-10\tau+3\tau^2)+\gamma_S(1-\tau)(\tau+3)(3-\tau^2)}$ (a) for γ in S , $\gamma_D > \frac{(3-\tau)(9-4\tau-\tau^2)}{45-46\tau+6\tau^2+2\tau^3+\tau^4}$ (b) for

δ in S , with (a) $\in (0.5, 1)$, (b) $\in (0.6, 1)$. Also, for $\gamma_D < \gamma_S$ (parallel trade from the low price to

high price-country, (a) implies $\gamma_S > \frac{\sqrt{-2538\tau+1591\tau^2-476\tau^3+159\tau^4-58\tau^5+9\tau^6+1377-27+31\tau-13\tau^2+\tau^3}}{2(1-\tau)(\tau+3)(3-\tau^2)}$,

which is only satisfied for high γ_S .

The difference between relative co-payment changes is positive⁵:

$$\begin{aligned} & \frac{c_{b,D}^*(\gamma, \gamma)}{c_{b,D}(\gamma)} - \frac{c_{b,D}^*(\delta, \gamma)}{c_{b,D}(\delta)} \Big|_{\delta_D = \frac{1}{\gamma_D} - 1} \\ &= \frac{2(1 - \gamma_D) \left[\begin{array}{l} 2\gamma_S(\tau + 3)(1 - \tau)(27 - 36\tau + 17\tau^2) \\ + \gamma_S^2(1 - \tau)^2(3 - 2\tau)(\tau + 3)^3 \\ - 4\tau\gamma_D(9 - 5\tau)[4 + \gamma_S(1 - \tau)(\tau + 3)] \end{array} \right]}{[2\gamma_D - 1]\Omega(\gamma, \gamma)\Omega(\delta, \gamma)} > 0. \end{aligned} \quad (\text{C.16})$$

The wholesale price is higher under indemnity insurance:

$$w_S^*(\delta, \gamma) = \frac{2(1 - \tau)[(9 - 5\tau) + \delta_D(9 - \tau)]}{\Omega(\delta, \gamma)} > \frac{2(1 - \tau)(9 - 5\tau)}{\Omega(\gamma, \gamma)} = w_S^*(\gamma, \gamma).$$

Given that indemnity insurance is applied in the source country S , the relative co-payment change in the destination country D under coinsurance is given as:

$$\frac{c_{b,D}^*(\gamma, \delta)}{c_{b,D}(\gamma)} = \frac{4\gamma_D(9 - 5\tau) + 4\tau(\tau + 3)(1 - \tau)}{\Omega(\gamma, \delta)}, \quad (\text{C.17})$$

the relative co-payment change under indemnity insurance is given as:

$$\frac{c_{b,D}^*(\delta, \delta)}{c_{b,D}(\delta)} = \frac{2(2(9 - 2\tau - 2\tau^2 - \tau^3) - 3\delta_D(9 - 4\tau - \tau^2))}{(1 - \delta_D)\Omega(\delta, \delta)}. \quad (\text{C.18})$$

⁵Under lower coinsurance rates, the difference between relative co-payment changes would be negative, suggesting that co-payments change to a greater extent under coinsurance rates. This is due to lower coinsurance rates implying high reimbursement amounts, which correspond to negative copayments, i.e. subsidies. Lower drug prices under parallel trade then result in an increase of the subsidy and a ratio of co-payments $\frac{c_{b,H}^*(\delta, \delta)}{c_{b,H}(\delta)}$ larger than 1, misleading into interpreting this as an increase of co-payments under parallel trade and a negative difference between relative co-payment changes as a higher reduction of co-payments under coinsurance rates.

The difference between relative co-payment changes then is given as

$$\begin{aligned}
& \frac{c_{b,D}^*(\gamma, \delta)}{c_{b,D}(\gamma)} - \frac{c_{b,D}^*(\delta, \delta)}{c_{b,D}(\delta)} \\
&= \frac{2 \left[\begin{array}{c} 2\gamma_D \delta_D (9 - 5\tau) (9 - \tau - \tau^2 + \tau^3) \\ + \delta_D (1 - \tau) (\tau + 3) (81 - 99\tau + 25\tau^2 + 7\tau^3 + 2\tau^4) \\ - 6(\tau + 3) (9 - 5\tau) (\tau - 1)^2 (1 - \gamma_D) \end{array} \right]}{[1 - \delta_D] \Omega(\gamma, \delta) \Omega(\delta, \delta)}. \tag{C.19}
\end{aligned}$$

For identical prices, it is positive:

$$\begin{aligned}
& \frac{c_{b,D}^*(\gamma, \delta)}{c_{b,D}(\gamma)} - \frac{c_{b,D}^*(\delta, \delta)}{c_{b,D}(\delta)} \Big|_{\delta_D = \frac{1}{\gamma_D} - 1} \\
&= \frac{2(1 - \gamma_D) \left[\begin{array}{c} (1 - \tau) (\tau + 3) (81 - 99\tau + 25\tau^2 + 7\tau^3 + 2\tau^4) \\ + 4\tau\gamma_D (9 - 5\tau) (7 - 2\tau - \tau^2) \end{array} \right]}{[2\gamma_D - 1] \Omega(\gamma, \delta) \Omega(\delta, \delta)} > 0. \tag{C.20}
\end{aligned}$$

The wholesale price is higher under indemnity insurance:

$$w_S^*(\delta, \delta) = \frac{2(1 - \tau) [(9 - 5\tau) + \delta_D (9 - \tau)]}{\Omega(\delta, \delta)} > \frac{2(1 - \tau) (9 - 5\tau)}{\Omega(\gamma, \delta)} = w_S^*(\gamma, \delta).$$

The relative change in public pharmaceutical expenditure under coinsurance is given as:

$$\frac{E_D^*(\gamma, \cdot)}{E_D(\gamma)} = \frac{(1 - \gamma_D) \left(p_{b,D}^*(\gamma, \cdot) q_{b,D}^*(\gamma, \cdot) + p_{\beta,D}^*(\gamma, \cdot) q_{\beta,D}^*(\gamma, \cdot) \right)}{(1 - \gamma_D) p_{b,D}(\gamma) q_{b,D}(\gamma)}. \tag{C.21}$$

Given that coinsurance is applied in the source country S , the relative change in public pharmaceutical expenditure is

$$\frac{E_D^*(\gamma, \gamma)}{E_D(\gamma)} = \frac{[\Omega(\gamma, \gamma)]^2 - \gamma_S^2 (9 - 5\tau) (\tau + 3)^2 (1 - \tau)^3}{[\Omega(\gamma, \gamma)]^2} < 1, \tag{C.22}$$

given that indemnity insurance is applied in the source country S , the relative change in public

pharmaceutical expenditure is

$$\frac{E_D^*(\gamma, \delta)}{E_D(\gamma)} = \frac{[\Omega(\gamma, \delta)]^2 - (9 - 5\tau)(\tau + 3)^2(1 - \tau)^3}{[\Omega(\gamma, \delta)]^2} < 1. \quad (\text{C.23})$$

C.3.2 Changes in Co-payments and Public Pharmaceutical Expenditure in the Source Country

Under coinsurance, the change in co-payments in the destination country S is given as:

$$\begin{aligned} \Delta c_{b,S}(\gamma, \cdot) &= c_{b,S}^*(\gamma) - c_{b,S}(\gamma) \\ &= \gamma_S (p_{b,S}^*(\gamma, \cdot) - p_{b,S}(\gamma)) > 0, \end{aligned} \quad (\text{C.24})$$

since $p_{b,S}^*(\gamma, \cdot) > p_{b,S}(\gamma)$. Under indemnity insurance, the change in co-payments in S is given as

$$\begin{aligned} \Delta c_{b,S}(\delta, \cdot) &= c_{b,S}^*(\delta) - c_{b,S}(\delta) \\ &= p_{b,S}^*(\delta, \cdot) - p_{b,S}(\delta) > 0, \end{aligned} \quad (\text{C.25})$$

since $p_{b,S}^*(\delta, \cdot) > p_{b,S}(\delta)$.

Given that coinsurance is applied in the destination country D , the relative co-payment change in the source country S under coinsurance is given as:

$$\frac{c_{b,S}^*(\gamma, \gamma)}{c_{b,S}(\gamma)} = \frac{4\gamma_D(9 - 5\tau) + \gamma_S(1 - \tau)(27 - 4\tau + \tau^2)}{\Omega(\gamma, \gamma)} \quad (\text{C.26})$$

and under indemnity insurance as:

$$\frac{c_{b,S}^*(\gamma, \delta)}{c_{b,S}(\delta)} = \frac{(1 - \tau)(27 - 4\tau + \tau^2) + 4\gamma_D(9 - 5\tau)(1 - \delta_S) - \delta_S(1 - \tau)(\tau + 3)^2}{[1 - \delta_S]\Omega(\gamma, \delta)}, \quad (\text{C.27})$$

with the difference between relative co-payment changes then amounting to:

$$\begin{aligned} & \frac{c_{b,S}^*(\gamma, \gamma)}{c_{b,S}(\gamma)} - \frac{c_{b,S}^*(\gamma, \delta)}{c_{b,S}(\delta)} \\ = & \frac{2(9-5\tau)(1-\tau) \left[\begin{array}{c} -4\gamma_D(9-5\tau)(1-\gamma_S(1-\delta_S)) \\ -\gamma_S\delta_S(1-\tau)(\tau+3)^2 \end{array} \right]}{[1-\delta_S]\Omega(\gamma, \gamma)\Omega(\gamma, \delta)}. \end{aligned} \quad (\text{C.28})$$

Assuming identical drug prices allows to express the reimbursement amount under indemnity insurance in terms of the coinsurance rate in the following way:

$$p_{b,S}(\gamma) = \frac{1}{2\gamma_S} = \frac{1+\delta_S}{2} = p_{b,S}(\delta) \iff \delta_S = \frac{1}{\gamma_S} - 1. \quad (\text{C.29})$$

Similar to assuming identical drug prices in the destination country D implying high coinsurance rates, this standard of comparison for the source country requires rather high coinsurance rates due to the positive co-payment condition under indemnity insurance. Taking into account that consumers are required to co-pay a positive amount, the reimbursement amount is δ_S restricted to $\delta_S < \widetilde{\delta}_S$, with $\widetilde{\delta}_S(\gamma, \delta) = \frac{4\gamma_D(9-5\tau)+(1-\tau)(27-4\tau+\tau^2)}{(4\gamma_D(9-5\tau)+(1-\tau)(\tau+3)^2)}$ and $\widetilde{\delta}_S(\delta, \delta) = \frac{45-55\tau+19\tau^2-\tau^3}{27-21\tau+\tau^2+\tau^3}$. In combination with $\delta_S = \frac{1}{\gamma_S} - 1$, this implies $\gamma_S > \frac{(4\gamma_D(9-5\tau)+(1-\tau)(\tau+3)^2)}{8\gamma_D(9-5\tau)+(1-\tau)(2\tau^2+2\tau+36)}$ (a) for γ in D , $\gamma_S > \frac{27-21\tau+\tau^2+\tau^3}{72-76\tau+20\tau^2}$ (b) for δ in D , with (a) $\in (0.25, 0.5)$, (b) $\in (0.375, 0.5)$. Admittedly, this also allows for rather moderate coinsurance rates, but this would correspond to restrictions on the degree of vertical product differentiation τ or the coinsurance rate in the destination country. The general case, without further restrictions, however, assumes high coinsurance rates.

