ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Birg, Laura

Conference Paper

Pharmaceutical Regulation under Market Integration through Parallel Trade

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Birg, Laura (2022) : Pharmaceutical Regulation under Market Integration through Parallel Trade, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/264128

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Pharmaceutical Regulation under Market Integration through Parallel Trade

Laura Birg*

February 2022

Abstract

I study the effect of parallel trade (cross border resale of goods without the authorization of the manufacturer) on pharmaceutical regulation in a North-South framework. Governments in both countries may restrict prices directly via price caps or limit third-party payer reimbursement for the drug via reimbursement limits.

Parallel trade may relax regulation in the source country of parallel imports (South) and intensify regulation in the destination country (North): In the source country, parallel trade may relax regulation both under a price cap and a reimbursement limit under certain conditions. In the destination country, parallel trade has no effect on the level of regulation under a price cap, and it intensifies regulation under a reimbursement limit.

Parallel trade may change regulatory preferences: Under no parallel trade, both the source and destination country set price caps. Under parallel trade, the source country sets a price cap, but the destination country sets a reimbursement limit, thereby enforcing a higher price cap in the South. This implies higher drug prices under parallel trade in both source and destination country.

JEL classification: F12, I11, I18

Keywords: pharmaceutical regulation, price cap, reimbursement limit, parallel trade

1 Introduction

Parallel trade refers to the cross border resale of goods without the authorization of the manufacturer (Maskus, 2000). This type of arbitrage is a response to international price differences

*Department of Human and Health Sciences, University of Bremen, Grazer Str 2, 28359 Bremen, Germany, lbirg@uni-bremen.de

(Kyle, 2010). In the European Union, parallel trade is a common phenomenon, especially for pharmaceuticals, for which price differences may reach up to 300% (Kanavos & Costa-Font, 2005).

One major source of drug price differences are regulatory differences, both in the level of regulation and the type of regulatory instruments applied.[1] Commonly applied regulatory instruments are price caps and (internal) reference prices. Price caps aim at restricting monopoly pricing and limit the prices (of covered drugs) directly by establishing price ceilings.[2] (Internal) reference prices are intended to increase price sensitivity by limiting reimbursement by third-party payers.[3] Under (internal) reference pricing, price setting is free, but the limitation of reimbursement implies that patients have to pay the difference between drug price and reimbursement limit. In the European Union, almost all member states use or have used price caps and/or reference prices (see Espin & Rovira (2007) or Carone, Schwierz & Xavier (2012) for an overview of regulatory interventions in the European Union).

By creating differences in drug prices, regulatory differences drive the direction and volume of parallel trade.[4] At the same time, parallel trade limits firms' ability to set different prices in different countries. In destination countries of parallel imports, parallel trade may reduce drug prices by providing lower-priced substitutes and enhancing competition in pharmaceutical markets.[5]

Price interdependencies under parallel trade may create the incentive for firms to delay or avoid launches in low-price countries, i.e., potential source countries of parallel imports. This allows firms to (temporarily) retain high prices in other countries (see Danzon, Wang & Wang, 2005; Kyle, 2007).[6] For instance, there is an average launch delay of 10 months in the EU 15,

[1] According to 168 TFEU, health policy is in the national competence of member states. The Price Transparency Directive (Directive 89/195/EC), which provides rules for the control of pharmaceutical prices, is the only existing harmonization measure in the field of pharmaceutical price regulation and reimbursement.

[2] Price caps may be based on prices of the same drug in other countries or prices of therapeutic alternatives.

[3] Typoically, reimbursement limits are set for a group of substitutable drugs. Substitutability may be defined with respect to the active ingredient, therapeutic category or the therapeutic function (Lopez-Casanovas & Puig-Junoy, 2000).

[4] Typical source countries of parallel imports are low price-countries with strict price regulation, e.g., Greece, Italy, Portugal, and Spain, while destination countries are high price-countries with relatively free price setting, e.g., Denmark, Germany, the Netherlands, and Sweden (Kanavos & Costa-Font, 2005). In 2016, pharmaceutical parallel trade had a volume of € 5.2 bn (EFPIA, 2018). In the destination countries, the share of parallel imports in pharmacy market sales ranged between 8.2% in the Netherlands, 8.5% in Germany, 12.9% in Sweden, and 25.5 % in Denmark in 2018 (EFPIA, 2020).

[5] Empirical evidence on the price decreasing effect of parallel trade in destination countries is ambiguous. While some studies (Kanavos et al., 2004, Kyle, Allsbrook & Schulman, 2008) find no evidence for price competition generated by parallel trade or stronger price competition than in countries without parallel trade, others (West & Mahon, 2003, Ganslandt & Maskus, 2004, Granlund & Köksal, 2011, Duso, Herr & Suppliet, 2014) find that parallel trade may generate competitive pressure on drug prices.

[6] Danzon, Wang & Wang (2005), Danzon & Epstein (2012), Kyle (2011) suggest that stricter regulation and/or

ranging from 3.5 months in Germany to 18.9 months in Belgium (Heuer, Mejer & Neuhaus, 2007). This suggests that pharmaceutical firms' export decisions are shaped by parallel trade and regulatory decisions (Bennato & Valletti, 2014).

As firms may delay (or even limit supply) to low-price countries, governments in source countries of parallel imports may refrain from strict regulation if they take into account the impact on the firm's decision to export to their country (Pecorino, 2002; Königbauer, 2004; Grossman & Lai, 2008; Bennato & Valletti, 2014). Pecorino (2002) shows that under parallel trade, a pharmaceutical manufacturer will make fewer concessions in drug price bargaining in a potential source country of parallel imports. Königbauer (2004) argues that parallel imports may discipline national regulators in the European Union, as the manufacturer's threat of not supplying countries with low drug prices prevents free-riding on the pharmaceutical firms' R&D investment by setting low prices. Similarly, Grossman & Lai (2008) show that the pace of innovation may be faster under parallel trade because of the change in regulatory choices. Bennato & Valletti (2014) find that a withdrawal from price regulation in a source country of parallel imports may increase R&D investment.

While the literature on the effect of parallel trade on pharmaceutical regulation has focused on the impact of price caps in source countries of parallel imports, this paper extends the analysis of the effect of parallel trade on pharmaceutical regulation to regulation in both the destination and the source country and to two potential regulatory instruments, price caps and reimbursement limits. This set-up allows differentiating between the effect of parallel trade on regulation in export-oriented destination countries and the effect of parallel trade on regulation in import-oriented source countries. Moreover, the analysis of two common regulatory instruments allows studying the choice of instruments under parallel trade.

By changing regulatory decisions in destination and/or source country of parallel imports, parallel trade may result in regulatory convergence or regulatory divergence. In the former case, parallel trade may also erode price differences between countries, in particular, if price differences are driven by regulatory differences. In this sense, market integration through parallel trade may replace policy harmonization. Given that virtually no harmonization instrument is in place in the European Union, parallel trade may be of particular importance.

The effect of parallel trade on regulation is linked to the policy issue of access to pharmaceu-

parallel trade may result in greater launch delays. Also, approval procedures for new drugs and approval times may contribute to non-launches (Houy & Jelovac, 2019).

ticals. Parallel trade may affect access in the source country of parallel imports if manufacturers decide to limit or avoid selling in low-price countries. Moreover, also through its impact on regulatory decisions parallel trade may affect access – if governments change regulatory decisions in response to parallel trade, this also changes drug prices and with it access to the drug.

In the source country (hereafter South), parallel trade may relax regulation under both regulatory instruments. The manufacturer's threat of not exporting to the South may increase the price cap or the reimbursement limit under certain conditions.[7] In the destination country (North), parallel trade may intensify regulation under a reimbursement limit and has no effect on the strictness of regulation under a price cap. Under a price cap, the government in the North prefers a price cap of zero, maximizing consumer surplus and minimizing third-party payer expenditure. Under a reimbursement limit, the government in the North prefers a high level of reimbursement, as a potential decrease in the reimbursement limit would decrease both consumer surplus and third-party payer expenditure.

Parallel trade may change the choice of regulatory instruments: Under no parallel trade, both the North and the South prefer price caps over reimbursement limits. Welfare in both countries decreases in the drug price. Price caps allow governments to set a drug price of zero, while reimbursement limits result in higher drug prices. Under parallel trade, the North applies a reimbursement limit, but the South sets a price cap. Choosing a reimbursement limit allows the North to enforce a lower level of regulation and thus a higher price in the South. The reimbursement limit in the North allows enforcing a higher drug price in the South than under price caps. The South, as under no parallel trade, prefers a price caps to the reimbursement limit, as it allows attaining a lower drug price. This is, under endogenous health policy choice, parallel trade results in a lower level of regulation and higher drug prices. Compared to the equilibrium without parallel trade, in which both countries set price caps of zero, drug prices under parallel trade are higher. This also worsens access in both the source and destination country.

The paper is organized as follows. The next section introduces the model. Section 3 studies the effect of regulation on the export decision. Section 4 analyzes the effect of parallel trade on the level of regulation for a given instrument. Section 5 studies the effect of parallel trade on the choice of regulatory instruments. Section 6 presents discussions of the model. Section 7

[7] Under a price cap, parallel trade relaxes regulation unless the North also sets a price cap. Under a reimbursement limit, parallel trade relaxes regulation, if countries are sufficiently different in demand.

concludes.

2 Model

Consider two countries $j = N, S$, North and South, which differ in demand.[8] An innovative firm, which sells an on-patent drug is located in the North. The firm may decide not to sell to the South. In both countries, welfare-maximizing governments may restrict prices directly via price caps or limit third-party payer reimbursement for the drug via reimbursement limits.

2.1 Firm

In the North, an innovative firm F is located, which sells an on-patent drug. The firm always sells in the North (if the price is non-negative), but the firm may decide whether or not to sell the South based on the price in the South and the potential impact on the price in the North under parallel trade. Patent protection provides the firm with a monopoly in both countries. The firm produces at constant marginal cost, which is normalized to zero.

Two pricing regimes are possible depending on the (non) occurence of parallel trade: If parallel trade does not take place, the firm may price-discriminate between the North and the South and set country-specific drug prices p_j. If parallel trade takes place, the firm sets a uniform price p_{NS}. Therefore, parallel trade (as costless arbitrage) enforces a uniform drug price in both countries (as in Pecorino, 2002; Valletti, 2006; Roy & Saggi, 2012; Bennato & Valletti, 2014).

Under parallel trade, the firm may avoid selling at a uniform price in both countries by not exporting to the South. Then the firm foregoes sales in the South but may set a country-specific price in the North. This is, under parallel trade the firm faces the trade-off between selling at a uniform price in both countries or selling at a country-specific price in the North and foregoing sales in the South.

2.2 Consumers

In both countries, there is a unit mass of consumers differing in willingness to pay for the drug. Each consumer demands either one or zero units of the drug. This implies that the market is not covered and changes in the drug price result in quantity changes. The utility derived from

[8] Countries differ in both the maximum willingness to pay and price elasticity.

no drug consumption is zero. A consumer i in country j who buys one unit of drug obtains a net utility of

$$U(\theta_{ij}, c_j) = \theta_{ij} - c_j, \tag{1}$$

where θ_{ij} is the gross valuation and c_j the country-specific drug copayment. The (preference) parameter θ can be interpreted as willingness to pay. Assume that θ is uniformly distributed over the interval $[0, \mu_j]$ in country $j = N, S$, with $\mu_N = \mu \geq \mu_S = 1$. Heterogeneity among consumers in θ may stem from, for instance, differences in the severity of the condition, prescription practices or insurance coverage (see e.g., Brekke, Holmas & Straume, 2011). The parameter μ_j can be interpreted as the maximum willingness to pay. In the following, μ will be referred to as demand parameter.

The marginal consumer in country j who is indifferent between buying the drug or not has a gross valuation $\theta_{ij}^* = c_j$. Hence, demand in country N is given as $q_N = \frac{1}{\mu}(\mu - c_N)$, and demand in country S is given as $q_S = 1 - c_S$.

2.3 Governments

In both countries, third-party payers (health insurance, health insurance programs or social insurance, etc.) cover drug costs partially. The copayment is price dependent: Consider that consumers pay a fraction γ_j, $\gamma_j \in (0, 1)$, out-of-pocket (coinsurance)[9]. Under coinsurance, the drug copayment (and thus, the relevant price for consumers) is $c_j = \gamma_j p_j$, the reimbursement by third-party payers is $\left(1 - \gamma_j\right) p_j$.

In both countries, governments may restrict prices directly via price caps or limit third-party payer reimbursement for the drug via reimbursement limits. Under price caps, governments restrict drug prices to maximum prices P_j. Under reimbursement limits, governments restrict reimbursement by third-party payers to reimbursement limits r_j. Firms are free to set prices, but third-party payers reimburse the drug according to the reimbursement limit, which changes consumer copayments to $c_j^R = \gamma_j r_j + p_j - r_j$ and reimbursement to $\left(1 - \gamma_j\right) r_j$.

Governments maximize domestic welfare, which is defined as the (unweighted) sum of consumer surplus and the firm's profit net of third-party payer reimbursement in the North and

[9] Typically, coinsurance rates are not subject to frequent changes. Coinsurance rates are set to balance the protection from risks of financially losses due to sickness and (prevention of) moral hazard in utilizing health care products and services. In Germany, e.g., the coinsurance rate has not been changed since 2004.

consumer surplus net of third-party payer reimbursement in the South:

$$\begin{aligned} W_N &= CS_N + \pi - E_N, \\ W_S &= CS_S - E_S. \end{aligned} \tag{2}$$

Consumer surplus in country j is given as $CS_j = \frac{1}{\mu_j} \int_{\theta_{ij}^*}^{\mu_j} (\theta_{ij} - c_j)\, d\theta$, the firm's profit is given as $\pi = q_N p_N + q_S p_S$, third-party payer reimbursement in country j is defined as $E_j = (1-\gamma_j) p_j q_j$ under no regulation and price caps[10] and $E_j = (1-\gamma_j) r_j q_j$ under reimbursement limits.[11]

2.4 Structure of the Model

In this set-up, there are two differences between countries: First, countries N and S differ in demand and demand elasticity (due to differences in maximum willingness to pay and coinsurance rates). Differences in μ_j and/or γ_j generate differences in drug prices and trigger parallel trade. Second, as the North is export-oriented, while the South is import-oriented, and accordingly welfare, government objectives, and regulatory decisions differ between countries.

Consider the following timing: In stage 0, governments in the North and the South choose the regulatory instrument**,** i.e., price caps or reimbursement limits. In stage 1, the firm decides whether or not to sell to the South. In stage 3, governments set price caps or reimbursement limits. In the latter case, the firm sets prices in stage 4.

3 The Effect of Parallel Trade on the Export Decision

In this section, the effect of parallel trade on the firm's export decision is analyzed in a scenario without regulation to illustrate how the export decision is the basis for the firm's threat of non-supply.

Equilibrium prices, quantities, and the firm's profit can be found in Appendix A.1. The superscript x denotes variables under no parallel trade.

[10] Under no regulation and price caps, the third-party payer reimburses the drug based on the drug price.
[11] Under reimbursement limits, the third-party payer reimburses the drug based on the reimbursement limit.

3.1 No Parallel Trade

When parallel trade is not allowed, the firm can price discriminate. The firm maximizes

$$\pi^x_{N,S} = q_N\left(p^x_N\right)p^x_N + q_S\left(p^x_S\right)p^x_S. \tag{3}$$

The firm price sets country-specific prices $p^x_N = \frac{\mu}{2\gamma_N}$ and $p^x_S = \frac{1}{2\gamma_S}$.

Equilibrium drug prices decrease in coinsurance rates γ_j. Also, the drug price in the North increases in the maximum willingness to pay μ.

Differences in drug prices stem from differences in coinsurance rates and/or differences in demand. In particular, the price in the North is higher than the price in the South, if demand in the North is sufficiently high, $\mu > \widetilde{\mu^x} = \frac{\gamma_N}{\gamma_S}$, and/or if the coinsurance rate (and accordingly price elasticity) in the North is sufficiently low relative to that in the South, $\gamma_N < \widetilde{\gamma^x_N} = \mu\gamma_S$.[12] This implies that both differences in demand and differences in coinsurance rates may be drivers of parallel trade.[13] Parallel trade is profitable if $p^x_N > p^x_S$. In what follows, assume $\gamma_N < \widetilde{\gamma^x_N} = \mu\gamma_S$, i.e., parallel trade takes place.

3.2 Parallel Trade

If parallel trade is allowed, the possibility to price-discriminate is undermined, and the firm sells at a uniform price p_{NS} in both countries. The firm's profit is

$$\pi_{N,S} = \left(q_N\left(p_{NS}\right) + q_S\left(p_{NS}\right)\right)p_{NS}, \tag{4}$$

which is maximized for the price $p_{NS} = \frac{\mu}{\gamma_N + \mu\gamma_S}$.

Parallel trade decreases the price in the North and increases the price in the South ($p^x_N > p_{NS} > p^x_S$).[14] Higher demand μ aggravates the price changes generated by parallel trade ($\frac{\partial(p_{NS}-p_N)}{\partial\mu} > 0$, $\frac{\partial(p_{NS}-p_S)}{\partial\mu} > 0$).

Under parallel trade, the firm may decide not to sell to the South to avoid selling at a uniform price. The firm sells to both countries if $\Delta = \pi_{N,S} - \pi^x_N \geq 0$, i.e., the profit from

[12] If demand in both countries is identical ($\mu = 1$), $p^x_N > p^x_S$ if $\gamma_N < \gamma_S$. If coinsurance rates in both countries are identical ($\gamma_N = \gamma_S = \gamma$), $p^x_N > p^x_S$ if $\mu > 1$.

[13] Jelovac & Bordoy (2005) study the welfare consequences of parallel trade if parallel trade is driven by differences in coinsurance rates ("health systems") or differences in willingness to pay ("health needs"). They show that when countries differ in health systems, parallel trade decreases welfare; when countries differ in health needs, parallel trade increases welfare. Note that Jelovac & Bordoy (2005) consider these differences between countries separately.

[14] This also implies that the quantity in the North increases, while the quantity in the South decreases.

selling at a uniform price p_{NS} to both countries is at least as high as the profit from selling at a country-specific price p_N^x only to the North. The firm trades off accepting a lower price in the North and selling in both countries and setting a country-specific price for the North and foregoing sales in the South.

Parallel trade only occurs within a limited range of parameters. Two conditions have to be met: The first condition is that a sufficiently high price difference between N and S makes parallel trade profitable. The second condition is that the drug has to be sold to the South in the first place. This is, by not supplying the South, the firm can deter parallel trade.

3.3 Welfare

Welfare in the North is given as $W_N^x = CS_N^x + \pi - E_N^x$, welfare in the South is given as $W_S^x = CS_S^x - E_S^x$.

In the North, parallel trade decreases the drug price. The lower drug price increases consumer surplus and decreases third-party payer expenditure. The firm's profit decreases, as parallel trade generates distortions from the profit-maximizing prices in both countries, i.e., both the profit from sales to the North and the profit from sales to the South are lower under parallel trade. If demand in the North is sufficiently low, and prices in both countries under no parallel trade are rather similar, the latter effect is less strong. Therefore, parallel trade increases welfare in the North if demand is sufficiently low, i.e.$W_N > W_N^x$ if $\mu < \widetilde{\mu_{W_N}}$.

In the South, parallel trade increases the drug price. The higher drug price decreases consumer surplus but decreases third-party payer expenditure (the quantity effect is stronger than the price effect). Higher demand in the North increases the uniform price under parallel trade, and the latter effect is stronger. Thus, parallel trade increases welfare in the South if demand in the North is sufficiently high, i.e., $W_S > W_S^x$, if $\mu > \widetilde{\mu_{W_S}}$. Note that parallel trade increases welfare in both countries if copayments in the South are sufficiently low relative to copayments in the North.

4 Choice of Regulation Level - Second-stage Outcome

In this section, I study the effect of parallel trade on the level of regulation for a given regulatory instrument. The regulatory instrument is treated as exogenous in this section but will be endogenized in section 5. I denote no regulation by the superscript $\varnothing$, a price cap by the

superscript P, a reimbursement limit by the superscript R, and policy choice by $\Psi = \{\varnothing, P, R\}$. Superscripts Ψ_N, Ψ_S denote equilibrium policy choices in the North and South. There are two symmetric equilibria (PP, RR) and two asymmetric equilibria (PR, RP). In addition, there are two equilibria, in which only the South limit prices or reimbursement ($\varnothing R$, $\varnothing P$). In what follows, I will discuss the symmetric equilibria as well as the equilibria in which only the South regulates first, followed by the asymmetric equilibria.

Under no parallel trade, policy choices in both countries are independent, under parallel trade, policy choices are linked via the uniform drug price.

4.1 Price Caps in Both Countries (PP) or the South only ($\varnothing P$)

Consider first a scenario, in which both governments or only the government in the South set price caps.

Price caps define maximum prices that can be charged. In what follows, consider binding price caps in both countries, i.e. $P_N^{x,(.)} < p_N^x$, $P_S^{x(.)} < p_S^x$ under no parallel trade and $P_N^{(.)}$, $P_S^{(.)} < p_{NS}$ under parallel trade.

Under price caps, the firm first decides whether to sell to the South; then governments set price caps.[15] Price caps and welfare can be found in Appendix A.2.

Parallel trade may create spillovers in the regulatory decision: Under price caps, the price cap in the South becomes the global price cap if it is lower than the price or even the price cap in the North. While under no parallel trade, the domestic government in the North is in charge of the setting the price cap, under parallel trade, the pricing decision may be shifted to the government in the South. This depends on which government sets the lower price cap.

Moreover, parallel trade may distort regulatory decisions. In the North, the government may not be able to set a high price cap, if the price or price cap in the South is lower. In the South, the government cannot set a very low price cap if the firm's decision to export is endogenous. The firm's threat not to supply the South under parallel trade may force the government to set a higher price cap.

4.1.1 No Parallel Trade ($x, \varnothing P$, x, PP)

Consider first that parallel trade is not allowed. As drug prices and policy choices are independent, the equilibria in which i) only the government in the South ($x, \varnothing P$) sets a price cap

[15] There is no stage in which the firm sets prices as price caps delegate price setting to governments.

$P_S^{x,\varnothing P}$ and in which ii) both governments in the North and in the South set price caps $P_N^{x,P}$, $P_S^{x,P}$ (x, PP), are discussed jointly.

The firm's profit is given as

$$\pi_{N,S}^{x,(.)P} = \begin{cases} q_N\left(p_N^{x,\varnothing P}\right) p_N^{x,\varnothing P} + q_S\left(P_S^{x,\varnothing P}\right) P_S^{x,\varnothing P} & \text{for} \quad \varnothing P \\ q_N\left(P_N^{x,PP}\right) P_N^{x,PP} + q_S\left(P_S^{x,PP}\right) P_S^{x,PP} & \text{for} \quad PP. \end{cases} \tag{5}$$

For equilibrium $\varnothing P$, the firm may set the drug price freely in the North, while in the South the government sets a price cap $P_S^{x,\varnothing P}$. The firm sets the drug price $p_N^{x,\varnothing P} = \frac{\mu}{2\gamma_N}$ in the North. In the South, welfare decreases in the price cap, because a lower drug price increases consumer surplus and decreases third-party payer expenditure. The government in the South sets $P_S^{x,\varnothing P*} = 0$.

For equilibrium PP, governments in the North and the South set price caps $P_N^{x,PP}$, $P_S^{x,PP}$. In both countries, welfare decreases in the price cap. In the North, a lower price cap increases demand, increasing consumer surplus and decreasing third-party payer expenditure. At the same time, lower price caps in both countries decrease the firm's profit. The former effect dominates the latter and welfare decreases in the drug price. In the South, welfare decreases in the price cap.

The government in the North sets $P_N^{x,PP*} = 0$, the government in the South sets $P_S^{x,PP*} = 0$. Price caps of zero imply that neither consumers nor health insurance have to pay for the drug and the market is covered.

Lemma 1 summarizes regulatory preferences in North and South for price caps.

Lemma 1 *Under price caps, both North and South prefer a high level of regulation, i.e., price caps of zero.*

4.1.2 Parallel Trade and Price Cap in the South ($\varnothing P$)

Consider that parallel trade is allowed and only the South sets a price cap P_S^P. The firm's profit is given as

$$\pi_{N,S}^{\varnothing P} = \left(q_N\left(P_S^{\varnothing P}\right) + q_S\left(P_S^{\varnothing P}\right)\right) P_S^{\varnothing P}. \tag{6}$$

In this case, the price in the North is neither determined by the firm nor the government in the North but is given by the price cap in the South: The government in the South sets the global price.

Under parallel trade, the firm may avoid selling at the price cap set by the government in the South by not selling to the South. The firm sells to the South under parallel trade if the profit from selling at the price cap P_S^P in both markets is at least as high than selling at the (free) price p_N^x in the North only, i.e., $\Delta^{\varnothing P} = \pi_{N,S}^{\varnothing P} - \pi_N^x \geq 0$. A (strict) price cap aggravates the trade-off associated with the export decision: The lower the price cap, the higher is the difference in profits for the firm's two strategies and the less likely it is that the firm sells to the South. Note that for a price cap of zero (as it is under no parallel trade), the firm abstains from exporting to the South. Only if $P_S^{\varnothing P} > \widehat{P_S^{\varnothing P}}$, i.e., the price cap is sufficiently high, the firm sells to the South.

Welfare in the South decreases in the price cap $P_S^{\varnothing P}$, as a lower price cap increases consumer surplus and decreases third-party payer expenditure. The government in the South maximizes W_S^P subject to $\Delta^{\varnothing P} = \pi_{N,S}^{\varnothing P} - \pi_N^{x,(.)} \geq 0$, i.e., the firm selling to the South. The government in the South would prefer a price cap of zero. But given that the firm would not export to the South at a very low price cap, the government in South sets $P_S^{\varnothing P*} \geq 0$, which is the lowest price cap compatible with the firm selling in the South. Thus, parallel trade relaxes regulation by preventing the government in the South from setting a very low price cap.

These results suggest, in line with empirical evidence, that strict regulation increases the threat of the firm not supplying the South under parallel trade (Danzon & Epstein, 2008; Verniers, Stremerscha & Croux, 2011; Costa-Font, McGuire & Varol, 2014). Source countries face a trade-off under parallel trade: accept high prices and benefit from (safe) drug supply or regulate prices and face the risk of not being supplied. The endogenous export decision of the firm stemming from the trade-off described above results in a credible threat generated by price spillovers under parallel trade. As a result, source countries of parallel imports may abstain from very strict regulation under parallel trade (Bennato & Valletti, 2014; Grossman & Lai, 2008). This may also explain the observation of Kanavos & Costa-Font (2005) that source countries of parallel import such as France, Italy, and Portugal have changed to a lower level of regulation.

4.1.3 Parallel Trade and Price Caps in the North and South (PP)

Consider that parallel trade is allowed and both countries set price caps. In this case, the lower price cap becomes the global price cap. This is, de facto only one government decides on the price cap in both countries. Parallel trade constrains regulatory decisions in the sense that the

price cap has to be not only lower than the firm's price but also lower than the price cap in the other country to be binding.

The firm's profit is given as

$$\pi_{N,S}^{PP} = \begin{cases} q_N^{PP}\left(P_N^{PP}\right) P_N^{PP} + q_S^{PP}\left(P_S^{PP}\right) P_S^{PP} & \text{if} \quad P_N^{PP} \leq P_S^{PP} \\ \left(q_N^{PP}\left(P_S^{PP}\right) + q_S^{PP}\left(P_S^{PP}\right)\right) P_S^{PP} & \text{if} \quad P_N^{PP} > P_S^{PP}. \end{cases} \tag{7}$$

If $P_N^{PP} \leq P_S^{PP}$, the price cap in the North is P_N^{PP}, the price cap in the South is P_S^{PP}. In this case, the higher price cap does not spill over to the North. If $P_N^{PP} > P_S^{PP}$, the price cap in the South is the global price cap. In the former case, if $P_N^{PP} \leq P_S^{PP}$, $\Delta^{PP}\big|_{P^{PP}=P_N^{PP}} = \pi_{N,S}^{PT,P} - \pi_N^{P,(.)} > 0$, i.e., the firm sells to the South at any price cap P_S^{PP}, as selling in the South does not imply accepting a lower price also for sales in the North. Also, in this case, the government in the North sets a lower price than the government in the South, and parallel trade does not take place so that a potentially lower price from the South cannot spill over to the South. In the latter case, if $P_N^{PP} > P_S^{PP}$, the price cap in the South is the global price cap. The firm sells in the South, if the profit from selling at the price cap P_S^P in both markets is at least as high than selling at the (free) price p_N^x in the North only, i.e., $\Delta^{PP}\big|_{P^{PP}=P_S^{PP}} = \pi_{N,S}^{PP} - \pi_N^{xPP} \geq 0$, that is, if $P_S^{PP} \geq \widehat{P_{S,\Delta^{PP}}^{PP}}$, i.e., the price cap is sufficiently high.

As before, welfare in both countries decreases in the price cap. Lower price caps increase consumer surplus, decrease third-party payer expenditure and decrease the firm's profit, increasing welfare in both countries. Governments set price caps $P_N^{PP*} = 0$, $P_S^{PP*} = 0$. Under these equilibrium price caps, the firm sells to the South as sales to the South do not come at the cost of accepting a lower price for sales in the North as well. The price cap in the North if the firm decided not to export to the South would be $P_N^{xPP*} = 0$, this is, it would be thus the same.

This is, the preference of the North for a low price cap deprives the firm of its threat of non-supply to the South, and thus the firm cannot use this threat to press for a higher price cap in the South. If both countries set price caps, parallel trade does not induce a lower level of regulation in either country. The North's preference for a low price cap allows the South to also choose a low price cap without being constrained by the firm's threat of not exporting to the South.

Proposition 1 summarizes the effect of parallel trade on the level of regulation for a price cap only in the South and price caps in both countries.

Proposition 1 *i) If only the government in the South sets a price cap, parallel trade relaxes regulation. ii) If both governments in the North and South set price caps, parallel trade has no effect on the level of regulation in both countries.*

The effect of parallel trade on regulation depends on whether only the South sets price caps or both countries do. If only the South sets a price cap, parallel trade relaxes regulation. If both countries apply price caps, parallel trade does not affect regulatory decisions. Then the preference of the North for a low price cap deprives the firm of its threat of non-supply and enables the South to also set a low price cap. The firm can only make use of the threat of non-supply if it has the option of selling at a higher price in the North. The possibility of parallel trade to prevent a high level of regulation in source countries is thus contingent on the regulatory decision in the North.