The difference between relative co-payment changes is negative:

$$\begin{aligned} & \frac{c_{b,S}^*(\gamma, \gamma)}{c_{b,S}(\gamma)} - \frac{c_{b,S}^*(\gamma, \delta)}{c_{b,S}(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1} \\ = & \frac{-2\gamma_S(1-\tau)(1-\gamma_S)(9-5\tau) \left[8\gamma_D(9-5\tau) + (1-\tau)(\tau+3)^2 \right]}{[2\gamma_S - 1]\Omega(\gamma, \gamma)\Omega(\gamma, \delta)} < 0 \end{aligned} \quad (\text{C.30})$$

As a higher ratio of relative co-payments implies higher price increases, this corresponds to co-payments being increased less under coinsurance.

The wholesale price is higher under coinsurance:

$$w_S^*(\gamma, \gamma) = \frac{2(1-\tau)(9-5\tau)}{\Omega(\gamma, \gamma)} > \frac{2(1-\tau)(9-5\tau)}{\Omega(\gamma, \delta)} = w_S^*(\gamma, \delta). \quad (\text{C.31})$$

Given that indemnity insurance is applied in the destination country D , the relative co-payment change in the source country S under coinsurance is given as:

$$\frac{c_{b,S}^*(\delta, \gamma)}{c_{b,S}(\gamma)} = \frac{4(9-5\tau) + \gamma_S(1-\tau)(27-4\tau+\tau^2) + 2\gamma_S\delta_D(1-\tau)(9-\tau)}{\Omega(\delta, \gamma)}, \quad (\text{C.32})$$

and under indemnity insurance as:

$$\frac{c_{b,S}^*(\delta, \delta)}{c_{b,S}(\delta)} = \frac{\left[\begin{array}{c} (1-\tau)(27-4\tau+\tau^2) + 4(9-5\tau) \\ -\delta_S [4(9-5\tau) + (1-\tau)(\tau+3)^2] + 2\delta_D(1-\tau)(9-\tau) \end{array} \right]}{(1-\delta_S)\Omega(\delta, \delta)}. \quad (\text{C.33})$$

The difference between relative co-payment changes then amounts to:

$$\begin{aligned} & \frac{c_{b,S}^*(\delta, \gamma)}{c_{b,S}(\gamma)} - \frac{c_{b,S}^*(\delta, \delta)}{c_{b,S}(\delta)} \\ &= \frac{2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)] \left[\begin{array}{c} -4(9-5\tau)(1-\gamma_S(1-\delta_S)) \\ -\gamma_S\delta_S(1-\tau)(\tau+3)^2 \end{array} \right]}{(1-\delta_S)\Omega(\delta, \gamma)\Omega(\delta, \delta)}. \end{aligned} \quad (\text{C.34})$$

Assuming identical drug prices under segmented markets, the difference between relative

co-payment changes is negative:

$$\begin{aligned}
& \frac{c_{b,S}^*(\delta, \gamma)}{c_{b,S}(\gamma)} - \frac{c_{b,S}^*(\delta, \delta)}{c_{b,S}(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1} \\
&= \frac{2\gamma_S(1-\tau)(1-\gamma_S) \left[\begin{array}{l} (9-5\tau)(81-\tau^3-5\tau^2-43\tau) \\ +\delta_D(9-\tau)(81-\tau^3-5\tau^2-43\tau) \end{array} \right]}{(1-2\gamma_S)\Omega(\delta, \gamma)\Omega(\delta, \delta)} < 0. \quad (\text{C.35})
\end{aligned}$$

This is, co-payments increase less under coinsurance.

The wholesale price is higher under coinsurance:

$$\begin{aligned}
w_S^*(\delta, \gamma) &= \frac{2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)]}{\Omega(\delta, \gamma)} \\
&> \frac{2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)]}{\Omega(\delta, \delta)} = w_S^*(\delta, \delta). \quad (\text{C.36})
\end{aligned}$$

In the source country, the relative change in public pharmaceutical expenditure under coinsurance is given as:

$$\frac{E_S^*(\cdot, \gamma)}{E_S(\gamma)} = \frac{(1-\gamma_S)p_{b,S}^*(\cdot, \gamma)q_{b,S}^*(\cdot, \gamma)}{(1-\gamma_S)p_{b,S}(\gamma)q_{b,S}(\gamma)}. \quad (\text{C.37})$$

Given that coinsurance is applied in the destination country D , the relative change in public pharmaceutical expenditure is:

$$\frac{E_S^*(\gamma, \gamma)}{E_S(\gamma)} = \frac{[\Omega(\gamma, \gamma)]^2 - 4\gamma_S^2(\tau-1)^2(5\tau-9)^2}{[\Omega(\gamma, \gamma)]^2} < 1, \quad (\text{C.38})$$

given that indemnity insurance is applied in country D , the relative change in public pharmaceutical expenditure is:

$$\frac{E_S^*(\delta, \gamma)}{E_S(\gamma)} = \frac{[\Omega(\delta, \gamma)]^2 - 4\gamma_S^2(1-\tau)^2 \left[\begin{array}{l} (9-5\tau)^2 + \delta_D^2(9-\tau)^2 \\ +2\delta_D(9-\tau)(9-5\tau) \end{array} \right]}{[\Omega(\delta, \gamma)]^2} < 1. \quad (\text{C.39})$$

For coinsurance in country D , the difference between relative health expenditure changes under coinsurance and indemnity insurance is given as:

$$\begin{aligned}
& \frac{E_S^*(\gamma, \gamma)}{E_S(\gamma)} - \frac{E_S^*(\gamma, \delta)}{E_S(\delta)} \\
= & \frac{2(5\tau^2 - 14\tau + 9) \left[\begin{array}{c} 16\gamma_D^2(9 - 5\tau)^2 - \gamma_S^2(9 - 16\tau - \tau^2)(\tau + 3)^2(1 - \tau)^2 \\ -8\gamma_S\gamma_D(1 - \tau)(9 - 5\tau)(\gamma_S(9 - 5\tau) - (\tau + 3)^2) \\ -2\gamma_S^2\delta_S(9 - 5\tau)(1 - \tau)\Omega(\gamma, \delta) \end{array} \right]}{(\delta_S + 1)[\Omega(\gamma, \gamma)]^2\Omega(\gamma, \delta)} \quad (\text{C.40})
\end{aligned}$$

Assuming identical drug prices as a standard of comparison, the difference between relative health expenditure changes is positive:

$$\begin{aligned}
& \frac{E_S^*(\gamma, \gamma)}{E_D(\gamma)} - \frac{E_S^*(\gamma, \delta)}{E_D(\delta)} \Big|_{\delta_S = \frac{1}{\gamma_S} - 1} \\
= & \frac{2\gamma_S(5\tau^2 - 14\tau + 9) \left[\begin{array}{c} 16\gamma_D^2(5\tau - 9)^2 + \gamma_S^2(\tau - 1)^2(\tau + 3)^4 \\ +8\tau\gamma_S\gamma_D(1 - \tau)(\tau + 11)(9 - 5\tau) \\ -2\gamma_S(9 - 5\tau)(\tau + 3)^2(\tau - 1)^2 \end{array} \right]}{[\Omega(\gamma, \gamma)]^2\Omega(\gamma, \delta)} > 0. \quad (\text{C.41})
\end{aligned}$$

$$\begin{aligned}
& \left[\begin{array}{c} 16\gamma_D^2(5\tau - 9)^2 + 8\tau\gamma_S\gamma_D(1 - \tau)(\tau + 11)(9 - 5\tau) \\ +\gamma_S^2(\tau - 1)^2(\tau + 3)^4 - 2\gamma_S(9 - 5\tau)(\tau + 3)^2(\tau - 1)^2 \end{array} \right] > 0 \\
\text{if } \gamma_D > (1 - \tau) & \frac{\sqrt{\gamma_S(9 - 5\tau)^3 [2(\tau + 3)^2 - \gamma_S(17\tau + 2\tau^2 + 9)]} - \tau\gamma_S(\tau + 11)(9 - 5\tau)}{4(9 - 5\tau)^2}.
\end{aligned}$$

The positive co-payment condition requires $\gamma_D > 0.5$. Combination with the condition above

results in

$$\begin{aligned} & \frac{1}{4}(1-\tau) \frac{\sqrt{\gamma_S(9-5\tau)^3 \left[2(\tau+3)^2 - \gamma_S(17\tau+2\tau^2+9) \right]} - \tau\gamma_S(\tau+11)(9-5\tau)}{(5\tau-9)^2} < 0.5 \\ \rightarrow \gamma_D & < (1-\tau) \frac{\sqrt{\gamma_S(9-5\tau)^3 \left[2(\tau+3)^2 - \gamma_S(17\tau+2\tau^2+9) \right]} - \tau\gamma_S(\tau+11)(9-5\tau)}{4(5\tau-9)^2}, \end{aligned}$$

which is not possible.