4.2 Reimbursement Limits in Both Countries (RR) or the South only ($\varnothing R$)

Consider now a scenario, in which both governments set or only the government in the South sets reimbursement limits. This implies that governments do not regulate prices directly but third-party payer reimbursement. The regulation of reimbursement separates the copayment paid by the consumer from the amount financed by the third-party payer, and from the price received by the firm. Governments restrict reimbursement to $r_j \leq p_j^R$ in country j. The firm is free in setting prices. Reimbursement limits change the consumer copayment in country j to $c_j^R = \gamma_j r_j + p_j^R - r_j$ and reimbursement to $\left(1-\gamma_j\right) r_j$. Consider binding reimbursement limits, i.e. $r_N \leq \widetilde{r_N}$ so that $p_N^{R(.)} - r_N \geq 0$ and $r_S < \widetilde{r_S}$ so that $p_S^{(.)R} - r_S \geq 0$.

Consider that governments set reimbursement limits to maximize welfare and can commit to reimbursement limits. In the first stage, governments set reimbursement limits; in the second stage, the firm decides whether to export and in the third stage the firm sets prices. Reimbursement limits and welfare can be found in Appendix A.3.

4.2.1 No Parallel Trade ($x, \varnothing R$, x, RR)

Consider first that parallel trade is not allowed. As drug prices and policy choices are independent, the equilibria in which i) only the government in the South ($x, \varnothing R$) sets reimbursement limit $r_S^{x,R}$ and in which ii) both governments in the North and the South set reimbursement

limit $r_N^{x,R}$, $r_S^{x,R}$ (x, RR) are discussed jointly. The firm's profit is given as

$$\pi_{N,S}^{x,(.)R} = \begin{cases} q_N\left(p_N^{x,\varnothing R}\right)p_N^{x,\varnothing R} + q_S\left(p_S^{x,\varnothing R}\right)p_S^{x,\varnothing R} & \text{for} \quad \varnothing R \\ q_N\left(p_N^{x,RR}\right)p_N^{x,RR} + q_S\left(p_S^{x,RR}\right)p_S^{x,RR} & \text{for} \quad RR. \end{cases} \tag{8}$$

In equilibrium $\varnothing R$, the firm sets drug prices $p_N^{x,\varnothing R} = \frac{\mu}{2\gamma_N}$ and $p_S^{x,\varnothing R} = \frac{1+r_S^{x,\varnothing R}(1-\gamma_S)}{2}$.[16] The drug price in the South increases in the reimbursement limit. Welfare in the South decreases in the reimbursement limit. A (potential) decrease in the reimbursement limit increases the consumer copayment for a given drug price and decreases the quantity demanded, decreasing consumer surplus and decreasing third-party payer expenditure. The overall effect of a (potential) decrease in the reimbursement limit on welfare is positive. The government in the South sets $r_S^{x,\varnothing R*} = 0$. Patients pay the full price of the drug out-of-pocket. The government in the South removes insurance coverage by limiting reimbursement.

In equilibrium RR, the firm sets drug prices $p_N^{xR} = \frac{\mu+r_N(1-\gamma_N)}{2}$ and $p_S^{xR} = \frac{1+r_S(1-\gamma_S)}{2}$.[17] In the North, welfare increases in the reimbursement limit. A (potential) decrease in the reimbursement limit increases the consumer copayment for a given drug price[18] and decreases the quantity demanded. This decreases consumer surplus, the firm's profit, and decrease third-party payer expenditure, with an overall negative effect on welfare. Therefore, the government sets the highest possible binding reimbursement limit, $r_N^{x,RR*} = \widetilde{r_N^{x,\varnothing RR}}$. As $p_N^{x,RR}\left(\widetilde{r_N^{x,\varnothing RR}}\right) = \widetilde{r_N^{x,\varnothing RR}}$, there is no additional copayment for patients. This implies that the government in N chooses not to limit the reimbursement, but reimburses the full drug price.

In the South, welfare decreases in the reimbursement limit. The government in the South sets $r_S^{x,RR*} = 0$. The government in the South removes insurance coverage.

This shows that without parallel trade, regulatory preferences differ among the two countries, with the North preferring a very low level of regulation, while the South prefers a very high level of regulation.

Lemma 2 summarizes regulatory preferences in North and South for reimbursement limits.

Lemma 2 *Under reimbursement limits, the North prefers a low level of regulation, i.e., reimbursement of the drug price, and the South prefer a high level of regulation, i.e., a reimbursement*

[16] A binding reimbursement limit implies $r_S \leq \widetilde{r_S^{x,\varnothing R}} = \frac{1}{\gamma_S+1}$.

[17] Binding reimbursement limits imply $r_N \leq \widetilde{r_N^{x,RR}} = \frac{\mu}{\gamma_N+1}$ and $r_S \leq \widetilde{r_S^{x,RR}} = \frac{1}{\gamma_S+1}$.

[18] The drug price increases in the reimbursement limit. A decrease in the reimbursement limit decreases the drug price, but the decrease in the drug price does not fully compensate the decrease in the reimbursement limit and the consumer copayment increases.

limit of zero.

4.2.2 Parallel Trade and Reimbursement Limit in the South ($\varnothing R$)

Consider now that parallel trade is allowed and only the South sets a reimbursement limit $r_S^{\varnothing R}$. The firm sells at a uniform price p_{NS}^R in both countries. The firm's profit is

$$\pi_{N,S}^{\varnothing R} = \left(q_N\left(p_{NS}^{\varnothing R}\right) + q_S\left(p_{NS}^{\varnothing R}\right)\right) p_{N,S}^{\varnothing R}. \tag{9}$$

The firm sets the uniform price $p_{NS}^{\varnothing R} = \frac{\mu\left(2+r_S^{\varnothing R}(1-\gamma_S)\right)}{2(\gamma_N+\mu)}$.[19] The drug price increases in the reimbursement limit in the South. The firm exports to the South under parallel trade if the profit from selling to both countries at a uniform price is at least as high as the profit from selling to only the North at the country-specific price (and not selling to the South), i.e. $\Delta^{\varnothing R} = \pi_{N,S}^{\varnothing R} - \pi_N^{x,\varnothing R} \geq 0$.

The government in the South maximizes $W_S^{\varnothing R}$ subject to $\Delta^{\varnothing R} \geq 0$, i.e., the firm exporting to the South. Depending on demand in the North, two cases are possible: For $\mu < \widetilde{\mu^{\varnothing R}}$, the government sets a reimbursement limit of $r_S^{\varnothing R*}\Big|_{\mu<\widetilde{\mu^{\varnothing R}}} = 0$; for $\mu > \widetilde{\mu^{\varnothing R}}$, the government sets a reimbursement limit of $r_S^{\varnothing R*}\Big|_{\mu>\widetilde{\mu^{\varnothing R}}} \geq 0$.

Parallel trade does not affect the level of regulation in the South if demand in the North is sufficiently low. In this case, the difference in the uniform price under parallel trade (and regulation in the South) and the price that the firm would charge in the North if it decided not to sell to the South is rather small, so that the benefit of additional sales to the South set off the cost of accepting a lower uniform price under parallel trade. A high level of regulation and the corresponding decrease in the uniform price then have a rather small effect on the firm's export decision.

Parallel trade relaxes regulation, i.e., it increases the reimbursement amount, if demand in the North is sufficiently high. In this case, the price the firm would charge in the North if it did not export to the South and the uniform price under parallel trade are rather different. If a high level of regulation lowers the price in the South and hence the uniform price under parallel trade even more, the firm does not export to the South. In this case, similar to the case of price caps, parallel trade prevents the government in the South from setting a very low reimbursement limit.

[19] A binding reimbursement limit implies $r_S \leq \widetilde{r_S^{\varnothing R}} = \frac{2\mu}{2\gamma_N+\mu(1+\gamma_S)}$.

Also under reimbursement limits, parallel trade has the potential to relax regulation but only if sufficiently high differences in demand and/or coinsurance rates between countries put exports to the South at risk. If countries are sufficiently similar, parallel trade does not relax regulation under reimbursement limits. The firm's threat of not selling to the South may not help to achieve a higher reimbursement level if small differences in demand do not provide the firm with a profitable outside option when not selling in the South.

Compared to price caps, the potential of parallel trade to prevent a high level of regulation under reimbursement limits is lower. Price caps enforce maximum prices set by the government. Free pricing under reimbursement limits implies that the firm may avoid a very low drug price even for a high level of regulation, but the firm loses the threat of non-supply at the same time.

4.2.3 Parallel Trade and Reimbursement Limits in the North and South (RR)

Consider now that parallel trade is allowed and both governments set reimbursement limits r_N^{RR} and r_S^{RR}. The firm's profit is

$$\pi_{N,S}^{RR} = \left(q_N\left(p_{NS}^{RR}\right) + q_S^{RR}\left(p_{N,S}^{RR}\right)\right) p_{NS}^{RR}. \tag{10}$$

The firm sets the uniform price $p_{NS}^{RR} = \frac{2\mu+\mu r_S^{RR}(1-\gamma_S)+r_N^{RR}(1-\gamma_N)}{2(\mu+1)}$.[20] The drug price increases in both reimbursement limits. The firm exports to the South under parallel trade if the profit from selling to both countries at a uniform price is at least as high as the profit from selling to only the North at a country-specific price (and not selling to the South)., i.e., $\Delta^{RR} = \pi_{N,S}^{RR} - \pi_N^{xR} \geq 0$. If demand in the North is sufficiently low ($\mu \leq \widehat{\mu_{\Delta^{RR}}}$), the firm sells to the South.

The government in the North maximizes W_N^{RR}, the government in the South maximizes W_S^{RR} subject to $\Delta^{RR} \geq 0$, i.e., the firm exporting to the South. There is strategic interaction of reimbursement limits via the uniform price under parallel trade. Optimal (unconstrained) reimbursement limits are strategic complements. An increase in the reimbursement limit in the South increases the drug price, which in turn c.p. increases the copayment and reduces the quantity demanded in the North. The increase of the copayment and decrease in quantity reduces consumer surplus and third-party payer expenditure. The firm's profit decreases as well. An increase of the reimbursement limit in the North then reduces the copayment and increases the quantity demanded, countervailing the effect on consumer surplus, third-party

[20] Reimbursement limits are binding for $r_N \leq \widetilde{r_N^{RR}} = \frac{\mu\left(2+r_S^{RR}(1-\gamma_S)\right)}{\gamma_N+2\mu+1}$ and $r_S^{RR} < \widetilde{r_S^{RR}} = \frac{2\mu+r_N^{RR}(1-\gamma_N)}{\mu+\mu\gamma_S+2}$.

payer expenditure, and the firm's profit. Similarly, an increase in the reimbursement limit in the North increases the copayment and reduces the quantity demanded in the South, decreasing consumer surplus and third-party payer expenditure. If the government in the North then raises the reimbursement limit, the reduction of the copayment and increase in the quantity demanded increase consumer surplus and third-party payer expenditure.

Depending on demand in the North, two cases are possible. First, if countries are sufficiently similar ($\mu < \widetilde{\mu^{RR}}$), countries set (unconstrained) reimbursement limits to maximize welfare. The government in the South sets a reimbursement limit of zero ($r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} = 0$) (as under no parallel trade). The government in the North sets a reimbursement limit ($r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} > 0$) that is lower than under no parallel trade. The lower drug price under parallel trade decreases the copayment and increases demand for the drug c.p., which increases consumer surplus and increases third-party payer expenditure.[21] Then the government in the North reduces the reimbursement limit to countervail this effect, reducing third-party payer expenditure. Second, if countries differ in market size sufficiently ($\mu \geq \widetilde{\mu^{RR}}$), the government in the South sets a positive reimbursement limit ($r_S^{RR*}\big|_{\mu\geq\widetilde{\mu^{RR}}} > 0$), which is compatible with the firm exporting to the South. The North sets a lower reimbursement limit than under no parallel trade ($r_N^{RR*}\big|_{\mu\geq\widetilde{\mu^{RR}}} > 0$) for the same reason as described above.

Proposition 2 summarizes the effect of parallel trade on the level of regulation for a reimbursement limit only in the South and reimbursement limits in both countries.

Proposition 2 *i) If only the government in the South sets a reimbursement limit, parallel trade does not affect the level of regulation if $\mu < \widetilde{\mu^R}$ and decreases the level of regulation if $\mu \geq \widetilde{\mu^R}$. ii) If both governments in the North and the South set reimbursement limits, parallel trade increases the level of regulation in the North and a) does not affect the level of regulation in the South if $\mu < \widetilde{\mu^{RR}}$ and b) decreases the level of regulation in the South if $\mu \geq \widetilde{\mu^{RR}}$.*

If governments set reimbursement limits, parallel trade results in regulatory convergence. If demand in the North is sufficiently small, parallel trade decreases the reimbursement limit in the North and has no effect on the reimbursement limit in the South. If demand in the North is sufficiently large ($\mu > \widetilde{\mu_{RR}}$), parallel trade decreases the reimbursement limit in the North and increases the reimbursement limit in the South. In both cases, parallel trade generates regulatory convergence.

[21] Note that there is no price effect for third-party payer expenditure as the reimbursement limit r_N is the basis for reimbursement per unit purchased.

This is, there may be no need for explicit harmonization of regulatory decisions but market integration through parallel trade may align regulatory decisions under reimbursement limits.

While under both price caps and reimbursement limits, parallel trade may relax regulation, market outcomes under the two instruments differ: Under reimbursement limits, more relaxed regulation implies higher reimbursement limits and lower copayments (i.e., a shift of the financing burden from the patient to the insurer); whereas under price caps, relaxed regulation results in higher price caps, that is, higher copayments and higher reimbursement (i.e., an increase in the transfer from patient and insurer to the firm).

4.3 "Asymmetric" Equilibria

4.3.1 No Parallel Trade, Price Cap in the North, Reimbursement Limit in the South (x, PR)

Consider first that parallel trade is not allowed and the government in the North sets a price $P_N^{x,PR}$ and the government in the South sets a reimbursement limit $r_S^{x,PR}$. The firm's profit is given as

$$\pi_{N,S}^{x,PR} = q_N\left(P_N^{x,PR}\right) P_N^{x,PR} + q_S\left(p_S^{x,PR}\right) p_S^{x,PR}. \tag{11}$$

Under no parallel trade, pricing and regulatory decisions are independent. Therefore, the price for the North is the same as in the equilibrium $\varnothing P$ or PP, the price for the South is the same as in the equilibrium $\varnothing R$ or RR. In the North, the government sets $P_N^{x,PR*} = 0$. In the South, the firm sets the drug price $p_S^{x,PR} = \frac{1+r_S(1-\gamma_S)}{2}$[22] and the government in the South sets $r_S^{xR*} = 0$.

4.3.2 Parallel Trade and Price Cap in the North, Reimbursement Limit in the South (PR)

Consider now that parallel trade is allowed and the government in the North sets a price cap P_N^{PR} and the government in the South sets a reimbursement limit r_S^{PR}. The firm's profit is given as

$$\pi_{N,S}^{PR} = q_N\left(P_N^{PR}\right) P_N^{PR} + q_S\left(p_S^{PR}\right) p_S^{PR} \tag{12}$$

The price cap in the North is P_N^{PR}, the (freely set) drug price in the South is p_S^{PR}. In this case, a binding price cap in the North is lower than the drug price in the South and parallel

[22] Binding reimbursement limits imply $r_S \leq \widetilde{r_S} = \frac{1}{\gamma_S+1}$.

trade does not occur. If $P_N^{PR} > p_S^{PR}$, the price cap in the North would not be binding and the drug price in the South is the global price.