For indemnity insurance in country D , the difference between relative health expenditure changes under coinsurance and indemnity insurance amounts to:

$$\begin{aligned} & \frac{E_S^*(\delta, \gamma)}{E_S(\gamma)} - \frac{E_S^*(\delta, \delta)}{E_S(\delta)} \\ & = \frac{2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)] \begin{bmatrix} [\Omega(\delta, \gamma)]^2 \\ -2\Omega(\delta, \delta)\gamma_S^2(1-\tau)(9-5\tau)(\delta_S+1) \\ -2\Omega(\delta, \delta)\gamma_S^2(1-\tau)\delta_D(9-\tau)(\delta_S+1) \end{bmatrix}}{(\delta_S+1)\Omega(\delta, \delta)[\Omega(\delta, \gamma)]^2}. \end{aligned} \quad (\text{C.42})$$

Under identical drug prices as standard of comparison, this is positive:

$$\begin{aligned} & \frac{E_S^*(\delta, \gamma)}{E_S(\gamma)} - \frac{E_S^*(\delta, \delta)}{E_S(\delta)} \Big|_{\gamma_S = \frac{1}{\delta_S+1}} \\ & = \frac{2(1-\tau)[(9-5\tau) + \delta_D(9-\tau)] \begin{bmatrix} (27+5\tau-15\tau^2-\tau^3)\Omega(\delta, \delta) \\ +2\delta_S(\tau+3)(9-5\tau)\Omega(\delta, \delta) \\ -2\delta_D\delta_S(1-\tau)(9-\tau)\Omega(\delta, \delta) \\ -2\delta_D(1-\tau)(9-\tau)\Omega(\delta, \delta) \\ +16\delta_S^2(5\tau-9)^2 \end{bmatrix}}{(\delta_S+1)\Omega(\delta, \delta)[\Omega(\delta, \gamma)]^2} > 0, \end{aligned} \quad (\text{C.43})$$

$$\begin{aligned}
& \left[\begin{array}{c} (27 + 5\tau - 15\tau^2 - \tau^3) \Omega(\delta, \delta) \\ +2\delta_S (\tau + 3) (9 - 5\tau) \Omega(\delta, \delta) \\ -2\delta_D \delta_S (1 - \tau) (9 - \tau) \Omega(\delta, \delta) \\ -2\delta_D (1 - \tau) (9 - \tau) \Omega(\delta, \delta) \\ +16\delta_S^2 (5\tau - 9)^2 \end{array} \right] < 0 \\
\text{if } \delta_D > & \frac{1}{2} \frac{\Omega(\delta, \delta) (27 + 5\tau - 15\tau^2 - \tau^3) + 16\delta_S^2 (5\tau - 9)^2}{(1 - \tau) (9 - \tau) (\delta_S + 1) \Omega(\delta, \delta)} \\
& + \frac{1}{2} \frac{2\delta_S (\tau + 3) (9 - 5\tau) \Omega(\delta, \delta)}{(1 - \tau) (9 - \tau) (\delta_S + 1) \Omega(\delta, \delta)},
\end{aligned}$$

violated is by the positive co-payment-condition even for $\delta_S = 0$.

C.4 Identical Co-payments as Standard of Comparison

To make coinsurance and indemnity insurance comparable, the main body of the chapter assumes identical drug prices under segmented markets as standard of comparison. This implies a certain threshold of coinsurance rates.

Identical co-payments under segmented markets may provide another standard of comparison, which does not assume a certain level of coinsurance rates, but is subject to another, probably more severe limitation.

Identical co-payments under segmented markets correspond to identical quantities consumed and is intended to give a notion of financial exposure of patients and access to pharmaceuticals. It includes the different insurance effect of both cost-sharing systems. This is illustrated by the comparison between insurance and no insurance under segmented markets, under coinsurance, the insurance effect entirely absorbed by manufacturer, under indemnity insurance, the insurance effect benefits both manufacturer and consumers. Identical co-payments entail the reimbursement amount under indemnity insurance being zero:

$$\gamma_D p_{b,D}(\gamma) = \gamma_D \frac{1}{2\gamma_D} = \frac{1 + \delta_D}{2} - \delta_D = p_{b,D}(\delta) - \delta_D \iff \delta_D = 0. \quad (\text{C.44})$$

Under coinsurance, the insurance effect is entirely absorbed by the manufacturer and for consumers, coinsurance has the same effect as no reimbursement (effective prices under coinsurance correspond to effective prices under no insurance). Assuming identical co-payments then transfers this effect to indemnity insurance as well. But for coinsurance, the manufacturer is only able to absorb the insurance effect entirely under segmented markets, under parallel trade, competition prevents it from doing so. Assuming identical co-payments under segmented markets not only transfers this insurance-absorbance-effect to the indemnity insurance scheme under segmented markets, but also to the indemnity insurance scheme under parallel trade, corresponding to assuming no reimbursement under indemnity insurance for both cases and thus comparing coinsurance with no insurance. This results in co-payments decreasing always more under coinsurance, as coinsurance provides reimbursement, whereas indemnity insurance under this standard of comparison does not.

C.5 Degree of Product Differentiation

The locally sourced version and the parallel import are perceived as vertical differentiated products. Differences in appearing and packaging or uncertainty whether both versions of the drug are really identical cause consumers to associate a lower quality with the parallel import. Health policy may impact on the extent to which both versions are perceived as qualitative different, for instance, by providing information on the equivalence of parallel imports and locally sourced drugs, by specifying packaging standards and setting information requirements. The following section analyzes the impact of a change in the degree of product differentiation, more precisely, the extent of the quality difference between the two versions of the drug on the consequences of parallel trade.

As an increase in product differentiation parameter τ results in the two versions of the drug being perceived as more remote substitutes and accordingly, reduces the competitive pressure from parallel trade in the destination country, the double marginalization effect in the origin

country gains importance. The wholesale price is lowered

$$\frac{\partial w_S^*(\gamma)}{\partial \tau} < 0, \frac{\partial w_S^*(\delta)}{\partial \tau} < 0. \quad (\text{C.45})$$

In country D , the price of the locally sourced version increases in τ under both cost-sharing schemes:

$$\frac{\partial p_{b,D}^*(\gamma)}{\partial \tau} > 0, \frac{\partial p_{b,D}^*(\delta)}{\partial \tau} > 0. \quad (\text{C.46})$$

There are two effects of an increasing τ on the price for the parallel import, a direct negative effect from the best-response function and a positive strategic effect from the price of the locally sourced version. Under both cost-sharing schemes, the negative direct effect dominates and accordingly, the price of the parallel import decreases in τ :

$$\frac{\partial p_{\beta,D}^*(\gamma)}{\partial \tau} < 0, \frac{\partial p_{\beta,D}^*(\delta)}{\partial \tau} < 0. \quad (\text{C.47})$$

Under both cost-sharing schemes, both quantities decrease in τ :

$$\frac{\partial q_{b,D}^*(\gamma)}{\partial \tau} < 0, \frac{\partial q_{b,D}^*(\delta)}{\partial \tau} < 0, \quad (\text{C.48})$$

$$\frac{\partial q_{\beta,D}^*(\gamma)}{\partial \tau} < 0, \frac{\partial q_{\beta,D}^*(\delta)}{\partial \tau} < 0. \quad (\text{C.49})$$

With respect to the locally sourced version, the increase of the drug price causes a decrease in quantity sold. For the parallel import, even with a lower drug price, less consumers buy the parallel import, as the gross valuation of the parallel import decreases ($\theta(1 - \tau)$) so that consumers with a low valuation no longer purchase the drug.

With an increase in the price of the locally sourced version and a decrease in quantity, the increase in the degree of vertical product differentiation results in a lower competition effect. This is, the parameter τ not only relates to the degree of product differentiation, but also determines the degree of competition between the manufacturer and intermediary in country D .

If $\tau = 0$, both versions of the drug are considered to be of the same quality and competition in country D amounts to Bertrand competition with homogenous products. The wholesale price is set at its maximum. The intermediary prices at marginal cost, i.e. the wholesale price, and the manufacturer's best response is to set the same price. Market coverage depends on the degree of reimbursement, the total quantity of the drug sold is $\frac{\gamma_S + 2\gamma_D}{\gamma_S + 4\gamma_D}$ and $\frac{3}{5}(1 + \delta_D)$ under coinsurance and indemnity insurance, respectively. For $\tau = 1$, no quality is associated with the parallel import and gross utility from its purchase is zero. Consequently, there is a monopoly for the manufacturer. Under coinsurance, the case of $\tau = 1$ corresponds to the situation of segmented markets, when parallel trade is not allowed. The manufacturer sets the monopoly price $p_{b,D}^* = \frac{1}{2\gamma_D}$ and sells the monopoly quantity of $q_{b,D}^* = \frac{1}{2}$. Under indemnity insurance, the case of $\tau = 1$ is not equivalent to the situation of segmented markets, as demand for the locally sourced versions for $\tau = 1$ does not include the impact of the reimbursement amount and thus does not translate into the demand under no parallel trade. The manufacturer sets a price $p_{b,D}^*(\delta) = \frac{(\delta_D + 2)}{4}$, which is lower than the price, if parallel trade is not allowed, and sells a quantity $q_{b,D}^*(\delta) = \frac{(\delta_D + 2)}{4}$.

In country S , the price of the locally sourced version decreases in τ under both cost-sharing schemes:

$$\frac{\partial p_{b,S}^*(\gamma)}{\partial \tau} < 0, \quad \frac{\partial p_{b,S}^*(\delta)}{\partial \tau} < 0, \quad (\text{C.50})$$

while the quantity sold increases in τ :

$$\frac{\partial q_{b,S}^*(\gamma)}{\partial \tau} > 0, \quad \frac{\partial q_{b,S}^*(\delta)}{\partial \tau} > 0. \quad (\text{C.51})$$

As the wholesale price decreases and accordingly, marginal cost for the intermediary, the drug price decreases and consumers who did not buy before now buy the drug. This is, with respect to country S , via a reduction of the wholesale price, the increase in the degree of vertical product differentiation mitigates the double marginalization effect.