In the South, the firm sets the drug price $p_S^{PR} = \frac{1+r_S^{PR}(1-\gamma_S)}{2}$. The firm always sells to the South, i.e., $\Delta^{PR} = \pi_{N,S}^{PR} - \pi_N^{xP} > 0$. Selling in the South does not imply accepting a lower price for sales also in the North, as parallel trade does not take place. In the North, welfare decreases in the price cap and therefore the government sets a price cap of $P_N^{PR*} = 0$. In the South, welfare decreases in the reimbursement limit and thus the government sets $r_S^{PR,P*} = 0$. Parallel trade does not affect regulatory decisions in both countries.

4.3.3 No Parallel Trade, Reimbursement Limit in the North, Price Cap in the South (RP)

Consider that parallel trade is not allowed and the government in the North sets a reimbursement limit $r_N^{x,RP}$ and the government in the South sets a price cap $P_S^{x,RP}$. The firm's profit is given as

$$\pi_{N,S}^{x,PR} = q_N\left(p_N^{x,RP}\right)P_N^{x,RP} + q_S\left(P_S^{x,RP}\right)P_S^{x,RP}. \tag{13}$$

Under no parallel trade, pricing and regulatory decisions are independent. Therefore, the price for the North is the same as in the equilibrium $\varnothing R$ or RR; the price for the South is the same as in the equilibrium $\varnothing P$ or PP. In the North, the firm sets the drug price $p_N^{x,RP} = \frac{\mu+r_N^{x,RP}(1-\gamma_N)}{2}$[23] and the government sets $r_N^{x,RP*} = \widetilde{r_N}$. In the South, the government sets the price cap $P_S^{x,RP*} = 0$.

4.3.4 Parallel Trade and Reimbursement Limit in the North, Price Cap in the South (RP)

Consider that parallel trade is allowed and the government in the North sets a reimbursement limit r_N^{RP} and the government in the South sets a price cap P_S^{RP}. The firm's profit is given as

$$\pi_{N,S}^{PR} = q_N\left(P_S^{RP}\right)P_S^{RP} + q_S\left(P_S^{RP}\right)P_S^{RP}. \tag{14}$$

The price cap in the South is the global price cap. This implies that the drug price in the North is P_S^{RP}.[24] The firm sells to the South if the profit from selling to both countries at a

[23] Binding reimbursement limits imply $r_N \leq \widetilde{r_N} = \frac{\mu}{\gamma_N+1}$.

[24] A binding reimbursement limit implies $r_N \leq \widetilde{r_N^{RP}} = P_S^{RP}$.

uniform price is at least as high as the profit from selling to only the North at the drug price p_N^{RP} (and not selling to the South), i.e., $\Delta^{RP} = \pi_{N,S}^{RP} - \pi_N^{xP} \geq 0$. In the North, welfare increases in the reimbursement limit. Therefore, the government sets the highest possible binding reimbursement limit, $r_N^{x,RP*} = \widetilde{r_N^{x,\varnothing RP}}$. The government in the South maximizes W_S^{RP} subject to $\Delta^{RP} = \pi_{N,S}^{RP} - \pi_N^{xP} \geq 0$, i.e., the firm selling to the South. Given that the government in the South would prefer a price cap of zero, but the firm would not export to the South at a very low price cap, the government in South sets $P_S^{RP*} > 0$, which is the lowest price cap compatible with the firm selling in the South. Parallel trade prevents the government in the South from setting a very low price cap. As in the equilibrium $\varnothing P$, parallel trade prevents the government in the South from setting a very low price cap.

Proposition 3 summarizes the effect of parallel trade on the level of regulation for asymmetric equilibria with a price cap in one country and a reimbursement limit in the other country.

Proposition 3 *i) If the government in the North sets a price cap and the government in the South sets a reimbursement limit, parallel trade does not affect the level of regulation in either country. ii) If the government in the North sets a reimbursement limit and the government in the South sets a price cap, parallel trade increases the level of regulation in the North and decreases the level of regulation in the South.*

Table 1 summarizes all regulatory choices under no parallel trade and parallel trade.

North, South	$\varnothing$	P	R
$\varnothing$		$P_S^{x,\varnothing P*} = 0$ $P_S^{\varnothing P*} > 0$	$r_S^{x,\varnothing R*} = 0$ $r_S^{\varnothing R*}\big\|_{\mu<\widetilde{\mu^{\varnothing R}}} = 0$ if $\mu < \widetilde{\mu^{\varnothing R}}$ $r_S^{\varnothing R*}\big\|_{\mu>\widetilde{\mu^{\varnothing R}}} \geq 0$ if $\mu > \widetilde{\mu^{\varnothing R}}$
P		$P_N^{x,PP*} = 0$, $P_S^{x,PP*} = 0$ $P_N^{PP*} = 0$, $P_S^{PP*} = 0$	$P_N^{x,P*} = 0$, $r_S^{xR*} = 0$ $P_N^{PR*} = 0$, $r_S^{PR,P*} = 0$
R		$r_N^{x,RP*} = \widetilde{r_N}$, $P_S^{x,RP*} = 0$ $r_N^{x,RP*} = \widetilde{r_N^{x,\varnothing RP}}$, $P_S^{RP*} > 0$	$r_N^{x,RR*} = \widetilde{r_N^{x,\varnothing RR}}$, $r_S^{x,RR*} = 0$ $r_N^{RR*}\big\|_{\mu<\widetilde{\mu^{RR}}} > 0$, $r_S^{RR*}\big\|_{\mu<\widetilde{\mu^{RR}}} = 0$ if $\mu < \widetilde{\mu^{RR}}$ $r_N^{RR*}\big\|_{\mu>\widetilde{\mu^{RR}}} > 0$, $r_S^{RR*}\big\|_{\mu>\widetilde{\mu^{RR}}} > 0$ if $\mu > \widetilde{\mu^{RR}}$

Table 1: Equilibrium regulatory choices

I can now characterize the complete effect of parallel trade on regulation for price caps and reimbursement limits: Under price caps, parallel trade has no effect on the level of regulation in the North and decreases the level of regulation in the South if either the North does not regulate or if the North sets a reimbursement limit. If the North also sets a price cap, parallel

trade does not affect the level of regulation in the South, as the North's preference for a low price cap allows the South to also set a low price cap without risking to the firm not selling to the South. Under reimbursement limits, parallel trade increases the level of regulation in the North to counteract the effect of a lower drug price on expenditure and to further decrease the drug price to increase welfare. Parallel trade decreases the level of regulation in the South if countries are sufficiently different in terms of demand and if the North does not set a price cap. Parallel trade does not affect the level of regulation in the South if countries are sufficiently similar in terms of demand or if the North sets a price cap. The firm's threat of not selling to the South may not achieve a lower level of regulation if small differences in demand or the North's preference for a low price cap do not provide the firm with a profitable outside alternative when selling only in the North.

5 Choice of Regulatory Instrument - First-stage Outcome

Consider now the choice of regulatory instruments in both countries.

5.1 No Parallel Trade

Under no parallel trade, the choice of the regulatory instrument in both countries is independent. The choice of the regulatory instrument in the South affects welfare in the North via the firm's profit, but there is no strategic effect of the North's choice of policy instrument on the policy choice in the South or vice versa. In the South, welfare is the same, irrespective of the instrument chosen by the North.

Both countries prefer price caps over reimbursement limits. In both countries, welfare decreases in the drug price. Price caps allow governments to set a drug price of zero, whereas reimbursement limits with the free pricing always generate a higher drug price. In addition, under price caps, governments may increase consumer surplus and decrease third-party payer expenditure at the same time. Under reimbursement limits, however, there is a trade-off between increasing consumer surplus and decreasing expenditure: A decrease in the reimbursement limit increases the copayment, thus decreasing third-party payer expenditure, but also decreasing consumer surplus.

5.2 Parallel Trade

Under parallel trade, the choices of the regulatory instrument are linked via the price. If the South sets a price cap, the (lower) price cap translates directly to the drug price in the North. If the South sets a reimbursement limit, the uniform price under parallel trade increases in the reimbursement limit in the South.

The North chooses a price cap if the South set a reimbursement limit and the North chooses a reimbursement limit if the South set a price cap. If the South sets a reimbursement limit, the North chooses a price cap, as it generates a lower drug price and thus higher welfare in the North. Compared to the equilibrium in which both countries choose reimbursement limits, welfare in the North is higher as the drug price is lower, consumer surplus is higher, and third-party payer expenditure is lower. If the South sets a price cap, the North sets a reimbursement limit to enforce a higher price cap in the South. The North accepts a higher drug price and thus lower consumer surplus and higher third-party payer reimbursement at the benefit of boosting the firm's profit. Compared to the equilibrium in which both countries choose price caps, welfare in the North is higher as the firm's profit is higher as the drug price in the South is higher, although consumer surplus is lower and third-party payer expenditure is higher.

The South always prefers price caps over reimbursement limits, irrespective of the policy choice in the North. Welfare decreases in the drug price and price caps allow the government in the South to achieve a lower drug price as compared to a reimbursement limit.

In the resulting equilibrium, the North sets a reimbursement limit, while the South sets a price cap. In this equilibrium, the South sets a price cap higher than zero that is compatible with the firm exporting to the South. This price cap also applies in the North, where the government sets a reimbursement limit equal to the price cap.

Parallel trade diversifies regulation in the sense that it results in different instruments being applied in the two countries. Compared to the equilibrium under no parallel trade, where both countries set price caps of zero, parallel trade relaxes regulation. This suggests that the potential of parallel trade to relax regulation also holds under endogenous health policy choice.

Proposition 2 summarizes the choice of regulatory instruments.

Proposition 4 *Under no parallel trade, both the North and the South choose price caps in equilibrium. Under parallel trade, the North chooses a reimbursement limit, and the South chooses a price cap in equilibrium.*

6 Discussion

6.1 Country Differences

The countries North and South differ in two respects. Demand in the North is higher than in the South and the firm's profits are considered only in the social welfare function of the North. The demand assumption drives the magnitude of the effects. The profit assumption is crucial for the effect of parallel trade on the level of regulation under price caps compared to reimbursement limits.

If the firm were a multinational enterprise located in a third country or if the government did not take the firm's profit into account, regulatory preferences in the North would change. In this case, the North would only care about consumer surplus and third-party payer expenditure, just like the South. Then the case of reimbursement limits in both countries would be similar to the case of price caps: The North would prefer a high level of regulation, i.e., a reimbursement limit of zero. For reimbursement limits in both countries, parallel trade had no effect on the level of regulation. Under endogenous regulatory instrument choice, the North would then choose a price cap (of zero), and parallel trade would not affect the choice of regulatory instruments in both countries.

If the government in the North considered only a part of the profits in the social welfare function, e.g., local profits of a multinational firm, results would be qualitatively the same as in the main part of the paper.

The profit assumption in the paper could also reflect the fact that the government in the North takes into account that the firm must be able to fund and should have the ability to develop new drugs while the government in the South behaves mypoic (or free rides).

Differences in demand drive the magnitude of the effect of parallel trade. The price difference, and, therefore, the threat of non-supply, increase in μ. A higher demand difference aggravates the effect of a threat of non-supply on regulatory decisions.

6.2 Governments Objectives

Governments might give more weight to minimizing third-party payer expenditure than maximizing welfare. In this case, the South would have a similar preference for strict regulation under both instruents, as a low price cap and a low reimbursement limit decrease third-party payer expenditure. The North, however, would prefer a low price cap but would also choose a

low reimbursement limit if minimizing third-party payer expenditure was more important. In this case, the effect of parallel trade on regulation under reimbursement limits would be similar than the effect under price caps, a preference for a high level of regulation in the North would eliminate the firm's threat of non-supply to the South. This allows the South to pursue strict regulation without risking the non-supply.

Without parallel trade, both governments would prefer strict price caps. This could raise the question of why governments are then interested in parallel trade at all. The model suggests that under certain conditions, parallel trade results in welfare gains. In the European Union, however, parallel trade is based on Articles 34 and 35 of the Treaty on the Functioning of the European Union (TFEU) and the related decisions of the European Court of Justice. From the perspective of an individual member state, it is not open for disposition.

The model is static. Both governments take drug development as a given. In particular, they do not consider the impact of regulation on drug development. Since new drug development is a long-term process, this could be explained by the myopia of governments. Of course, it is conceivable that governments also consider the long-term dynamic effects (see, e.g., Grossman & Lai, 2006). This factor weakens the preference for low prices in both countries. However, the tradeoff between short- and long-term effects is beyond the model and is left for future research.

6.3 Cooperative Governments

Governments might also cooperate in choosing the level of regulation or setting regulatory instruments. Under price caps, both countries choose a price cap of zero, but a higher price cap in the South would increase welfare in the North (via the firm's profit). Under cooperation (and a potential side payment) the South could potentially be motivated to increase its price cap, thereby reducing consumer surplus and increasing third-party payer expenditure, while welfare in the North would increase due to an increase in the firm's profit. Under reimbursement limits, cooperation would relax regulation in both countries. This would increase the drug price and reduce the quantity demanded, thereby reducing consumer surplus, reducing third-party payer expenditure, and increasing the firm's profit.

6.4 Direction and Volume of Parallel Imports

The model assumes that pricing interdependence under parallel trade induces the manufacturer to set a uniform price under parallel trade. This assumption can also be found in Pecorino

(2002), Valletti (2006), Roy & Saggi (2012), and Bennato & Valletti (2014). Empirical evidence on whether parallel trade erodes price differences, however, is ambiguous. Competition from parallel trade may affect only a part of the market in the North, e.g., because some consumers associate a lower quality with the parallel import due to differences in appearance and packaging (Maskus, 2000) or because manufacturers try to restrict parallel trade by differentiating products across countries (Kyle, 2011). In this case, only a part of the market in the North would be affected by parallel trade and the firm could set a different (and higher) price for the part of the market that is isolated from parallel trade. Results from the model would qualitatively be similar but weaker. For example, for a significant difference in demand between the North and the South, it would be optimal for the firm to sell to the North only and not export to the South in order to retain a high price in the North and to avoid the spillover of the lower price in the South to part of the market in the North.

The model only considers parallel trade from the South to the North. If the drug price in the North is lower than the drug price in the South, parallel trade transmits the lower price in the North to the South. Two cases can be distinguished: If both countries set price caps, both prices are zero, so parallel trade has no effect. If the North sets a price cap and the South applies a reimbursement limit, parallel trade has no effect on the price in the North but results in a lower price in the South. The main results of this paper holds in this case.

If the firm does not export to the South, parallel trade from the North to the South could make the drug available in the South. However, this undermines the firm's decision not to export to the South if the price in the South is too low. This mechanism weakens the threat of the drug being unavailable in the South and allows the government in the South to set higher prices. However, parallel imports typically take place if a locally sourced version of a drug is available.

7 Conclusion

In this paper, I have studied the effect of parallel trade on pharmaceutical regulation in a North-South framework. An innovative firm located in the North can sell its drug only in the North or in both countries. Governments may set price caps or may limit reimbursement for the drug.

Parallel trade may relax regulation in the source country under both regulatory instruments: Under price caps, it only decreases the level of regulation if the destination country

does not set a price cap. Under reimbursement limits, it only does so if both countries are sufficiently different with respect to demand and if the destination country does not set a price cap. The manufacturer's threat of not supplying the source country of parallel imports requires a profitable alternative, i.e., the loss in profit from not selling in the source country has to be compensated by selling at a higher price in the destination country.

Parallel trade may intensify regulation in the destination country but only under reimbursement limits. Under price caps, the destination country prefers a high level of regulation, i.e., a price cap of zero, as it increases consumer surplus and decreases third-party payer expenditure. Under reimbursement limits, the destination country prefers a low level of regulation, i.e., reimbursement of the full drug price, as a decrease in the reimbursement limit would decrease third-party payer expenditure, but also consumer surplus. [25] In response to the increase in third-party payer expenditure due to the lower drug price under parallel trade, the government in the destination country decreases the reimbursement limit. By relaxing regulation in the source country and/or intensifying regulation in the destination country, parallel trade results in regulatory convergence. This implies, that for given regulatory instruments, parallel trade may be a substitute for policy harmonization in some cases, e.g., if both countries set reimbursement limits, but not in others, e.g., if both countries set price caps. Whether parallel trade may replace policy harmonization depends on whether the firm has a profitable outside option when not selling in the South, i.e., the regulatory choice in the destination country or on differences in demand between countries.

Parallel trade may change regulatory preferences: Under no parallel trade, both the North and the South prefer price caps over reimbursement limits. Welfare in both countries decreases in the drug price. Price caps allow governments to set a drug price of zero, while free pricing under reimbursement limits yields a higher drug price. Under parallel trade, the South sets a price cap, as it allows attaining a lower drug price. The North applies a reimbursement limit, as it allows the North to enforce a higher price cap in the South. The North accepts a higher drug price and thus lower consumer surplus and higher third-party payer reimbursement at the benefit of boosting the firm's profit. The reimbursement limit in the North allows enforcing a higher drug price in the South than under price caps. This result complements the result in the previous literature on the effect of parallel trade on regulation, namely that governments in source countries may refrain form strict regulation if they take into account the impact of

[25] Under both instruments, a high level or regulation decreases the firm's profit.

regulatory decisions on the firm's decision to supply the respective country. For endogenous health policy choice, both source and destination countries of parallel imports choose less strict regulation with higher drug prices. This result also implies that the ability of parallel trade to replace policy harmonization is no longer dependent on differences in demand between countries. Moreover, this effect seems not to be brought about by the firm's profit maximization but is also in the interest of the welfare-maximizing government in the destination country.

Under endogenous health policy choice, parallel trade results in a lower level of regulation and higher drug prices. Compared to the equilibrium without parallel trade, in which both countries set price caps of zero, drug prices under parallel trade are higher. This also worsens access in both the source and destination country.

The firm's decision not to sell to the South and thus both the effect of parallel trade on the level regulation in the South and the effect of parallel trade on the choice of regulatory instrument in the North are driven by the decrease of the firm's profit. The lower profit of the firm may also induce a dynamic consequence if lower profits translate to less investment in innovation. This is left for further research.

References

[1] Bennato, A. R. & Valletti, T. (2014): Pharmaceutical innovation and parallel trade. International Journal of Industrial Organization 33, 83 - 92.

[2] Brekke, K. R.; Holmas, T. & Straume, O. R. (2011): Reference pricing, competition, and pharmaceutical expenditures: theory and evidence from a natural experiment, Journal of Public Economics 95, 624 - 638.

[3] Carone, G.; Schwierz C. & Xavier, A. (2012): Cost-containment policies in public pharmaceutical spending in the EU. European Commission Economic Paper 461, http://ec.europa.eu/economy_finance/publications/economic_paper/2012/pdf/ecp_461_en.pdf.

[4] Danzon, P. M. (2001): Reference Pricing: Theory and Evidence, in: Lopez-Casanovas, G. & Jonsson, B. (eds.): Reference Pricing and Pharmaceutical Policy: Perspectives on Economics and Innovation, Barcelona, 86 - 126.

[5] Danzon, P. M. & Epstein, A. J. (2012): Effects of Regulation on Drug Launch and Pricing in Interdependent Markets, in: Bolin, K. & Kaestner, R. (ed.): The Economics of Medical Technology, Emerald Group Publishing Limited, 35 - 71.

[6] Danzon, P.M.; Wang, Y. R. & Wang, L. (2005): The impact of price regulation on the launch delay of new drugs-evidence from twenty-five major markets in the 1990s, Health Economics, 14(3), 269 - 292.

[7] Duso, T.; Herr, A. & Suppliet, M. (2014.). The Welfare Impact Of Parallel Imports: A Structural Approach Applied To The German Market For Oral Anti-Diabetics. Health Economics 23, 1036 - 1057.

[8] EFPIA (2020): The Pharmaceutical Industry in figures - Edition 2020, https://efpia.eu/media/554521/efpia_pharmafigures_2020_web.pdf.

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

[10] 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://ec.europa.eu/enterprise/sectors/healthcare/files/docs/study_pricing_2007/andalusian_school_public_h.

[11] Ganslandt, M. & Maskus, K. E. (2004): Parallel imports and the pricing of pharmaceutical products: evidence from the European Union, Journal of Health Economics 23, 1035 - 1057.

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

[13] Granlund, D. & Köksal, M. Y. (2011): EU enlargement, parallel trade and price competition in pharmaceuticals: what's to blame? ; derogation or perception?. Gothenburg University Working Paper No. 499.

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

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

[16] Heuer, A.; Mejer, M. & Neuhaus, J. (2007): The National Regulation of Pharmaceutical Markets and the Timing of New Drug Launches in Europe. Kiel Advanced Studies Working Paper No. 437.

[17] Houy, N. & Jelovac, I. (2019). Comparing approval procedures for new drugs. The World Economy 42(5), 1598 - 1619.

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

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

[20] 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.

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

[22] Kyle M. K. (2007): Pharmaceutical Price Controls and Entry Strategies, The Review of Economics and Statistics 89(1), 88 - 99.

[23] Kyle, M. K. (2010): Parallel Trade in Pharmaceuticals: Firm Responses and Competition Policy, in Hawk, B. E. (ed.) International Antitrust Law and Policy: Fordham Competition Law 2009.

[24] Kyle, M. (2011): Strategic Responses to Parallel Trade. The B.E. Journal of Economic Analysis & Policy 11, 1 - 34.

[25] Lopez-Casanovas, G. & Puig-Junoy, J. (2000): Review of the Literature on Reference Pricing, Health Policy 54, 87 - 123.

[26] Maskus, K. E. (2000): Parallel imports. The World Economy, 23, 1269 - 1284.

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

[28] Pecorino, P. (2002): Should the US allow prescription drug reimports from Canada?, Journal of Health Economics, 21(4), 699 - 708.

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

[30] Roy, S. & Saggi, K. (2012): Equilibrium parallel import policies and international market structure, Journal of International Economics 87, 262 - 276.

[31] Valletti, T. (2006): Differential pricing, parallel trade, and the incentive to invest. Journal of International Economics 70, 314 - 324.

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

Appendix

A.1 The Effect of Parallel Trade on the Export Decision

No Parallel Trade

Under no parallel trade, the firm's profit is $\pi^x_{N,S} = \frac{1}{\mu}\left(\mu - \gamma_N p^x_N\right) p^x_N + \left(1 - \gamma_S p^x_S\right) p^x_S$.

The equilibrium price in N is $p^x_N = \frac{\mu}{2\gamma_N}$; the equilibrium price in S is $p^x_S = \frac{1}{2\gamma_S}$. The price in N is higher than the price in S if μ is sufficiently high or γ_N is sufficiently low

($p^x_N - p^x_S = \frac{\mu\gamma_S - \gamma_N}{2\gamma_N\gamma_S} > 0$, if $\mu > \widetilde{\mu^x} = \frac{\gamma_N}{\gamma_S} \vee \gamma_N < \widetilde{\gamma^x_N} = \mu\gamma_S$). The equilibrium profit is $\pi^x_{N,S} = \frac{\gamma_N + \mu\gamma_S}{4\gamma_N\gamma_S}$.

Welfare in N is $W^x_N = \underbrace{\frac{\mu}{8}}_{CS_N} + \underbrace{\frac{(\gamma_N + \mu\gamma_S)}{4\gamma_N\gamma_S}}_{\pi^x_{N,S}} - \underbrace{\frac{(1-\gamma_N)\,\mu}{4\gamma_N}}_{E_N} = \frac{3\mu\gamma_S + 2}{8\gamma_S}$;

welfare in S is $W^x_S = \underbrace{\frac{1}{8}}_{CS_S} - \underbrace{\frac{1}{4\gamma_S}(1-\gamma_S)}_{E_N} = \frac{3\gamma_S - 2}{8\gamma_S}$.

Parallel Trade

Under parallel trade, the firm's profit is $\pi_{N,S} = \frac{1}{\mu}\left(\mu - \gamma_N p_{NS}\right) p_{NS} + \left(1 - \gamma_S p_{NS}\right) p_{NS}$.

The equilibrium uniform price is $p_{NS} = \frac{\mu}{\gamma_N + \mu\gamma_S}$. Compared to the prices under no parallel trade, the uniform price under parallel is lower than the price in N if μ is sufficiently high ($p_{NS} - p_N = -\frac{\mu(\mu\gamma_S - \gamma_N)}{2\gamma_N(\gamma_N + \mu\gamma_S)} < 0$, if $\mu > \widehat{\mu_p}$) and the uniform price under parallel is higher than the price in S if μ is sufficiently high

($p_{NS} - p_S = \frac{\mu\gamma_S - \gamma_N}{2\gamma_S(\gamma_N + \mu\gamma_S)} > 0$ if $\mu > \widehat{\mu_p}$). The equilibrium profit is $\pi^{PT}_{N,S} = \frac{\mu}{\gamma_N + \mu\gamma_S}$.

The firm exports to S if μ is sufficiently low or γ_N is sufficiently high ($\Delta = \pi_{N,S} - \pi_N = \mu\frac{3\gamma_N - \mu\gamma_S}{4\gamma_N(\gamma_N + \mu\gamma_S)} > 0$, if $\mu < \widehat{\mu_\Delta} = 3\frac{\gamma_N}{\gamma_S} \vee \gamma_N > \widehat{\gamma_{N_\Delta}} = \frac{\mu\gamma_S}{3}$).

Welfare in N is $W_N = \underbrace{\frac{\gamma_S^2\mu^3}{2\left(\gamma_N + \mu\gamma_S\right)^2}}_{CS_N} + \underbrace{\frac{\mu\left(\gamma_S\mu^2 + \gamma_N\right)}{\left(\gamma_N + \mu\gamma_S\right)^2}}_{\pi^x_{N,S}} - \underbrace{(1-\gamma_N)\frac{\gamma_S\mu^2}{\left(\gamma_N + \mu\gamma_S\right)^2}}_{E_N} = \frac{\mu\left(\mu^2\gamma_S^2 + 2\gamma_N\mu\gamma_S + 2\gamma_N\right)}{2(\gamma_N + \mu\gamma_S)^2}$; welfare in S is $W_S = \underbrace{\frac{\gamma_N^2}{2\left(\gamma_N + \mu\gamma_S\right)^2}}_{CS_S} - \underbrace{(1-\gamma_S)\frac{\gamma_N\mu}{\left(\gamma_N + \mu\gamma_S\right)^2}}_{E_N} = \frac{\gamma_N(\gamma_N - 2\mu + 2\mu\gamma_S)}{2(\gamma_N + \mu\gamma_S)^2}$. Compared to welfare under no parallel trade, welfare in N is higher under parallel trade, if μ is sufficiently low

($W_N - W^x_N = \frac{\left(2\gamma_N + \mu^2\gamma_S^2 - \mu\gamma_S(2 - 3\gamma_N)\right)(\mu\gamma_S - \gamma_N)}{8\gamma_S(\gamma_N + \mu\gamma_S)^2} > 0$, if $\mu < \widetilde{\mu_{W_N}} = \frac{2 - 3\gamma_N - \sqrt{(2 - 9\gamma_N)(2 - \gamma_N)}}{2\gamma_S}$) and welfare in S is higher under parallel trade if μ is sufficiently high ($W_S - W^x_S = \frac{(\mu\gamma_S(2 - 3\gamma_S) - 2\gamma_N - \gamma_N\gamma_S)(\mu\gamma_S - \gamma_N)}{8\gamma_S(\gamma_N + \mu\gamma_S)^2} > 0$, if $\mu > \widetilde{\mu_{W_S}} = \frac{\gamma_N(\gamma_S + 2)}{\gamma_S(2 - 3\gamma_S)}$).

A.2 Choice of Regulation Level - Second-stage Outcome

Price Caps in Both Countries (PP) or the South only ($\varnothing P$)

No Parallel Trade ($\varnothing P$, PP)

Under no parallel trade and a price cap the South only ($\varnothing P$), the firm's profit is

$\pi^{x,\varnothing P}_{N,S} = q_N\left(p^{x,\varnothing P}_N\right) p^{x,\varnothing P}_N + q_S\left(P^{x,\varnothing P}_S\right) P^{x,\varnothing P}_S$.

The equilibrium price in N is $p^{x,\varnothing P}_N = \frac{\mu}{2\gamma_N}$.

Welfare in N is $W_N^{x,\varnothing P} = \underbrace{\frac{\mu}{8}}_{CS_N^{x,\varnothing P}} + \underbrace{\frac{\mu}{4\gamma_N} + P_S^{x,\varnothing P}\left(1-\gamma_S P_S^{x,\varnothing P}\right)}_{\pi^{x,\varnothing P}} - \underbrace{\frac{(1-\gamma_N)\,\mu}{4\gamma_N}}_{E_N^{x,\varnothing P}} = \frac{3\mu}{8} + P_S^{x,\varnothing P}\left(1-\gamma_S P_S^{x,\varnothing P}\right)$.

Under no parallel trade and price caps in both countries (PP), the firm's profit is

$\pi_{N,S}^{x,PP} = q_N\left(P_N^{x,PP}\right)P_N^{x,PP} + q_S\left(P_S^{x,PP}\right)P_S^{x,PP}$.

Welfare in N is

$W_N^{x,PP} = \underbrace{\frac{\left(\mu-\gamma_N P_N^{x,PP}\right)^2}{2\mu}}_{CS_N^{x,PP}} + \underbrace{\frac{1}{\mu}\left(\mu-\gamma_N P_N^{x,PP}\right)P_N^{x,PP} + P_S^{x,PP}\left(1-\gamma_S P_S^{x,PP}\right)}_{\pi^{x,PP}} - \underbrace{(1-\gamma_N)\,\frac{1}{\mu}\left(\mu-\gamma_N P_N^{x,PP}\right)P_N^{x,PP}}_{E_N^{x,PP}}$

$= \frac{\mu^2-\gamma_N^2\left(P_N^{x,PP}\right)^2}{2\mu} + P_S^{x,PP}\left(1-\gamma_S P_S^{x,PP}\right)$.

The welfare maximizing price cap in N is $P_N^{x,PP*} = 0$.

In both cases, welfare in S is

$W_S^{x,(.)P} = \underbrace{\frac{\left(1-\gamma_S P_S^{x,(.)P}\right)^2}{2}}_{CS_S^{x,(.)P}} - \underbrace{(1-\gamma_S)\left(1-\gamma_S P_S^{x,(.)P}\right)P_S^{x,(.)P}}_{E_S^{x,(.)P}} = \frac{\left(1-\gamma_S P_S^{x,(.)P}\right)\left(1-P_S^{x,(.)P}(2-\gamma_S)\right)}{2}$.

The welfare maximizing price cap in S is $P_S^{x,(.)P*} = 0$.

In the equilibrium with a price cap in the South only ($\varnothing P$), welfare in N is $W_N^{x,\varnothing P}\left(P_N^{x,\varnothing P}, P_S^{x,(.)P}\right) = \frac{3}{8}\mu$. In the equilibrium with price caps in both countries (PP), welfare in N is $W_N^{x,PP}\left(P_N^{x,PP}, P_S^{x,(.)P}\right) = \frac{1}{2}\mu$. In both cases, welfare in S is $W_S^{x,(.)P}\left((.), P_S^{x,(.)P}\right) = \frac{1}{2}$.