Consequently, as an increase in the extent of quality difference between the two versions

of the drug reduces both the competition effect in the destination country D and the double marginalization effect, in other words, it mitigates the negative implications of parallel trade on the manufacturer's profit, it is in the interest of the manufacturer to maintain a high degree of perceived differentiation between the two versions of the drug. For instance, the manufacturer may attempt to increase cross-country product differentiation, in order for the parallel import to appear more differentiated with respect to the locally sourced version.

Appendix D

Appendix to Chapter 6

D.1 The Effect of Parallel Trade

D.1.1 Demand

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

$$\theta_D^{b,0} - \gamma_D p_{b,D} = 0 \Leftrightarrow \theta_D^{b,0} = \gamma_D p_{b,D}. \quad (\text{D.1})$$

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

$$q_{b,D}^* = 1 - \gamma_D p_{b,D}. \quad (\text{D.2})$$

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

If parallel trade is legal (international exhaustion of intellectual property rights), consumers in country D have the choice between the locally sourced version (b) from the domestic intermediary I_D or the parallel import (β) from the foreign intermediary I_S . The marginal consumer

who is indifferent between buying the locally sourced version b and the parallel import β has a gross valuation $\theta_D^{b,\beta}$, given by

$$\theta_D^{b,\beta} - \gamma_D p_{b,D}^* = \theta_D^{b,\beta} (1 - \tau) - \gamma_D p_{\beta,D}^* \Leftrightarrow \theta_D^{b,\beta} = \frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau}, \quad (\text{D.3})$$

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

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

Consequently, in country D , if the parallel import is available, demand for the locally sourced version b and for the parallel import β is given by

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

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

$$\theta_S^{b,0} - \gamma_S p_{b,S} = 0 \Leftrightarrow \theta_S^{b,0} = \gamma_S p_{b,S}. \quad (\text{D.6})$$

Accordingly, in country S demand for b is given by

$$q_{b,S} = 1 - \gamma_S p_{b,S}. \quad (\text{D.7})$$

D.1.2 Equilibrium without Parallel Trade

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

prices in both countries – are independent.

The manufacturers profit is given as

$$\pi_M = \underbrace{w_D (1 - \gamma_D p_{b,D})}_{\pi_{w_b,D}} + \underbrace{w_S (1 - \gamma_S p_{b,S})}_{\pi_{w_b,S}} + \phi_D + \phi_S, \quad (\text{D.8})$$

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

For the intermediary I_D , profit is given as:

$$\pi_{I_D} = \underbrace{(p_{b,D} - w_D) (1 - \gamma_D p_{b,D})}_{\pi_{b,D}} - \phi_D \quad (\text{D.9})$$

and for the intermediary I_S as:

$$\pi_{I_S} = \underbrace{(p_{b,S} - w_S) (1 - \gamma_S p_{b,S})}_{\pi_{b,S}} - \phi_S, \quad (\text{D.10})$$

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

In country D , the intermediary I_D maximizes (D.9) with respect to $p_{b,S}$. The first order condition to this problem is

$$\underbrace{(1 - \gamma_D p_{b,D})}_I + \underbrace{(p_{b,D} - w_D) \left(\underbrace{-\gamma_S}_{\frac{\partial q_{b,D}}{\partial p_{b,D}}} \right)}_{II} = 0, \quad (\text{D.11})$$

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

In country S , the intermediary I_S maximizes (D.10) with respect to $p_{b,S}$. The first order

condition to this problem is

$$(1 - \gamma_S p_{b,S}) + (p_{b,S} - w_S) \underbrace{(-\gamma_S)}_{\frac{\partial q_{b,S}}{\partial p_{b,S}}} = 0, \quad (\text{D.12})$$

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

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

$$\phi_D = \pi_{b,D} = \frac{(1 - w_D\gamma_D)^2}{4\gamma_D} \text{ and } \phi_S = \pi_{b,S} = \frac{(1 - w_S\gamma_S)^2}{4\gamma_S}. \quad (\text{D.13})$$

in order to extract the intermediaries' profits. In the absence of parallel trade and for segmented markets, the manufacturer's optimal strategy is to set the wholesale price equal to the marginal cost of production, i.e. $w_D = w_S = 0$ ¹. 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

$$p_{b,D} = \frac{1}{2\gamma_D} \text{ and } p_{b,S} = \frac{1}{2\gamma_S}. \quad (\text{D.14})$$

Equilibrium quantities are

$$q_{b,D} = \frac{1}{2}, \quad q_{b,S} = \frac{1}{2}. \quad (\text{D.15})$$

The manufacturer's profit is

$$\pi_M = \frac{(1 - \gamma_D)^2}{4\gamma_D} + \frac{(1 - \gamma_S)^2}{4\gamma_S}. \quad (\text{D.16})$$

¹This result can also be obtained by substituting (D.13) and equilibrium prices into (D.8) and maximizing with respect to w_D and w_S .

D.1.3 Equilibrium with Parallel Trade

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

The manufacturer's profit is given as

$$\begin{aligned} \pi_M^* = & \underbrace{w_D^* \left(1 - \frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_{\pi_{w_b,D}^*} + \underbrace{w_S^* (1 - \gamma_S p_{b,S}^*)}_{\pi_{w_b,S}^*} \\ & + \underbrace{w_S^* \left(\frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1-\tau)} \right)}_{\pi_{w_\beta}^*} + \phi_D^* + \phi_S^*, \end{aligned} \quad (\text{D.17})$$

where $\pi_{w_b,D}^*$ denotes the wholesale profit from the intermediary I_D 's sales in country D , $\pi_{w_b,S}^*$ the wholesale profit from the intermediary I_S 's sales in country S , $\pi_{w_\beta}^*$ the wholesale profit from the intermediary I_S 's sales as parallel imports in country D , and ϕ_D^* and ϕ_S^* the fixed fees.

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

The intermediaries' profits are given as

$$\pi_{I_D}^* = \underbrace{(p_{b,D}^* - w_D^*) \left(1 - \frac{\gamma_D(p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_{\pi_{b,D}^*} - \phi_D^*, \quad (\text{D.18})$$

$$\text{and } \pi_{I_S}^* = \underbrace{(p_{b,S}^* - w_S^*) (1 - \gamma_S p_{b,S}^*)}_{\pi_{b,S}^*} + \underbrace{(p_{\beta,D}^* - w_S^*) \left(\frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} - \frac{\gamma_D p_{\beta,D}^*}{(1 - \tau)} \right)}_{\pi_{\beta,D}^*} - \phi_S^*, \quad (\text{D.19})$$

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

In country D , the domestic intermediary I_D maximizes (D.18) with respect to $p_{b,D}^*$ which yields the first order condition

$$\underbrace{\left(1 - \frac{\gamma_D (p_{b,D}^* - p_{\beta,D}^*)}{\tau} \right)}_I + \underbrace{(p_{b,D}^* - w_D) \left(-\frac{\gamma_D}{\tau} \right)}_{II} = 0 \quad (\text{D.20})$$

and the best response function

$$p_{b,D}^* = \frac{1}{2\gamma_D} (\tau + p_{\beta,D}^* \gamma_D + \gamma_D w_D) \quad (\text{D.21})$$

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

The foreign intermediary I_S maximizes (D.19) with respect to $p_{\beta,D}^*$ which yields the first order condition

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

and the best response function

$$p_{\beta,D}^* = \frac{1}{2} (w_S^* + p_{b,D}^* (1 - \tau)). \quad (\text{D.22})$$

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

In country S , the intermediary maximizes (D.19) with respect to $p_{b,S}^*$. The first order condition to this maximization problem is

$$(1 - \gamma_S p_{b,S}^*) + (p_{b,S}^* - w_S^*) (-\gamma_S) = 0, \quad (D.23)$$

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

With fixed fees of

$$\phi_D^* = \frac{(2\tau + \gamma_D w_S^* - \gamma_D w_D^* (\tau + 1))^2}{\underbrace{\tau \gamma_D (\tau + 3)^2}_{\pi_{b,D}^*}} \quad (D.24)$$

$$\text{and } \phi_S^* = \underbrace{\frac{(1 - w_S^* \gamma_S)^2}{4\gamma_S}}_{\pi_{b,S}^*} + \underbrace{\frac{(\gamma_D w_D^* (1 - \tau) + (1 - \tau) \tau - \gamma_D w_S^* (1 + \tau))^2}{\tau \gamma_D (1 - \tau) (\tau + 3)^2}}_{\pi_{\beta,D}^*} \quad (D.25)$$

the manufacturer extracts the intermediaries' total profits.

Substituting (D.24), (D.25), and equilibrium prices into (D.17) and maximizing with respect to w_D^* and w_S^* gives the wholesale prices $w_D^* = \frac{(1-\tau)(2\tau + \gamma_D w_S^*)}{\gamma_D(3\tau + 1)}$ and $w_S^* = \frac{2(1-\tau)(5\tau - \tau^2 + 2\gamma_D w_D^*(1-\tau))}{4\gamma_D(3\tau + 1) + \tau\gamma_S(1-\tau)(\tau + 3)^2}$. Wholesale prices mutually reinforce one another; a higher wholesale price in the destination country, w_D^* induces a higher wholesale price in the source country, w_S^* and vice versa.

Equilibrium wholesale prices are given as:

$$w_D^* = \frac{2(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau + 1)(1-\tau))} \text{ and } w_S^* = \frac{2(1-\tau)}{4\gamma_D + \gamma_S(3\tau + 1)(1-\tau)}. \quad (D.26)$$

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

retail prices and reduce competition from parallel trade in the destination country.

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

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

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

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

first order condition:

$$\begin{aligned}
\frac{\partial \pi_M^*}{\partial w_D^*} &= \underbrace{\left(-\frac{2\gamma_D w_D^* (\tau + 1) - 2\tau - \gamma_D w_S^*}{\tau (\tau + 3)} \right)}_{\frac{\partial \pi_{w_b,D}^*}{\partial w_D^*}} + \underbrace{w_S^* \left(\frac{\gamma_D}{\tau (\tau + 3)} \right)}_{\frac{\partial \pi_{w_\beta}^*}{\partial w_D^*}} \\
&+ \underbrace{\left(-\frac{2(\tau + 1)(2\tau + \gamma_D w_S^* - \gamma_D w_D^* (\tau + 1))}{\tau (\tau + 3)^2} \right)}_{\frac{\partial \phi_D^*}{\partial w_D^*}} \\
&+ \underbrace{\left(\frac{2(\tau(1 - \tau) - \gamma_D w_S^* (\tau + 1) + \gamma_D w_D^* (1 - \tau))}{\tau (\tau + 3)^2} \right)}_{\frac{\partial \phi_S^*}{\partial w_D^*}} = 0. \tag{D.27}
\end{aligned}$$

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

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

$$\frac{\partial \pi_{w_b,D}^*}{\partial w_D^*} = \underbrace{\left(\frac{2\tau + \gamma_D w_S^* - \gamma_D w_D^* (\tau + 1)}{\tau (\tau + 3)} \right)}_{\frac{\partial w_D^*}{\partial w_D^*} q_{b,H}^*} + \underbrace{w_D^* \left(-\frac{2\gamma_D}{(\tau + 3)\tau} \right)}_{w_D^* \frac{\partial q_{b,H}^*}{\partial p_{b,H}^*} \frac{\partial p_{b,H}^*}{\partial w_D^*}} + \underbrace{w_D^* \left(\frac{\gamma_D (1 - \tau)}{(\tau + 3)\tau} \right)}_{w_D^* \frac{\partial q_{b,H}^*}{\partial p_{\beta,H}^*} \frac{\partial p_{\beta,H}^*}{\partial w_D^*}}. \tag{D.28}$$

The first part illustrates the wholesale profit-increasing effect of a higher wholesale price per unit sold, while ignoring changes in quantity. The second part gives the standard direct effect

of an increase of w_D^* . Via the increase of the price for the locally sourced version, an increase of w_D^* decreases demand for the locally sourced version and the wholesale profit. The third part indicates the strategic effect: An increase in w_D^* raises the price for the parallel import as well, thus it increases indirectly the demand for the locally sourced version. Similarly, the impact of the choice of w_D^* on the other components of the manufacturer's profit can be decomposed.

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

$$\begin{aligned}
\frac{\partial \pi_M^*}{\partial w_S^*} &= \underbrace{\left(\frac{1}{\tau} \gamma_D \frac{w_D^*}{\tau + 3} \right)}_{\frac{\partial \pi_{w_b,D}^*}{\partial w_S^*}} + \underbrace{\left(-\frac{(2\gamma_S w_S^* - 1)}{2} \right)}_{\frac{\partial \pi_{w_b,S}^*}{\partial w_S^*}} \\
&+ \underbrace{\left(-\frac{2\gamma_D w_S^* (\tau + 1) - \tau (1 - \tau) - \gamma_D w_D^* (1 - \tau)}{\tau (\tau + 3) (1 - \tau)} \right)}_{\frac{\partial \pi_{w_\beta}^*}{\partial w_S^*}} + \underbrace{\left(\frac{2(2\tau + \gamma_D w_S^* - \gamma_D w_D^* (1 + \tau))}{\tau (\tau + 3)^2} \right)}_{\frac{\partial \phi_D^*}{\partial w_S^*}} \\
&+ \underbrace{\left(-\frac{(1 - \gamma_S w_S^*)}{2} + -\frac{2(\tau + 1)(\tau(1 - \tau) - \gamma_D w_S^*(\tau + 1) + \gamma_D w_D^*(1 - \tau))}{\tau(1 - \tau)(\tau + 3)^2} \right)}_{\frac{\partial \phi_S^*}{\partial w_S^*}} = 0. \quad (\text{D.29})
\end{aligned}$$

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

Equilibrium drug prices are

$$p_{b,D}^* = \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}, p_{\beta,D}^* = \frac{(1-\tau)(2\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \text{ and} \quad (\text{D.30})$$

$$p_{b,S}^* = \frac{4\gamma_D + 3\gamma_S(1-\tau^2)}{2\gamma_S(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (\text{D.31})$$

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

$$\frac{p_{b,D}^*}{p_{b,D}} = \frac{4(\gamma_D + \tau\gamma_S(1-\tau))}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 1. \quad (\text{D.32})$$

The total change of the drug price is the net of a competition effect and a double marginalization effect². In country S , the drug price is higher under parallel trade, as the wholesale price w_S is higher under parallel trade:

$$\frac{p_{b,S}^*}{p_{b,S}} = \frac{(4\gamma_D + 3\gamma_S(1-\tau^2))}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 1. \quad (\text{D.33})$$

Equilibrium quantities are

$$q_{b,D}^* = \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)}, q_{\beta,D}^* = \frac{(1-\tau)\gamma_S}{4\gamma_D + \gamma_S(3\tau+1)(1-\tau)} \text{ and} \quad (\text{D.34})$$

$$q_{b,S}^* = \frac{4\gamma_D - \gamma_S(1-3\tau)(1-\tau)}{2(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (\text{D.35})$$

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

²For $w_D^* = 0$, the drug price would be $p_{b,D}^*(w_D^* = 0) = \frac{2\tau(\tau+3)(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)} < \frac{2(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} = p_{b,D}^*$.

higher under parallel trade than under segmented markets.

$$\begin{aligned}\frac{q_{b,D}^*}{q_{b,D}} &= \frac{8(\gamma_D + \tau\gamma_S(1-\tau))}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 1, \text{ if } \gamma_D < \frac{(1-6\tau+5\tau^2)}{4}\gamma_S, \\ \frac{q_{b,D}^* + q_{\beta,D}^*}{q_{b,D}} &= \frac{8\gamma_D + 2\gamma_S(1-\tau)(4\tau+1)}{(4\gamma_D + \gamma_S(1-\tau)(3\tau+1))} > 1.\end{aligned}\quad (\text{D.36})$$

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

$$\frac{q_{b,S}^*}{q_{b,S}} = \frac{(4\gamma_D - \gamma_S(3\tau-1)(\tau-1))}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 1. \quad (\text{D.37})$$

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

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

$$\begin{aligned}\frac{\pi_M^*}{\pi_M} &= \frac{(\gamma_S^2\gamma_D + \gamma_S\gamma_D^2 - 4\gamma_S\gamma_D + \gamma_S + \gamma_D)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}{4\gamma_D^2 + \gamma_S\gamma_D(\tau+1)(5-3\tau) + 4\tau\gamma_S^2(1-\tau)} < 1, \\ \text{if } \gamma_S < \gamma_S^* &= \frac{4\gamma_D(3\tau+1)(1-\tau) + \gamma_D^2(\tau+1)(3\tau-5) - (1-\tau)^2 + \sqrt{\Omega}}{2\gamma_D(3\tau+1)(1-\tau)},\end{aligned}$$

with $\Omega = (1 - \tau)^4 - 8\gamma_D(3\tau + 1)(1 - \tau)^3 + 2\gamma_D^2(50\tau + 69\tau^2 + 13)(1 - \tau)^2 + \gamma_D^3(3 - 2\tau + 3\tau^2)(8(3\tau + 1)(1 - \tau) + \gamma_D(3 - 2\tau + 3\tau^2))$.

D.2 The Effect of Contract Choice on the Effectiveness of Regulation under Segmented Markets

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

D.2.1 Two-part tariff

No regulation

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

The manufacturer's profit is given as

$$\pi_M = \underbrace{w_D(1 - \gamma_D p_{b,D})}_{\pi_{w_b,D}} + \underbrace{w_S(1 - \gamma_S p_{b,S})}_{\pi_{w_b,S}} + \phi_D + \phi_S, \quad (\text{D.38})$$

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

For the intermediary I_D , profit is given as:

$$\pi_{I_D} = \underbrace{(p_{b,D} - w_D)(1 - \gamma_D p_{b,D})}_{\pi_{b,D}} - \phi_D \quad (\text{D.39})$$

and for the intermediary I_S as:

$$\pi_{I_S} = \underbrace{(p_{b,S} - w_S)(1 - \gamma_S p_{b,S})}_{\pi_{b,S}} - \phi_S, \quad (\text{D.40})$$

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

In country D , the intermediary I_D maximizes (D.39) with respect to $p_{b,D}$. The first order condition to this problem is

$$(1 - \gamma_D p_{b,D}) + (p_{b,D} - w_D)(-\gamma_D) = 0, \quad (\text{D.41})$$

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

In country S , the intermediary I_S maximizes (D.40) with respect to $p_{b,S}$. The first order condition to this problem is

$$(1 - \gamma_S p_{b,S}) + (p_{b,S} - w_S)(-\gamma_S) = 0, \quad (\text{D.42})$$

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

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

$$\phi_D = \pi_{b,D} = \frac{(1 - w_D\gamma_D)^2}{4\gamma_D} \text{ and } \phi_S = \pi_{b,S} = \frac{(1 - w_S\gamma_S)^2}{4\gamma_S}. \quad (\text{D.43})$$

in order to extract the intermediaries' profits. In the absence of parallel trade and for segmented

markets, the manufacturer's optimal strategy is to set the wholesale price equal to the marginal cost of production, i.e. $w_D = w_S = 0^3$. 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

$$p_{b,D} = \frac{1}{2\gamma_D} \text{ and } p_{b,S} = \frac{1}{2\gamma_S}. \quad (\text{D.44})$$

Equilibrium quantities are

$$q_{b,D} = \frac{1}{2}, \quad q_{b,S} = \frac{1}{2}. \quad (\text{D.45})$$

Maximum Wholesale Margins

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

$$p_{b,D}^\mu = w_D^\mu + (1 - \mu_D) m_{b,D}, \quad m_{b,D} = p_{b,D} - w_D = \frac{1}{2\gamma_D} \quad (\text{D.46})$$

$$p_{b,S}^\mu = w_S^\mu + (1 - \mu_S) m_{b,S}, \quad m_{b,S} = p_{b,S} - w_S = \frac{1}{2\gamma_S}. \quad (\text{D.47})$$

The manufacturers profit is given as

$$\pi_M^\mu = \underbrace{w_D^\mu (1 - \gamma_D p_{b,D}^\mu)}_{\pi_{w_b,D}^\mu} + \underbrace{w_S^\mu (1 - \gamma_S p_{b,S}^\mu)}_{\pi_{w_b,S}^\mu} + \phi_D^\mu + \phi_S^\mu, \quad (\text{D.48})$$

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

³This result can also be obtained by substituting (D.43) and equilibrium prices into (D.38) and maximizing with respect to w .