Compared to no parallel trade and no regulation, welfare in N is lower under a price cap in the South only ($\varnothing P$) ($W_N^{xP}\left(P_N^{xP*}, P_S^{xP*}\right) - W_N^x = -\frac{1}{4\gamma_S} < 0$) and welfare in N is higher under price caps in both countries (PP) if μ is sufficiently high ($W_N^{x,PP}\left(P_N^{x,PP}, P_S^{x,(.)P}\right) - W_N^x = \frac{\mu\gamma_S-2}{8\gamma_S} > 0$, if $\mu > \widetilde{\mu_{W_N^{xP}}} = \frac{2}{\gamma_S}$). Welfare in S is higher under a price cap in the South only ($\varnothing P$) and price caps in both countries (PP) ($W_S^{x,(.)P}\left((.), P_S^{x,(.)P}\right) - W_S^x = \frac{(\gamma_S+2)}{8\gamma_S} > 0$).

Parallel Trade and Price Cap in the South ($\varnothing P$)

Under parallel trade and a price cap in the South only, the firm's profit is

$\pi_{N,S}^{\varnothing P} = \frac{1}{\mu}\left(\mu - \left(\gamma_N P_S^{\varnothing P}\right)\right)P_S^{\varnothing P} + \left(1-\left(\gamma_S P_S^{\varnothing P}\right)\right)P_S^{\varnothing P}$.

The firm exports to S if the price cap $P_S^{\varnothing P}$ is sufficiently high

($\Delta^{\varnothing P} = \pi_{N,S}^{\varnothing P} - \pi_N^x = \frac{8\mu\gamma_N P_S^{\varnothing P} - 4\gamma_N\left(P_S^{\varnothing P}\right)^2(\gamma_N+\mu\gamma_S) - \mu^2}{4\mu\gamma_N} > 0$, if $P_S^{\varnothing P} > \widehat{P_S^{\varnothing P}} = \frac{4\mu\gamma_N - 2\mu\sqrt{\gamma_N(3\gamma_N-\mu\gamma_S)}}{4\gamma_N(\gamma_N+\mu\gamma_S)}$).

Welfare in N is $W_N^{\varnothing P} = \underbrace{\frac{\left(\mu-\gamma_N P_S^{\varnothing P}\right)^2}{2\mu}}_{CS_N^{\varnothing P}} + \underbrace{\frac{1}{\mu}\left(\mu-\gamma_N P_S^{\varnothing P}\right)P_S^{\varnothing P} + P_S^{\varnothing P}\left(1-\gamma_S P_S^{\varnothing P}\right)}_{\pi^{\varnothing P}} - \underbrace{(1-\gamma_N)\,\frac{1}{\mu}\left(\mu-\gamma_N P_N^{\varnothing P}\right)P_N^{\varnothing P}}_{E_N^{\varnothing P}}$

$= \frac{\mu^2+2P_S^{\varnothing P}\mu-\left(P_S^{\varnothing P}\right)^2\left(\gamma_N^2+2\mu\gamma_S\right)}{2\mu}$. Welfare in S is

$W_S^{\varnothing P} = \underbrace{\frac{\left(1-\gamma_S P_S^{\varnothing P}\right)^2}{2}}_{CS_S^{\varnothing P}} - \underbrace{(1-\gamma_S)\left(1-\gamma_S P_S^{\varnothing P}\right)P_S^{\varnothing P}}_{E_S^{\varnothing P}} = \frac{\left(1-\gamma_S P_S^{\varnothing P}\right)\left(1-P_S^{\varnothing P}(2-\gamma_S)\right)}{2}$.

For a price cap of zero, the firm does not export to S ($\Delta^{\varnothing P}\big|_{P_S^{\varnothing P}=0} = -\frac{\mu}{4\gamma_N} < 0$). The welfare maximizing price cap in S is $P_S^{\varnothing P*} = \frac{4\mu\gamma_N - 2\mu\sqrt{\gamma_N(3\gamma_N-\mu\gamma_S)}}{4\gamma_N(\gamma_N+\mu\gamma_S)}$.

For the welfare-maximizing price cap in the South, welfare in N is

$W_N^{\varnothing P}(P_S^{\varnothing P*}) = \frac{\mu\left(2\mu^2\gamma_S^2(2\gamma_N+1) - 3\mu\gamma_N\gamma_S(2-3\gamma_N) + \gamma_N^2(8-3\gamma_N) + 4\sqrt{3\gamma_N^2-\mu\gamma_N\gamma_S}(\mu\gamma_S-\gamma_N(1-\gamma_N))\right)}{8\gamma_N(\gamma_N+\mu\gamma_S)^2}$. Welfare in S is

$W_S^{\varnothing P}(P_S^{\varnothing P*}) = \frac{4\gamma_N^4 - 8\mu\gamma_N^3(1-\gamma_S) - \mu^2\gamma_N\gamma_S(2-\gamma_S)(-3\gamma_N+\mu\gamma_S) + 4\mu\gamma_N\sqrt{\gamma_N(3\gamma_N-\mu\gamma_S)}(\gamma_N-\mu\gamma_S(1-\gamma_S))}{8\gamma_N^2(\gamma_N+\mu\gamma_S)^2}$.

Compared to parallel trade and no regulation, welfare in N is higher if μ is suffciently high

$(W_N^{\varnothing P}\left(P_S^{\varnothing P*}\right) - W_N = \frac{\mu\left(2\mu^2\gamma_S^2 - \mu\gamma_N\gamma_S(6-\gamma_N) - 3\gamma_N^3 + 4\sqrt{3\gamma_N^2 - \mu\gamma_N\gamma_S}(\mu\gamma_S - \gamma_N(1-\gamma_N))\right)}{8\gamma_N(\gamma_N+\mu\gamma_S)^2} > 0$, if $\mu > \widetilde{\mu_{W_N^{\varnothing P}}}$, where $\widetilde{\mu_{W_N^{\varnothing P}}}$ is the solution to
$f_{W_N^{\varnothing P}}(\widetilde{\mu_{W_N^{\varnothing P}}}) = 2\mu^2\gamma_S^2 - \mu\gamma_N\gamma_S(6-\gamma_N) - 3\gamma_N^3 + 4\sqrt{3\gamma_N^2 - \mu\gamma_N\gamma_S}(\mu\gamma_S - \gamma_N(1-\gamma_N)) = 0)$. Welfare in S is higher if μ is suffciently low
$(W_S^{\varnothing P}\left(P_S^{\varnothing P*}\right) - W_S = \frac{\mu\left(-\mu^2\gamma_S^2(2-\gamma_S) + 3\mu\gamma_N\gamma_S(2-\gamma_S) + 4\sqrt{3\gamma_N^2 - \mu\gamma_N\gamma_S}(\gamma_N - \mu\gamma_S(1-\gamma_S))\right)}{8\gamma_N(\gamma_N+\mu\gamma_S)^2} > 0$ if $\mu < \widetilde{\mu_{W_S^{\varnothing P}}}$, where $\widetilde{\mu_{W_S^{\varnothing P}}}$ is the solution to
$f_{W_S^{\varnothing P}}(\widetilde{\mu_{W_S^{\varnothing P}}}) = -\mu^2\gamma_S^2(2-\gamma_S) + 3\mu\gamma_N\gamma_S(2-\gamma_S) + 4\sqrt{3\gamma_N^2 - \mu\gamma_N\gamma_S}(\gamma_N - \mu\gamma_S(1-\gamma_S)) = 0)$.

Parallel Trade and Price Cap in the North and South (PP)

Under parallel trade and price caps in both countries (PP), the firm's profit is

$$\pi_{N,S}^{PP} = \begin{cases} \frac{1}{\mu}\left(\mu - \left(\gamma_N P_N^{PP}\right)\right)P_N^{PP} + \left(1 - \left(\gamma_S P_S^{PP}\right)\right)P_S^{PP} & \text{if} \quad P_N^{PP} \leq P_S^{PP} \\ \left(\frac{1}{\mu}\left(\mu - \left(\gamma_N P_S^{PP}\right)\right) + \left(1 - \left(\gamma_S P_S^{PP}\right)\right)\right)P_S^{PP} & \text{if} \quad P_N^{PP} > P_S^{PP}. \end{cases}$$

If $P_N^{PP} \leq P_S^{PP}$, the firm always exports to S $(\Delta^{PP}\big|_{P_N^{PP} \leq P_S^{PP}} = \pi_{N,S}^{PP} - \pi_N^{xPP} = P_S^{PP}\left(1 - P_S^{PP}\gamma_S\right) > 0)$. If $P_N^{PP} > P_S^{PP}$, the firm exports to S if the price cap $P_S^{PT,PP}$ is suffciently high
$(\Delta^{PP}\big|_{P_N^{PP} > P_S^{PP}} = \pi_{N,S}^{PP} - \pi_N^{xPP} = \frac{P_S^{PP}\mu\left(1 - P_S^{PP}\gamma_S\right) - \left(P_N^{xPP} - P_S^{PP}\right)\left(\mu - \gamma_N\left(P_S^{PP} + P_N^{xPP}\right)\right)}{\mu} \geq 0,$
if $P_S^{PT,PP} > \widehat{P_{S,\Delta^{PP}}^{PP}} = \frac{\mu - \sqrt{\mu^2 - P_N^{xPP}\left(\mu - \gamma_N P_N^{xPP}\right)(\gamma_N + \mu\gamma_S)}}{\gamma_N + \mu\gamma_S})$. Under a price cap P_N^{xPP*} of zero, the firm exports to the South $(P_N^{xPP*} = 0 \to \widehat{P_{S,\Delta^{PP}}^{PP}}\left(P_N^{xPP*}\right) = 0)$.

If $P_N^{PP} \leq P_S^{PP}$, welfare in N is

$$W_N^{PP} = \underbrace{\frac{\left(\mu - \gamma_N P_N^{PP}\right)^2}{2\mu}}_{CS_N^{PP}} + \underbrace{\frac{1}{\mu}\left(\mu - \gamma_N P_N^{PP}\right)P_N^{PP} + P_S^{PP}\left(1 - \gamma_S P_S^{PP}\right)}_{\pi^{PP}} - \underbrace{(1-\gamma_N)\frac{1}{\mu}\left(\mu - \gamma_N P_N^{PP}\right)P_N^{PP}}_{E_N^{PP}}$$

$= \frac{\mu^2 - \gamma_N^2\left(P_N^{PP}\right)^2}{2\mu} + P_S^{PP}\left(1 - \gamma_S P_S^{PP}\right)$. If $P_N^{PP} > P_S^{PP}$, welfare in N is

$$W_N^{PP} = \underbrace{\frac{\left(\mu - \gamma_N P_S^{PP}\right)^2}{2\mu}}_{CS_N^{PP}} + \underbrace{\frac{1}{\mu}\left(\mu - \gamma_N P_S^{PP}\right)P_S^{PP} + P_S^{PP}\left(1 - \gamma_S P_S^{PP}\right)}_{\pi^{PP}} - \underbrace{(1-\gamma_N)\frac{1}{\mu}\left(\mu - \gamma_N P_N^{PP}\right)P_N^{PP}}_{E_N^{PP}}$$

$= \frac{\mu^2 + 2P_S^{PP}\mu - \left(P_S^{PP}\right)^2\left(\gamma_N^2 + 2\mu\gamma_S\right)}{2\mu}$.

The welfare maximizing price cap in N is $P_N^{PP*} = 0$.

Welfare in S is $W_S^{PP} = \underbrace{\frac{\left(1 - \gamma_S P_S^{PP}\right)^2}{2}}_{CS_S^{PP}} - \underbrace{(1-\gamma_S)\left(1 - \gamma_S P_S^{PP}\right)P_S^{PP}}_{E_S^{PP}} = \frac{\left(1 - \gamma_S P_S^{PP}\right)\left(1 - P_S^{PP}(2-\gamma_S)\right)}{2}$.

The welfare maximizing price cap in S is $P_S^{PP*} = 0$.

For welfare-maximizing price caps, welfare in N is $W_N^{PP}\left(P_N^{PP*}, P_S^{PP*}\right) = \frac{1}{2}\mu$; welfare in S is
$W_S^{PP}\left(P_N^{PP*}, P_S^{PP*}\right) = \frac{1}{2}$.

Compared to parallel trade and no regulation, welfare in N is lower
$(W_N^{PP}(P_N^{PP*}, P_S^{PP*}) - W_N = -\frac{\mu(2-\gamma_N)\gamma_N}{2(\gamma_N+\mu\gamma_S)^2} < 0)$, welfare in S is higher
$(W_S^{PP}(P_N^{PP*}, P_S^{PP*}) - W_S = \frac{\mu\left(\mu\gamma_S^2 + 2\gamma_N\right)}{2(\gamma_N+\mu\gamma_S)^2} > 0)$.

Reimbursement Limits in Both Countries (RR) or the South only ($\varnothing R$)

No Parallel Trade ($\varnothing R$, RR)

Under no parallel trade and a reimbursement limit in the South only ($\varnothing R$), the firm's profit is
$\pi_{N,S}^{x,\varnothing R} = q_N\left(p_N^{x,\varnothing R}\right)p_N^{x,\varnothing R} + q_S\left(p_S^{x,\varnothing R}\right)p_S^{x,\varnothing R}$.

The equilibrium price in N is $p_N^{x,\varnothing R}=\frac{\mu}{2\gamma_N}$, the equilibrium price in S is $p_S^{x,\varnothing R}=\frac{1+r_S^{x,\varnothing R}(1-\gamma_S)}{2}$, with $p_S^{x,\varnothing R}-r_S^{x,\varnothing R}\geq 0$, if $r_S^{x,\varnothing R}\leq \widetilde{r_S^{x,\varnothing R}}=\frac{1}{\gamma_S+1}$.

Welfare in N is $W_N^{x,\varnothing R}=\underbrace{\frac{\mu}{8}}_{CS_N^{x,\varnothing R}}+\underbrace{\frac{\mu}{4\gamma_N}+\frac{\left(1+r_S^{x,\varnothing R}(1-\gamma_S)\right)^2}{4}}_{\pi^{x,\varnothing R}}-\underbrace{\frac{(1-\gamma_N)\mu}{4\gamma_N}}_{E_N^{x,\varnothing R}}=\frac{3\mu}{8}+\frac{\left(1+r_S^{x,\varnothing R}(1-\gamma_S)\right)^2}{4}$.

Under no parallel trade and reimbursement limits in both countries (RR), the firm's profit is $\pi_{N,S}^{x,RR}=\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{x,RR}+\left(p_N^{x,RR}-r_N^{x,RR}\right)\right)\right)p_N^{x,RR}+\left(1-\left(\gamma_S r_S^{x,RR}+\left(p_S^{x,RR}-r_S^{x,RR}\right)\right)\right)p_S^{x,RR}$.

The equilibrium price in N is $p_N^{x,RR}=\frac{\mu+r_N^{x,RR}(1-\gamma_N)}{2}$, with $p_N^{x,RR}-r_N^{x,RR}\geq 0$, if $r_N^{x,RR}\leq \widetilde{r_N^{x,RR}}=\frac{\mu}{\gamma_N+1}$. The equilibrium price in S is $p_S^{x,RR}=\frac{1+r_S^{x,RR}(1-\gamma_S)}{2}$, with $p_S^{x,RR}-r_S^{x,RR}\geq 0$, if $r_S^{x,RR}\leq \widetilde{r_S^{x,RR}}=\frac{1}{\gamma_S+1}$.

Welfare in N is

$$W_N^{x,RR}=\underbrace{\frac{\left(\mu+r_N^{x,RR}(1-\gamma_N)\right)^2}{8\mu}}_{CS_N^{x,RR}}+\underbrace{\frac{\left(\mu+r_N^{x,RR}(1-\gamma_N)\right)^2}{4\mu}+\frac{\left(1+r_S^{x,RR}(1-\gamma_S)\right)^2}{4}}_{\pi_{N,S}^{x,RR}}-\underbrace{\frac{r_N(1-\gamma_N)\left(\mu+r_N^{x,RR}(1-\gamma_N)\right)}{2\mu}}_{E_N^{x,RR}}$$

$$=\frac{3\mu^2+2\mu\left(1+r_S^{x,RR}(1-\gamma_S)\left(2+r_S^{x,RR}(1-\gamma_S)\right)+r_N^{x,RR}(1-\gamma_N)\right)-\left(r_N^{x,RR}\right)^2(1-\gamma_N)^2}{8\mu}.$$

The welfare-maximizing reimbursement limit in N is $r_N^{x,RR*}=\widetilde{r_N^{x,RR}}=\frac{\mu}{1+\gamma_N}$.

For both cases, welfare in S is $W_S^{x,(.)R}=\underbrace{\frac{\left(1+r_S^{x,(.)R}(1-\gamma_S)\right)^2}{8}}_{CS_S^{x,(.)R}}-\underbrace{\frac{r_S^{x,(.)R}(1-\gamma_S)\left(1+r_S^{x,(.)R}(1-\gamma_S)\right)}{2}}_{E_S^{x,(.)R}}$

$$=\frac{\left(1+r_S^{x,(.)R}(1-\gamma_S)\right)\left(1-3r_S^{x,(.)R}(1-\gamma_S)\right)}{8}.$$

The welfare-maximizing reimbursement limit in S is $r_S^{x,(.)R*}=0$.

For a welfare-maximizing reimbursement limit in the South ($\varnothing R$), welfare in N is $W_N^{x,\varnothing R}\left(r_N^{x,\varnothing R*},r_S^{x,\varnothing R*}\right)=\frac{3}{8}\mu+\frac{1}{4}$.

For welfare-maximizing reimbursement limits in both countries (RR), welfare in N is $W_N^{x,RR*}\left(r_N^{x,RR*},r_S^{x,RR*}\right)=\frac{2\mu(2\gamma_N+1)+\gamma_N(\gamma_N+2)+1}{4(\gamma_N+1)^2}$. In both cases, welfare in S is $W_S^{x,(.)R}\left(r_N^{x,RR*},r_S^{x,(.)R*}\right)=\frac{(3\gamma_S-2)}{8}$.

Compared to no parallel trade and no regulation, welfare in N is lower under a reimbursement limit in the South ($\varnothing R$) ($W_N^{x,\varnothing R}\left(r_N^{x,\varnothing R*},r_S^{x,\varnothing R*}\right)-W_N^x=-\frac{(1-\gamma_S)}{4\gamma_S}<0$). Welfare in N is lower under reimbursement limits in both countries (RR) if μ is sufficiently low ($W_N^{x,RR}\left(r_N^{x,RR*},r_S^{x,(.)R*}\right)-W_N^x=-\frac{2(\gamma_N+1)^2(1-\gamma_S)-\mu\gamma_S(3\gamma_N+1)(1-\gamma_N)}{8\gamma_S(\gamma_N+1)^2}<0$, $\mu<\widetilde{\mu_{W_N^{xRR}}}=\frac{2(\gamma_N+1)^2(1-\gamma_S)}{\gamma_S(1-\gamma_N)(3\gamma_N+1)}$). Welfare in S is higher under a reimbursement limit in the South ($\varnothing R$) and under reimbursement limits in both countries (RR) ($W_S^{x,(.)R}\left(r_N^{x,RR*},r_S^{x,(.)R*}\right)-W_S^x=\frac{\left(3\gamma_S^2-5\gamma_S+2\right)}{8\gamma_S}>0$).

Parallel Trade and Reimbursement Limit in S ($\varnothing R$)

Under parallel trade and a reimbursement limit in the South only ($\varnothing R$), the firm's profit is $\pi_{N,S}^{\varnothing R}=\frac{1}{\mu}\left(\mu-\gamma_N p_{N,S}^{\varnothing R}\right)p_{N,S}^{\varnothing R}+\left(1-\left(\gamma_S r_S^{\varnothing R}+\left(p_{N,S}^{\varnothing R}-r_S^{\varnothing R}\right)\right)\right)p_{N,S}^{\varnothing R}$.

The equilibrium uniform price is $p_{N,S}^{\varnothing R}=\frac{\mu\left(2+r_S^{\varnothing R}(1-\gamma_S)\right)}{2(\gamma_N+\mu)}$, with $p_S^{\varnothing R}-r_S^{\varnothing R}\geq 0$ if $r_S\leq \widetilde{r_S^{\varnothing R}}=\frac{2\mu}{2\gamma_N+\mu(1+\gamma_S)}$. The equilibrium profit is $\pi_{N,S}^{\varnothing R}=\frac{\mu\left(r_S^{\varnothing R}(1-\gamma_S)+2\right)^2}{4(\gamma_N+\mu)}$.

The firm exports to S if μ is sufficiently high or $r_S^{\varnothing R}$ is sufficiently high ($\Delta^{\varnothing R}=\pi_{N,S}^{\varnothing R}-\pi_N^{x,\varnothing R}=\frac{\mu\left(3\gamma_N-\mu+\gamma_N r_S^{\varnothing R}(1-\gamma_S)\left(4+r_S^{\varnothing R}(1-\gamma_S)\right)\right)}{4\gamma_N(\gamma_N+\mu)}>0$, if $\mu<\widehat{\mu_{\Delta^R}}=3\gamma_N+\gamma_N r_S^{\varnothing R}(1-\gamma_S)\left(4+r_S^{\varnothing R}(1-\gamma_S)\right)\vee r_S^{\varnothing R}>\widehat{r_{S,\Delta^{\varnothing R}}}=\frac{\sqrt{\gamma_N(\mu+\gamma_N)}-2\gamma_N}{\gamma_N(1-\gamma_S)}$).

Welfare in N is

$$W_N^{\varnothing R} = \underbrace{\frac{\mu\left(2\mu - r_S^{\varnothing R}\gamma_N\left(1-\gamma_S\right)\right)^2}{8\left(\mu+\gamma_N\right)^2}}_{CS_N^{\varnothing R}} + \underbrace{\frac{\mu\left(r_S^{\varnothing R}\left(1-\gamma_S\right)+2\right)^2}{4\left(\mu+\gamma_N\right)}}_{\pi^{\varnothing R}} - \underbrace{\frac{\left(1-\gamma_N\right)\mu\left(2\mu-\gamma_N r_S^{\varnothing R}\left(1-\gamma_S\right)\right)\left(2+r_S^{\varnothing R}\left(1-\gamma_S\right)\right)}{4\left(\mu+\gamma_N\right)^2}}_{E_N^{\varnothing R}}$$
$$= \frac{4\mu^3+2\mu^2\left(4\gamma_N+2r_S^{\varnothing R}(1-\gamma_S)+\left(r_S^{\varnothing R}\right)^2(1-\gamma_S)^2\right)+\mu\gamma_N r_S^{\varnothing R}(1-\gamma_S)\left(4(3-\gamma_N)+r_S^{\varnothing R}(1-\gamma_S)(4-\gamma_N)\right)+8\mu\gamma_N}{8(\mu+\gamma_N)^2}.$$

Welfare in S is

$$W_S^{\varnothing R} = \underbrace{\frac{\left(2\gamma_N + r_S^{\varnothing R}\left(1-\gamma_S\right)\left(\mu+2\gamma_N\right)\right)^2}{8\left(\mu+\gamma_N\right)^2}}_{CS_S^{\varnothing R}} - \underbrace{\frac{r_S^{\varnothing R}\left(1-\gamma_S\right)\left(2\gamma_N + r_S^{\varnothing R}\left(1-\gamma_S\right)\left(\mu+2\gamma_N\right)\right)}{2\left(\mu+\gamma_N\right)}}_{E_S^{\varnothing R}}$$
$$= \frac{\left(2\gamma_N+r_S^{\varnothing R}(1-\gamma_S)(2\gamma_N+\mu)\right)\left(2\gamma_N-r_S^{\varnothing R}(1-\gamma_S)(2\gamma_N+3\mu)\right)}{8(\gamma_N+\mu)^2}.$$

Consider two cases: i) For $\mu < \widetilde{\mu^{\varnothing R}} = 3\gamma_N$, the welfare-maximizing reimbursement limit in S is $r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^{\varnothing R}}} = 0$, which is the same as under no parallel trade ($r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^{\varnothing R}}} - r_S^{x,(.)R*} = 0$). ii) For $\mu > \widetilde{\mu^{\varnothing R}}$, the welfare-maximizing reimbursement limit in S is $r_S^{\varnothing R*}\big|_{\mu>\widetilde{\mu^{\varnothing R}}} = \frac{\sqrt{\gamma_N(\mu+\gamma_N)}-2\gamma_N}{\gamma_N(1-\gamma_S)}$, which higher than under no parallel trade ($r_S^{\varnothing R*}\big|_{\mu>\widetilde{\mu^R}} - r_S^{x,(.)R*} = \frac{\sqrt{\gamma_N(\mu+\gamma_N)}-2\gamma_N}{\gamma_N(1-\gamma_S)} > 0$).

i) For $\mu < \widetilde{\mu^{\varnothing R}}$, welfare in N is $W_N^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^R}}\right) = \frac{\mu\left(4\mu^2+8\gamma_N\mu+8\gamma_N\right)}{8(\mu+\gamma_N)^2}$. For ii) $\mu > \widetilde{\mu^{\varnothing R}}$, welfare in N is $W_N^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu>\widetilde{\mu^R}}\right) = \frac{2\mu^2(2\gamma_N+1)+\mu\gamma_N(3\gamma_N+4)-4\mu\sqrt{\gamma_N(\mu+\gamma_N)}}{8\gamma_N(\mu+\gamma_N)}$.

i) For $\mu < \widetilde{\mu^{\varnothing R}}$, welfare in S is $W_S^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^R}}\right) = \frac{\gamma_N^2}{2(\mu+\gamma_N)^2}$. For ii) $\mu > \widetilde{\mu^{\varnothing R}}$, welfare in S is $W_S^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu>\widetilde{\mu^R}}\right) = \frac{4\sqrt{\gamma_N(\mu+\gamma_N)}(3\mu+4\gamma_N)-16\gamma_N^2-\mu(3\mu+20\gamma_N)}{8\gamma_N(\mu+\gamma_N)}$.

Compared to parallel trade and no regulation, i) welfare in N is lower for $\mu < \widetilde{\mu^{\varnothing R}} = 3\gamma_N$ ($W_N^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^R}}\right) - W_N = -\frac{\mu^2\gamma_N(2-\gamma_N)(1-\gamma_S)(\mu+2\gamma_N+\mu\gamma_S)}{2(\gamma_N+\mu\gamma_S)^2(\mu+\gamma_N)^2} < 0$) and ii) welfare in N is lower for $\mu > \widetilde{\mu^{\varnothing R}}$ and if μ is a sufficiently low ($W_N^{\varnothing R}\left(r_S^{R*}\big|_{\mu>\widetilde{\mu^R}}\right) - W_N$
$= -\frac{\mu\left(4\sqrt{\gamma_N^2+\mu\gamma_N}(\gamma_N+\mu\gamma_S)^2-2\mu^3\gamma_S^2-\mu^2\gamma_N\gamma_S(4+4\gamma_S-\gamma_N\gamma_S)+2\mu\gamma_N^2(-2\gamma_N-4\gamma_S+\gamma_N\gamma_S+3)+\gamma_N^3(4-3\gamma_N)\right)}{8\gamma_N(\gamma_N+\mu\gamma_S)^2(\mu+\gamma_N)} < 0$ if $\mu < \widetilde{\mu_{W_N^{\varnothing R}}}$, where $\widetilde{\mu_{W_N^{\varnothing R}}}$ is the solution to $f_{W_N^{\varnothing R}}(\widetilde{\mu_{W_N^{\varnothing R}}}) = 4\sqrt{\gamma_N^2+\mu\gamma_N}\left(\gamma_N+\mu\gamma_S\right)^2 - 2\mu^3\gamma_S^2 - \mu^2\gamma_N\gamma_S\left(4+4\gamma_S-\gamma_N\gamma_S\right)$
$+2\mu\gamma_N^2\left(-2\gamma_N-4\gamma_S+\gamma_N\gamma_S+3\right)+\gamma_N^3\left(4-3\gamma_N\right) = 0$). Welfare in S is higher
($W_S^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu<\widetilde{\mu^{\varnothing R}}}\right) - W_S = \frac{\gamma_N(1-\gamma_S)\mu^2(2\mu+3\gamma_N-\gamma_N\gamma_S)}{2(\gamma_N+\mu\gamma_S)^2(\mu+\gamma_N)^2} > 0$, $W_S^{\varnothing R}\left(r_S^{\varnothing R*}\big|_{\mu>\widetilde{\mu^{\varnothing R}}}\right) - W_S$
$= \frac{4\sqrt{\gamma_N^2+\mu\gamma_N}(3\mu+4\gamma_N)(\gamma_N+\mu\gamma_S)^2-3\mu^4\gamma_S^2-2\mu^3\gamma_N\gamma_S(10\gamma_S+3)+\mu^2\gamma_N^2\left(5-48\gamma_S-16\gamma_S^2\right)-8\mu\gamma_N^3(5\gamma_S+2)-20\gamma_N^4}{8\gamma_N(\gamma_N+\mu\gamma_S)^2(\mu+\gamma_N)} > 0$).

Parallel Trade and Reimbursement Limits in N and S (RR)

Under parallel trade and reimbursement limits in both countries, the firm's profit is

$\pi_{N,S}^{RR} = \frac{1}{\mu}\left(\mu - \left(\gamma_N r_N^{RR} + \left(p_{NS}^{RR} - r_N^{RR}\right)\right)\right)p_{NS}^{RR} + \left(1 - \left(\gamma_S r_S^{RR} + \left(p_{NS}^{RR} - r_S^{RR}\right)\right)\right)p_{NS}^{RR}$.

The equilibrium uniform price is $p_{NS}^{RR} = \frac{2\mu+\mu r_S^{RR}(1-\gamma_S)+r_N^{RR}(1-\gamma_N)}{2(\mu+1)}$, with $p_N^{RR} - r_N \geq 0$,

if $r_N \leq \widetilde{r_N^{RR}} = \frac{\mu\left(2+r_S^{RR}(1-\gamma_S)\right)}{\gamma_N+2\mu+1}$ and $p_S^{RR} - r_S \geq 0$, if $r_S < \widetilde{r_S^{RR}} = \frac{2\mu+r_N^{RR}(1-\gamma_N)}{\mu+\mu\gamma_S+2}$. The equilibrium profit is $\pi_{N,S}^{RR} = \frac{\left(2\mu+r_N^{RR}(1-\gamma_N)+\mu r_S^{RR}(1-\gamma_S)\right)^2}{4\mu(\mu+1)}$.