For the intermediary I_D , profit is given as:

$$\pi_{I_D}^\mu = \underbrace{(1 - \mu_D) m_{b,D} (1 - \gamma_D p_{b,D}^\mu)}_{\pi_{b,D}^\mu} - \phi_D^\mu \quad (D.49)$$

and for the intermediary I_S as:

$$\pi_{I_S}^\mu = \underbrace{(1 - \mu_S) m_{b,S} (1 - \gamma_S p_{b,S}^\mu)}_{\pi_{b,S}^\mu} - \phi_S^\mu, \quad (D.50)$$

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

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

$$\begin{aligned} \phi_D^\mu &= \pi_{b,D}^\mu = (1 - \mu_D) m_{b,D} (1 - \gamma_D (w_D^\mu + (1 - \mu_D) m_{b,D})) \\ \text{and } \phi_S^\mu &= \pi_{b,S}^\mu = (1 - \mu_S) m_{b,S} (1 - \gamma_S (w_S^\mu + (1 - \mu_S) m_{b,S})) \end{aligned} \quad (D.51)$$

in order to extract the intermediaries' profits. Substituting (D.51) and equilibrium prices (D.46, D.47) into (D.48) and maximizing with respect to w_D^μ and w_S^μ gives the wholesale prices:

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

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

Equilibrium drug prices are

$$p_{b,D}^\mu = \frac{1}{2\gamma_D} \text{ and } p_{b,S}^\mu = \frac{1}{2\gamma_S}. \quad (\text{D.52})$$

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

Equilibrium quantities are

$$q_{b,D}^\mu = \frac{1}{2}, \quad q_{b,S}^\mu = \frac{1}{2}. \quad (\text{D.53})$$

Mandatory Rebates

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

D.2.2 Linear Pricing

No Regulation

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

The manufacturers profit is given as

$$\pi_M = \underbrace{w_D (1 - \gamma_D p_{b,D})}_{\pi_{w_b,D}} + \underbrace{w_S (1 - \gamma_S p_{b,S})}_{\pi_{w_b,S}}, \quad (\text{D.54})$$

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

For the intermediary I_D , profit is given as:

$$\pi_{I_D} = \underbrace{(p_{b,D} - w_D) (1 - \gamma_D p_{b,D})}_{\pi_{b,D}} \quad (\text{D.55})$$

and for the intermediary I_S as:

$$\pi_{I_S} = \underbrace{(p_{b,S} - w_S)(1 - \gamma_S p_{b,S})}_{\pi_{b,S}}, \quad (\text{D.56})$$

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

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

$$(1 - \gamma_D p_{b,D}) + (p_{b,D} - w_D)(-\gamma_S) = 0, \quad (\text{D.57})$$

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

In country S , the intermediary I_S maximizes (D.56) with respect to $p_{b,S}$. The first order condition to this problem is

$$(1 - \gamma_S p_{b,S}) + (p_{b,S} - w_S)(-\gamma_S) = 0, \quad (\text{D.58})$$

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

Substituting equilibrium prices into (D.54) and maximizing with respect to w_D and w_S gives the wholesale prices:

$$w_D = \frac{1}{2\gamma_D} \text{ and } w_S = \frac{1}{2\gamma_S}.$$

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

Equilibrium drug prices are

$$p_{b,D} = \frac{3}{4\gamma_D} \text{ and } p_{b,S} = \frac{3}{4\gamma_S}. \quad (\text{D.59})$$

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

Equilibrium quantities are

$$q_{b,D} = \frac{1}{4}, \quad q_{b,S} = \frac{1}{4}. \quad (\text{D.60})$$

Maximum Wholesale Margins

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

$$p_{b,D}^\mu = w_D^\mu + (1 - \mu_D) m_{b,D}, \quad m_{b,D} = p_{b,D} - w_D = \frac{1}{4\gamma_D} \quad (\text{D.61})$$

$$p_{b,S}^\mu = w_S^\mu + (1 - \mu_S) m_{b,S}, \quad m_{b,S} = p_{b,S} - w_S = \frac{1}{4\gamma_S}. \quad (\text{D.62})$$

The manufacturers profit is given as

$$\pi_M^\mu = \underbrace{w_D^\mu (1 - \gamma_D p_{b,D}^\mu)}_{\pi_{w_b,D}^\mu} + \underbrace{w_S^\mu (1 - \gamma_S p_{b,S}^\mu)}_{\pi_{w_b,S}^\mu}, \quad (\text{D.63})$$

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

For the intermediary I_D , profit is given as:

$$\pi_{I_D}^\mu = \underbrace{(1 - \mu_D) m_{b,D}}_{\pi_{b,D}^\mu} (1 - \gamma_D p_{b,D}^\mu) \quad (\text{D.64})$$

and for the intermediary I_S as:

$$\pi_{I_S}^\mu = \underbrace{(1 - \mu_S) m_{b,S}}_{\pi_{b,S}^\mu} (1 - \gamma_S p_{b,S}^\mu), \quad (\text{D.65})$$

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

Substituting equilibrium prices (D.61, D.62) into (D.63) and maximizing with respect to w_D^μ and w_S^μ gives the wholesale prices:

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

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

Equilibrium drug prices are

$$p_{b,D}^\mu = \frac{(1 + \gamma_D (1 - \mu_D) m_{b,D})}{2\gamma_D} = \frac{(5 - \mu_D)}{8\gamma_D} \quad (\text{D.66})$$

$$\text{and } p_{b,S}^\mu = \frac{(1 + \gamma_S (1 - \mu_S) m_{b,S})}{2\gamma_S} = \frac{(5 - \mu_S)}{8\gamma_S}. \quad (\text{D.67})$$

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

Equilibrium quantities are

$$q_{b,D}^\mu = \frac{(3 + \mu_D)}{8}, \quad q_{b,S}^\mu = \frac{(3 + \mu_S)}{8}. \quad (\text{D.68})$$

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

Mandatory Rebates

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

$$w_D^\psi = (1 - \psi_D) w_D = (1 - \psi_D) \frac{1}{2\gamma_D} \quad (\text{D.69})$$

and in country S to:

$$w_S^\psi = (1 - \psi_S) w_S = (1 - \psi_S) \frac{1}{2\gamma_S}. \quad (\text{D.70})$$

For the intermediary I_D , profit is given as:

$$\pi_{I_D}^\psi = \underbrace{\left(p_{b,D}^\psi - w_D^\psi \right)}_{\pi_{b,D}^\psi} \left(1 - \gamma_D p_{b,D}^\psi \right) \quad (\text{D.71})$$

and for the intermediary I_S as:

$$\pi_{I_S}^\psi = \underbrace{\left(p_{b,S}^\psi - w_S^\psi \right)}_{\pi_{b,S}^\psi} \left(1 - \gamma_S p_{b,S}^\psi \right), \quad (\text{D.72})$$

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

In country D , the intermediary I_D maximizes (D.71) with respect to $p_{b,S}^\psi$. The first order condition to this problem is

$$\left(1 - \gamma_D p_{b,D}^\psi \right) + \left(p_{b,D}^\psi - w_D^\psi \right) (-\gamma_S) = 0, \quad (\text{D.73})$$

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

In country S , the intermediary I_S maximizes (D.72) with respect to $p_{b,S}^\psi$. The first order

condition to this problem is

$$\left(1 - \gamma_S p_{b,S}^\psi\right) + \left(p_{b,S}^\psi - w_S^\psi\right) (-\gamma_S) = 0, \quad (\text{D.74})$$

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

Equilibrium drug prices are

$$p_{b,D}^\psi = \frac{(3 - \psi_D)}{4\gamma_D} \text{ and } p_{b,S}^\psi = \frac{(3 - \psi_S)}{4\gamma_S}. \quad (\text{D.75})$$

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

Equilibrium quantities are

$$q_{b,D}^\psi = \frac{(1 + \psi_D)}{4}, \quad q_{b,S}^\psi = \frac{(1 + \psi_S)}{4}. \quad (\text{D.76})$$

Quantities are higher than under no regulation.

D.3 Equilibrium with Maximum Wholesale Margins

Under maximum wholesale margins, the regulatory body of country j restricts the markup surcharged to a fraction μ_j of the markup surcharged under unregulated markets⁴:

$$\begin{aligned} p_{b,D}^{*\mu} &= w_D^{*\mu} + (1 - \mu_D) m_{b,D}^*, \\ \text{with } m_{b,D}^* &= p_{b,D}^* - w_D^* = \frac{2\tau(\gamma_D + \tau\gamma_S(1 - \tau))}{\gamma_D(4\gamma_D + \gamma_S(3\tau + 1)(1 - \tau))}, \end{aligned} \quad (\text{D.77})$$

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

$$\begin{aligned}
p_{\beta,D}^{*\mu} &= w_S^{*\mu} + (1 - \mu_D) m_{\beta,D}^*, \\
\text{with } m_{\beta,D}^* &= p_{\beta,D}^* - w_S^* = \frac{\tau \gamma_S (\tau - 1)^2}{\gamma_D (4\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))},
\end{aligned} \tag{D.78}$$

$$\begin{aligned}
\text{and } p_{b,S}^{*\mu} &= w_S^{*\mu} + (1 - \mu_S) m_{b,S}^*, \\
\text{with } m_{b,S}^* &= p_{b,S}^* - w_S^* = \frac{4\gamma_D - \gamma_S (1 - 3\tau) (1 - \tau)}{2\gamma_S (4\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))}.
\end{aligned} \tag{D.79}$$

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

D.3.1 Maximum Wholesale Margin Regulation in the Destination Country

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

The manufacturer's profit is given as:

$$\begin{aligned}
\pi_M^{*\mu} &= \underbrace{w_D^{*\mu} \left(1 - \frac{\gamma_D (p_{b,D}^{*\mu} - p_{\beta,D}^{*\mu})}{\tau} \right)}_{\pi_{w_b,D}^{*\mu}} + \underbrace{w_S^{*\mu} \left(1 - \gamma_S p_{b,S}^{*\mu} \right)}_{\pi_{w_b,S}^{*\mu}} \\
&\quad + \underbrace{w_S^{*\mu} \left(\frac{\gamma_D (p_{b,D}^{*\mu} - p_{\beta,D}^{*\mu})}{\tau} - \frac{\gamma_D p_{\beta,D}^{*\mu}}{(1 - \tau)} \right)}_{\pi_{w_\beta}^{*\mu}} + \phi_D^{*\mu} + \phi_S^{*\mu},
\end{aligned} \tag{D.80}$$

with fixed fees given as

$$\phi_D^{*\mu} = 0 \text{ and } \phi_S^{*\mu} = \frac{(1 - w_S^{*\mu} \gamma_S)^2}{4\gamma_S}. \tag{D.81}$$

Substituting (D.81), and equilibrium prices into (D.80) and maximizing with respect to $w_D^{*\mu}$ and $w_S^{*\mu}$ gives the wholesale prices $w_D^{*\mu} = \frac{(\tau + 2\gamma_D w_S^{*\mu})}{2\gamma_D}$ and $w_S^{*\mu} = \frac{4\gamma_D w_D^{*\mu} (1 - \tau)}{4\gamma_D + \tau \gamma_S (1 - \tau)}$. Equilibrium

wholesale prices are given as

$$w_D^{*\mu} = \frac{(4\gamma_D + \tau\gamma_S(1-\tau))}{2\gamma_D(4\gamma_D + \gamma_S(1-\tau))} \text{ and } w_S^{*\mu} = \frac{2(1-\tau)}{(4\gamma_D + \gamma_S(1-\tau))}. \quad (\text{D.82})$$

Both wholesale prices are higher than under free pricing:

$$\begin{aligned} w_D^{*\mu} - w_D^* &= \frac{\tau \left(16\gamma_D^2 + 4\gamma_S\gamma_D(4\tau+1)(1-\tau) + \gamma_S^2(7\tau-3)(1-\tau)^2 \right)}{2\gamma_D(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0, \\ w_S^{*\mu} - w_S^* &= \frac{6\tau\gamma_S(1-\tau)^2}{(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0. \end{aligned} \quad (\text{D.83})$$

Equilibrium drug prices are:

$$\begin{aligned} p_{b,D}^{*\mu} &= \frac{(4\gamma_D + \tau\gamma_S(1-\tau))}{2\gamma_D(4\gamma_D + \gamma_S(1-\tau))}, \quad p_{\beta,D}^{*\mu} = \frac{2(1-\tau)}{(4\gamma_D + \gamma_S(1-\tau))} \\ \text{and } p_{b,S}^{*\mu} &= \frac{4\gamma_D + 3\gamma_S(1-\tau)}{2\gamma_S(4\gamma_D + \gamma_S(1-\tau))}. \end{aligned} \quad (\text{D.84})$$

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

$$\begin{aligned} p_{b,D}^{*\mu} - p_{b,D}^* &= -\frac{3\tau\gamma_S^2(1-\tau)^3}{2\gamma_D(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0 \\ p_{\beta,D}^{*\mu} - p_{\beta,D}^* &= \frac{(2\gamma_D - \gamma_S(1-\tau))\tau\gamma_S(1-\tau)^2}{\gamma_D(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \leq 0, \\ p_{b,S}^{*\mu} - p_{b,S}^* &= \frac{3\tau\gamma_S(1-\tau)^2}{(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0. \end{aligned} \quad (\text{D.85})$$

Equilibrium quantities are:

$$\begin{aligned}
q_{b,D}^{*\mu} &= \frac{1}{2}, \\
q_{\beta,D}^{*\mu} &= \frac{\gamma_S(1-\tau)}{2(4\gamma_D + \gamma_S(1-\tau))} \text{ and} \\
q_{b,S}^{*\mu} &= \frac{(4\gamma_D - \gamma_S(1-\tau))}{2(4\gamma_D + \gamma_S(1-\tau))}.
\end{aligned} \tag{D.86}$$

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

$$\begin{aligned}
q_{b,D}^{*\mu} - q_{b,D}^* &= \frac{\gamma_S(\tau-1)^2}{2(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0 \\
q_{\beta,D}^{*\mu} - q_{\beta,D}^* &= -\frac{\gamma_S(1-\tau)(4\gamma_D + \gamma_S(1-3\tau)(1-\tau))}{2(4\gamma_D + \gamma_S(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0 \\
q_{b,S}^{*\mu} - q_{b,S}^* &= -\frac{3\tau\gamma_S^2(1-\tau)^2}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(1-\tau))} < 0.
\end{aligned} \tag{D.87}$$

D.3.2 Maximum Wholesale Margin Regulation in the Source Country

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

The manufacturer's profit is given as:

$$\begin{aligned}
\pi_M^{*\mu} &= \underbrace{w_D^{*\mu} \left(1 - \frac{\gamma_D(p_{b,D}^{*\mu} - p_{\beta,D}^{*\mu})}{\tau} \right)}_{\pi_{w_b,D}^{*\mu}} + \underbrace{w_S^{*\mu} \left(1 - \gamma_S p_{b,S}^{*\mu} \right)}_{\pi_{w_b,S}^{*\mu}} \\
&\quad + \underbrace{w_S^{*\mu} \left(\frac{\gamma_D(p_{b,D}^{*\mu} - p_{\beta,D}^{*\mu})}{\tau} - \frac{\gamma_D p_{\beta,D}^{*\mu}}{(1-\tau)} \right)}_{\pi_{w_\beta}^{*\mu}} + \phi_D^{*\mu} + \phi_S^{*\mu},
\end{aligned} \tag{D.88}$$

with fixed fees given as

$$\begin{aligned}\phi_D^{*\mu} &= \frac{(2\tau + \gamma_D w_S^{*\mu} - \gamma_D w_D^{*\mu} (\tau + 1))^2}{\tau \gamma_D (\tau + 3)^2} \text{ and} \\ \phi_S^{*\mu} &= \frac{(\gamma_D w_D^{*\mu} (1 - \tau) + (1 - \tau) \tau - \gamma_D w_S^{*\mu} (1 + \tau))^2}{\tau \gamma_D (1 - \tau) (\tau + 3)^2}.\end{aligned}\quad (\text{D.89})$$

Substituting (D.89), and equilibrium prices into (D.88) and maximizing with respect to $w_D^{*\mu}$ and $w_S^{*\mu}$ gives the wholesale prices $w_D^{*\mu} = \frac{(1-\tau)(2\tau+\gamma_D w_S^{*\mu})}{\gamma_D(3\tau+1)}$ and $w_S^{*\mu} = \frac{(1-\tau)(\tau(14+5\tau+\tau^2)+2\gamma_D w_D^{*\mu}(1-\tau))}{2(\gamma_D(3\tau+1)+\tau\gamma_S(1-\tau)(\tau+3)^2)}$.

Equilibrium wholesale prices are given as

$$w_D^{*\mu} = \frac{(1 - \tau) (\gamma_D (2 - \tau) + 4\tau \gamma_S (1 - \tau))}{2\gamma_D (\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))} \text{ and } w_S^{*\mu} = \frac{(3\tau + 2) (1 - \tau)}{2(\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))}. \quad (\text{D.90})$$

Both wholesale prices are higher than under no regulation:

$$\begin{aligned}w_D^{*\mu} - w_D^* &= \frac{(1 - \tau)^2 (4\gamma_D - \gamma_S (2 - 3\tau) (1 - \tau))}{2(4\gamma_D + \gamma_S (3\tau + 1) (1 - \tau)) (\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))} > 0, \\ w_S^{*\mu} - w_S^* &= \frac{(3\tau + 1) (1 - \tau) (4\gamma_D - \gamma_S (2 - 3\tau) (1 - \tau))}{2(4\gamma_D + \gamma_S (3\tau + 1) (1 - \tau)) (\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))} > 0.\end{aligned}\quad (\text{D.91})$$

Equilibrium drug prices are:

$$\begin{aligned}p_{b,D}^{*\mu} &= \frac{(2 - \tau) \gamma_D + 4\tau \gamma_S (1 - \tau)}{2\gamma_D (\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))}, p_{\beta,D}^{*\mu} = \frac{(1 - \tau) (\gamma_D (2 + \tau) + 2\tau \gamma_S (1 - \tau))}{2\gamma_D (\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))} \\ \text{and } p_{b,S}^{*\mu} &= \frac{(3\tau + 2) (1 - \tau)}{2(\gamma_D + \gamma_S (3\tau + 1) (1 - \tau))}.\end{aligned}\quad (\text{D.92})$$

In the destination country, prices are higher than under free pricing. In the source country, the

price is lower than under free pricing:

$$\begin{aligned}
p_{b,D}^{*\mu} - p_{b,D}^* &= \frac{(1-\tau)(4\gamma_D - \gamma_S(2-3\tau)(1-\tau))}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0, \\
p_{\beta,D}^{*\mu} - p_{\beta,D}^* &= \frac{(\tau+1)(1-\tau)(4\gamma_D - \gamma_S(2-3\tau)(1-\tau))}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0, \\
p_{b,S}^{*\mu} - p_{b,S}^* &= -\frac{4\gamma_D^2 - \gamma_S\gamma_D(1-3\tau)(1-\tau) + \gamma_S^2(3\tau+1)(1-\tau)^2}{2\gamma_S(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0 \quad (\text{D.93})
\end{aligned}$$

Equilibrium quantities are:

$$\begin{aligned}
q_{b,D}^{*\mu} &= \frac{\gamma_D(2-\tau) + 4\tau\gamma_S(1-\tau)}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))}, \quad q_{\beta,D}^{*\mu} = \frac{\gamma_S(1-\tau) - \gamma_D}{(\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \\
\text{and } q_{b,S}^{*\mu} &= \frac{2\gamma_D + 3\tau\gamma_S(1-\tau)}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (\text{D.94})
\end{aligned}$$

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

$$\begin{aligned}
q_{b,D}^{*\mu} - q_{b,D}^* &= \frac{\gamma_D(1-\tau)(4\gamma_D - \gamma_S(2-3\tau)(1-\tau))}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \\
q_{\beta,D}^{*\mu} - q_{\beta,D}^* &= -\frac{\gamma_D(4\gamma_D - \gamma_S(2-3\tau)(1-\tau))}{(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} \\
q_{b,S}^{*\mu} - q_{b,S}^* &= \frac{4\gamma_D^2 - \gamma_S\gamma_D(1-3\tau)(1-\tau) + \gamma_S^2(3\tau+1)(1-\tau)^2}{2(\gamma_D + \gamma_S(3\tau+1)(1-\tau))(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))}. \quad (\text{D.95})
\end{aligned}$$

D.4 Mandatory Rebates

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

$$w_D^{*\psi} = (1 - \psi_D) w_D^* \quad (\text{D.96})$$

and in country S to:

$$w_S^{*\psi} = (1 - \psi_S) w_S^*. \quad (\text{D.97})$$

D.4.1 Mandatory Rebates in the Destination Country

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

The manufacturer's profit is given as:

$$\pi_M^{*\psi} = \underbrace{w_S^{*\psi} (1 - \gamma_S p_{b,S}^{*\psi})}_{\pi_{w_b,S}^{*\psi}} + \underbrace{w_S^{*\psi} \left(\frac{\gamma_D (p_{b,D}^{*\psi} - p_{\beta,D}^{*\psi})}{\tau} - \frac{\gamma_D p_{\beta,D}^{*\psi}}{(1 - \tau)} \right)}_{\pi_{w_\beta}^{*\psi}} + \phi_D^{*\psi} + \phi_S^{*\psi}. \quad (\text{D.98})$$

with the fixed fees given as

$$\phi_D^{*\psi} = \frac{(2\tau + \gamma_D w_S^{*\psi})^2}{\tau \gamma_D (\tau + 3)^2}, \quad \phi_S^{*\psi} = \frac{(1 - w_S^{*\psi} \gamma_S)^2}{4\gamma_S}. \quad (\text{D.99})$$

Substituting (D.99), and equilibrium prices into (D.98) and maximizing with respect to $w_S^{*\psi}$ gives the wholesale price:

$$w_S^{*\psi} = \frac{2\tau(1 - \tau)(5 - \tau)}{(4\gamma_D(3\tau + 1) + \tau\gamma_S(1 - \tau)(\tau + 3)^2)}. \quad (\text{D.100})$$

The wholesale price is lower than under free pricing:

$$w_S^{*\psi} - w_S^* = -\frac{8(1 - \tau)^3(\gamma_D + \tau\gamma_S(1 - \tau))}{(4\gamma_D(3\tau + 1) + \tau\gamma_S(1 - \tau)(\tau + 3)^2)(4\gamma_D + \gamma_S(3\tau + 1)(1 - \tau))} < 0. \quad (\text{D.101})$$

Equilibrium drug prices are given as:

$$\begin{aligned}
p_{b,D}^{*\psi} &= \frac{2\tau(\tau+3)(\gamma_D + \tau\gamma_S(1-\tau))}{\gamma_D(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}, \\
p_{\beta,D}^{*\psi} &= \frac{\tau(1-\tau)(8\gamma_D + \tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}, \\
p_{b,S}^{*\psi} &= \frac{(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(4\tau + \tau^2 + 19))}{2\gamma_S(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}. \tag{D.102}
\end{aligned}$$

All prices are lower than under free pricing:

$$\begin{aligned}
& p_{b,D}^{*\psi} - p_{b,D}^* \\
&= -\frac{4(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))(2\gamma_D(\tau+1) + \tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D[4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2][4\gamma_D + \gamma_S(3\tau+1)(1-\tau)]} < 0, \\
& p_{\beta,D}^{*\psi} - p_{\beta,D}^* \\
&= -\frac{2(1-\tau)^2(\gamma_D + \tau\gamma_S(1-\tau))(4\gamma_D + \tau\gamma_S(\tau+3)(1-\tau))}{\gamma_D[4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2][4\gamma_D + \gamma_S(3\tau+1)(1-\tau)]} < 0, \\
& p_{b,S}^{*\psi} - p_{b,S}^* \\
&= -\frac{4(1-\tau)^3(\gamma_D + \tau\gamma_S(1-\tau))}{[4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2][4\gamma_D + \gamma_S(3\tau+1)(1-\tau)]} < 0. \tag{D.103}
\end{aligned}$$

Equilibrium quantities are

$$\begin{aligned}
q_{b,D}^{*\psi} &= \frac{2(\tau+3)(\gamma_D + \tau\gamma_S(1-\tau))}{(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}, \\
q_{\beta,D}^{*\psi} &= \frac{(1-\tau)(\tau\gamma_S(\tau+3) - 2\gamma_D)}{(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}, \\
q_{b,S}^{*\psi} &= \frac{4\gamma_D(3\tau+1) - \tau\gamma_S(1-\tau)(1-8\tau-\tau^2)}{2(4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2)}. \tag{D.104}
\end{aligned}$$

In the destination country, mandatory rebates shift demand from the parallel import to the

locally sourced version. In the source country, the quantity is higher than under free pricing:

$$\begin{aligned}
& q_{b,D}^{*\psi} - q_{b,D}^* \\
= & \frac{2(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))(8\gamma_D + \gamma_S(\tau+3)(1-\tau)(\tau+1))}{[4\gamma_D + \gamma_S(3\tau+1)(1-\tau)] [4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2]} > 0, \\
& q_{\beta,D}^{*\psi} - q_{\beta,D}^* \\
= & -\frac{2(1-\tau)(\gamma_D + \tau\gamma_S(1-\tau))(4\gamma_D + \gamma_S(\tau+3)(1-\tau))}{[4\gamma_D + \gamma_S(3\tau+1)(1-\tau)] [4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2]} < 0 \\
& q_{b,S}^{*\psi} - q_{b,S}^* \\
= & \frac{4\gamma_S(1-\tau)^3(\gamma_D + \tau\gamma_S(1-\tau))}{[4\gamma_D + \gamma_S(3\tau+1)(1-\tau)] [4\gamma_D(3\tau+1) + \tau\gamma_S(1-\tau)(\tau+3)^2]} > 0. \quad (\text{D.105})
\end{aligned}$$

D.4.2 Mandatory Rebates in the Source Country

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

The manufacturer's profit is given as:

$$\pi_M^{*\psi} = w_D^{*\psi} \left(1 - \frac{\gamma_D(p_{b,D}^{*\psi} - p_{\beta,D}^{*\psi})}{\tau} \right) + \phi_D^{*\psi} + \phi_S^{*\psi}, \quad (\text{D.106})$$

with fixed fees given as

$$\begin{aligned}
\phi_D^{*\psi} &= \frac{\left(2\tau + \gamma_D w_S^{*\psi} - \gamma_D w_D^{*\psi} (\tau + 1) \right)^2}{\tau \gamma_D (\tau + 3)^2}, \\
\phi_S^{*\psi} &= \frac{\left(\gamma_D w_D^{*\psi} (1 - \tau) + (1 - \tau) \tau - \gamma_D w_S^{*\psi} (1 + \tau) \right)^2}{\tau \gamma_D (1 - \tau) (\tau + 3)^2}. \quad (\text{D.107})
\end{aligned}$$

Substituting (D.107), and equilibrium prices into (D.106) and maximizing with respect to

$w_D^{*\psi}$ gives the wholesale price:

$$w_D^{*\psi} = 2\tau \frac{(1-\tau)}{\gamma_D (3\tau+1)}. \quad (\text{D.108})$$

The wholesale price is lower than under free pricing:

$$w_D^{*\psi} - w_D^* = -\frac{2(\tau-1)^2}{(3\tau+1)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0. \quad (\text{D.109})$$

Equilibrium drug prices are given as

$$\begin{aligned} p_{b,D}^{*\psi} &= \frac{2\tau}{\gamma_D (3\tau+1)}, \\ p_{\beta,D}^{*\psi} &= \frac{(1-\tau)\tau}{\gamma_D (3\tau+1)}, \\ \text{and } p_{b,S}^{*\psi} &= \frac{1}{2\gamma_S}. \end{aligned} \quad (\text{D.110})$$

All prices are lower than under free pricing:

$$\begin{aligned} p_{b,D}^{*\psi} - p_{b,D}^* &= -\frac{2(1-\tau)}{(3\tau+1)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0, \\ p_{\beta,D}^{*\psi} - p_{\beta,D}^* &= -\frac{2(1-\tau)(\tau+1)}{(3\tau+1)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0, \\ p_{b,S}^{*\psi} - p_{b,S}^* &= -\frac{(1-\tau)}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0. \end{aligned} \quad (\text{D.111})$$

Equilibrium quantities are:

$$q_{b,D}^{*\psi} = 2\frac{\tau}{3\tau+1}, \quad q_{\beta,D}^{*\psi} = \frac{1}{3\tau+1}, \quad q_{b,S}^{*\psi} = \frac{1}{2}. \quad (\text{D.112})$$

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

$$\begin{aligned}
q_{b,D}^{*\psi} - q_{b,D}^* &= -2\gamma_D \frac{(1-\tau)}{(3\tau+1)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} < 0, \\
q_{\beta,D}^{*\psi} - q_{\beta,D}^* &= 4 \frac{\gamma_D}{(3\tau+1)(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0, \\
q_{b,S}^{*\psi} - q_{b,S}^* &= \gamma_S \frac{(1-\tau)}{(4\gamma_D + \gamma_S(3\tau+1)(1-\tau))} > 0.
\end{aligned} \tag{D.113}$$

Appendix E

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides Statt, dass ich meine Doktorarbeit „Parallel Trade of Pharmaceuticals. Conflicts in Health Policy Objectives and Regulatory Externalities in the EU Internal Market.“ selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe. Die Arbeit hat bisher in gleicher oder ähnlicher Form oder auszugsweise noch keiner Prüfungsbehörde vorgelegen.