The firm exports to S if μ is sufficiently low or r_S^{RR} is sufficiently high

($\Delta^{RR} = \pi_{N,S}^{RR} - \pi_N^{x,RR} = \frac{\left(2\mu+r_N^{RR}(1-\gamma_N)+\mu r_S^{RR}(1-\gamma_S)\right)^2-(\mu+1)\left(\mu+r_N^{x,RR}(1-\gamma_N)\right)^2}{4\mu(\mu+1)} > 0$, if $\mu < \widehat{\mu_{\Delta^{RR}}}$, with $\widehat{\mu_{\Delta^{RR}}}$ as the solution to $f_{\widehat{\mu_{\Delta^{RR}}}}\left(\widehat{\mu_{\Delta^{RR}}}\right) = \left(2\mu + r_N^{RR}\left(1-\gamma_N\right) + \mu r_S^{RR}\left(1-\gamma_S\right)\right)^2 - \left(\mu+1\right)\left(\mu + r_N^{RR}\left(1-\gamma_N\right)\right)^2$;

$\Delta^{RR} = \pi_{N,S}^{RR} - \pi_N^{x,RR} > 0$, if $r_S^{RR} > \widehat{r_{S,\Delta^{RR}}} = \frac{\left(\sqrt{\mu+1}\left(r_N^{x,RR}(1-\gamma_N)+\mu\right)-r_N^{RR}(1-\gamma_N)-2\mu\right)}{\mu(1-\gamma_S)}$). For $r_N^{x,RR*} = \frac{\mu}{1+\gamma_N}$, the critical size of μ is $\widehat{\mu_{\Delta^{RR}}}\left(r_N^{x,RR*}\right)$ as the solution to

$f_{\widehat{\mu_{\Delta^{RR}}}}\left(\widehat{\mu_{\Delta^{RR}}}, r_N^{RR*}\right) = \left(2\mu + r_N^{RR}\left(1-\gamma_N\right) + \mu r_S^{RR}\left(1-\gamma_S\right)\right)^2 - \left(\mu+1\right)\left(\mu + \frac{\mu(1-\gamma_N)}{1+\gamma_N}\right)^2 = 0$, and the critical reimbursement limit $\widehat{r_{S,\Delta^{RR}}}$ is $\widehat{r_{S,\Delta^{RR}}}\left(r_N^{x,RR*}\right) = \frac{\left(2\sqrt{\mu+1}\frac{\mu}{\gamma_N+1} - r_N^{RR}(1-\gamma_N)-2\mu\right)}{\mu(1-\gamma_S)}$).

Welfare in N is $W_N^{RR} = \underbrace{\frac{\left(2\mu^2 + r_N^{RR}(1-\gamma_N)(2\mu+1) - \mu r_S^{RR}(1-\gamma_S)\right)^2}{8\mu(\mu+1)^2}}_{CS_N^{RR}} +$

$\underbrace{\frac{\left(2\mu + r_N^{RR}(1-\gamma_N) + \mu r_S^{RR}(1-\gamma_S)\right)^2}{4\mu(\mu+1)}}_{\pi^{RR}} - \underbrace{\frac{r_N^{RR}(1-\gamma_N)\left(2\mu^2 + r_N^{RR}(1-\gamma_N)(2\mu+1) - \mu r_S^{RR}(1-\gamma_S)\right)}{2\mu(\mu+1)}}_{E_N^{RR}}$

$= \frac{4\mu^2\left(2\mu+\mu^2+2\right)+4\mu r_N^{RR}(1-\gamma_N)(\mu+2)-\left(r_N^{RR}\right)^2(1-\gamma_N)^2\left(6\mu+4\mu^2+1\right)+4\mu^2 r_S^{RR}(1-\gamma_S)(\mu+2)+\mu^2\left(r_S^{RR}\right)^2(1-\gamma_S)^2(2\mu+3)+2\mu r_N r_S^{RR}(1-\gamma_S)(1-\gamma_N)(2\mu+3)}{8\mu(\mu+1)^2}$.

Welfare in S is $W_S^{RR} = \underbrace{\frac{\left(2 - r_N^{RR}(1-\gamma_N) + r_S^{RR}(1-\gamma_S)(\mu+2)\right)^2}{8(\mu+1)^2}}_{CS_S^{RR}} - \underbrace{\frac{r_S^{RR}(1-\gamma_S)\left(2 - r_N^{RR}(1-\gamma_N) + r_S^{RR}(1-\gamma_S)(\mu+2)\right)}{2(\mu+1)}}_{E_S^{RR}}$

$= \frac{4-4r_N(1-\gamma_N)+\left(r_N^{RR}\right)^2(1-\gamma_N)^2-4\mu r_S^{RR}(1-\gamma_S)-\left(r_S^{RR}\right)^2(1-\gamma_S)^2(\mu+2)(3\mu+2)+2\mu r_N^{RR} r_S^{RR}(1-\gamma_S)(1-\gamma_N)}{8(\mu+1)^2}$.

Best response functions are $r_N^{RR}\left(r_S^{RR}\right) = \mu\frac{4+2\mu+r_S^{RR}(1-\gamma_S)(2\mu+3)}{\left(6\mu+4\mu^2+1\right)(1-\gamma_N)}$, $r_S^{RR}\left(r_N^{RR}\right) = -\mu\frac{2-r_N^{RR}(1-\gamma_N)}{(\mu+2)(3\mu+2)(1-\gamma_S)}$.

Consider two cases: i) For $\mu < \widetilde{\mu^{RR}}$, the welfare-maximizing reimbursement limit in N is

$r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} = \frac{2\mu(\mu+2)}{(1-\gamma_N)\left(6\mu+4\mu^2+1\right)}$ and the welfare-maximizing reimbursement limit in S is $r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} = 0$.

ii) For $\mu > \widetilde{\mu^{RR}}$, the welfare-maximizing reimbursement limit in N is $r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}} = \mu\frac{\sqrt{\mu+1}(2\mu+3)-(\gamma_N+1)(\mu+1)}{2\left(1-\gamma_N^2\right)(\mu+1)^2}$

and the welfare-maximizing reimbursement limit in S is $r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}} = \frac{\sqrt{\mu+1}\left(6\mu+4\mu^2+1\right)-(\mu+1)(4\mu+3)(\gamma_N+1)}{2(\mu+1)^2(1-\gamma_S)(\gamma_N+1)} > 0$.

$\widetilde{\mu^{RR}}$ is the solution to

$f_{\widetilde{\mu^{RR}}}\left(\widetilde{\mu^{RR}}\right) = -16\mu^4 - 16\mu^3\left(2-\gamma_N(\gamma_N+2)\right) - 4\mu^2\left(1-10\gamma_N(\gamma_N+2)\right)$

$+3\mu\left(7+22\gamma_N+11\gamma_N^2\right) + (3\gamma_N+4)(3\gamma_N+2) = 0.$

Compared to no parallel trade and reimbursement limits in both countries, the reimbursement limit in N is higher ($r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} - r_N^{x,RR*} = -\frac{\mu\left(4\mu^2(1-\gamma_N)+4\mu(1-2\gamma_N)-3-5\gamma_N\right)}{\left(1-\gamma_N^2\right)\left(4\mu^2+6\mu+1\right)} < 0$,

$r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}} - r_N^{x,RR*} = -\frac{\mu\left(3-\gamma_N-\sqrt{\mu+1}(2\mu+3)+2\mu^2(1-\gamma_N)+\mu(5-3\gamma_N)\right)}{2(\mu+1)^2\left(1-\gamma_N^2\right)} < 0$). The reimbursement limit in S is the same for $\mu < \widetilde{\mu^{RR}}$ and higher for $\mu > \widetilde{\mu^{RR}}$ ($r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}} - r_S^{x,RR*} = 0$,

$r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}} - r_S^{xR*} = \frac{\sqrt{\mu+1}\left(6\mu+4\mu^2+1\right)-(\mu+1)(4\mu+3)(\gamma_N+1)}{2(\mu+1)^2(1-\gamma_S)(\gamma_N+1)} > 0$).

For i) $\mu < \widetilde{\mu^{RR}}$, welfare in N is $W_N^{RR}\left(r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}\right) = \mu\frac{3\mu+2\mu^2+3}{6\mu+4\mu^2+1}$. For ii) $\mu > \widetilde{\mu^{RR}}$, welfare in N is $W_N^{RR}\left(r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}\right)$

$= \frac{3\mu\left(2\gamma_N+\gamma_N^2+2\right)+\mu^2\left(22\gamma_N+11\gamma_N^2+31\right)+12\mu^3\left(2\gamma_N+\gamma_N^2+3\right)+4\mu^4\left(2\gamma_N+\gamma_N^2+3\right)-2\sqrt{\mu+1}\mu(\gamma_N+1)\left(4\mu^2+6\mu+1\right)}{8(\mu+1)^3(\gamma_N+1)^2}$. For $\mu < \widetilde{\mu^{RR}}$, welfare in S is $W_S^{RR}\left(r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}\right) = \frac{(3\mu+1)^2}{2\left(6\mu+4\mu^2+1\right)^2}$. For $\mu > \widetilde{\mu^{RR}}$, welfare in S is

$W_S^{RR}\left(r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}\right)$

$= \frac{2\sqrt{\mu+1}(\gamma_N+1)\left(12\mu^3+28\mu^2+18\mu+3\right)-\left(10\gamma_N+5\gamma_N^2+6\right)-3\mu\left(14\gamma_N+7\gamma_N^2+11\right)-4\mu^2\left(14\gamma_N+7\gamma_N^2+17\right)-4\mu^3\left(6\gamma_N+3\gamma_N^2+13\right)-12\mu^4}{8(\mu+1)^3(\gamma_N+1)^2}$.

Compared to parallel trade and no regulation, welfare in N is lower for

i) $\mu < \widetilde{\mu^{RR}}$ ($W_N^{RR}\left(r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}\right) - W_N$

$= -\frac{\mu\left(2\mu\gamma_N(6-5\gamma_S-3\gamma_N)+2\gamma_N(1-3\gamma_N)-\mu^2\left(-8\gamma_N+4\gamma_N^2+5\gamma_S^2\right)\right)}{2(\gamma_N+\mu\gamma_S)^2\left(4\mu^2+6\mu+1\right)} < 0$). Welfare is lower for ii) $\mu > \widetilde{\mu^{RR}}$ and if μ is sufficiently low ($W_N^{RR}\left(r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}\right) - W_N = \frac{\Omega_{\Delta W_N^{RR}}}{8(\gamma_N+\mu\gamma_S)^2(\mu+1)^3(\gamma_N+1)^2} < 0$, if $\mu < \widetilde{\mu_{W_N^{RR}}}$, with

$\Omega_{\Delta W_N^{RR}} = -2\mu\sqrt{\mu+1}\left(4\mu^2+6\mu+1\right)(\gamma_N+1)(\gamma_N+\mu\gamma_S)^2 + 8\mu^6\gamma_S^2 + 8\mu^5\gamma_S(2\gamma_N+3\gamma_S)$

$+\mu^4\left(-8\gamma_N - 4\gamma_N^2 + 4\gamma_N^4 + 19\gamma_S^2 - \gamma_N^2\gamma_S^2 + 48\gamma_N\gamma_S - 2\gamma_N\gamma_S^2\right)$

$+\mu^3\left(-24\gamma_N - 12\gamma_N^2 + 12\gamma_N^4 + 2\gamma_S^2 - \gamma_N^2\gamma_S^2 + 38\gamma_N\gamma_S - 2\gamma_N\gamma_S^2 - 4\gamma_N^2\gamma_S - 2\gamma_N^3\gamma_S\right)$

$+\mu^2\gamma_N\left(-17\gamma_N + 4\gamma_S - 2\gamma_N^2 + 11\gamma_N^3 - 4\gamma_N\gamma_S - 2\gamma_N^2\gamma_S - 24\right) + \gamma_N\mu\left(-10\gamma_N - 2\gamma_N^2 + 3\gamma_N^3 - 8\right)$, where $\widetilde{\mu_{W_N^{RR}}}$ is the solution to $f_{W_N^{RR}}(\widetilde{\mu_{W_N^{RR}}}) = \Omega_{\Delta W_N^{RR}} = 0$). Welfare in S is higher for

i) $\mu < \widetilde{\mu^{RR}}$ ($W_S^{RR}\left(r_N^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu<\widetilde{\mu^{RR}}}\right) - W_S = \frac{\Omega_{\Delta W_S^{RR},1}}{2(\gamma_N+\mu\gamma_S)^2\left(4\mu^2+6\mu+1\right)^2} > 0$,

with $\Omega_{\Delta W_S^{RR},1} = 32\mu^5\gamma_N(1-\gamma_S)+\mu^4\left(96\gamma_N-16\gamma_N^2+9\gamma_S^2-96\gamma_N\gamma_S\right)-2\mu^3\left(-44\gamma_N+24\gamma_N^2-3\gamma_S^2+35\gamma_N\gamma_S\right)$
$+\mu^2\left(24\gamma_N-35\gamma_N^2+\gamma_S^2-12\gamma_N\gamma_S\right)-2\mu\gamma_N(3\gamma_N-1))$, welfare in S is higher for ii) $\mu > \widetilde{\mu^{RR}}$ and if μ is sufficiently high ($W_S^{RR}\left(r_N^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}, r_S^{RR*}\big|_{\mu>\widetilde{\mu^{RR}}}\right) - W_S = -\frac{\Omega_{\Delta W_S^{RR},2}}{8(\gamma_N+\mu\gamma_S)^2(\mu+1)^3(\gamma_N+1)^2} > 0$,
with $\Omega_{\Delta W_S^{RR},2} = 2\sqrt{\mu+1}(\gamma_N+1)(\gamma_N+\mu\gamma_S)^2\left(12\mu^3+28\mu^2+18\mu+3\right)$
$-12\mu^6\gamma_S^2-4\mu^5\gamma_S\left(6\gamma_N+13\gamma_S+3\gamma_N\gamma_S(\gamma_N+2)\right)$
$-4\mu^4\left(17\gamma_S^2-\gamma_N\left(\gamma_N+2\gamma_N^2+2\right)+\gamma_N\gamma_S\left(16\gamma_N+14\gamma_S+8\gamma_N^2+7\gamma_N\gamma_S+28\right)\right)$
$-\mu^3\left(33\gamma_S^2+8\gamma_N\left(\gamma_N+\gamma_N^2+2\gamma_N^3-3\right)+\gamma_N\gamma_S\left(160\gamma_N+42\gamma_S+80\gamma_N^2+21\gamma_N\gamma_S+160\right)\right)$
$-\mu^2\left(8\gamma_N\left(4\gamma_N+7\gamma_N^2+5\gamma_N^3-3\right)+6\gamma_S^2+\gamma_N\gamma_S\left(132\gamma_N+10\gamma_S+66\gamma_N^2+5\gamma_N\gamma_S+90\right)\right)$
$-\mu\gamma_N\left(\gamma_N\left(58\gamma_N+33\gamma_N^2+29\right)+20\gamma_S+18\gamma_N\gamma_S(\gamma_N+2)-8\right)-\gamma_N^2\left(18\gamma_N+9\gamma_N^2+10\right)$, where $\widetilde{\mu_{W_S^{RR}}}$ is the solution to $f_{W_S^{RR}}(\widetilde{\mu_{W_S^{RR}}}) = \Omega_{\Delta W_S^{RR},2} = 0$).

"Asymmetric" Equilibria

No Parallel Trade, Price Cap in the North, Reimbursement Limit in the South (x, PR)

Under no parallel trade, a price cap in the North and a reimbursement limit in the South, the firm's profit is $\pi_{N,S}^{x,PR} = \pi_{N,S}^{x,PR} = q_N\left(P_N^{x,PR}\right)P_N^{x,PR} + q_S\left(p_S^{x,PR}\right)p_S^{x,PR}$.

The equilibrium price in S is $p_S^{x,PR} = \frac{1+r_S^{x,PR}(1-\gamma_S)}{2}$, with $p_S^{x,PR} - r_S^{x,PR} \geq 0$, if $r_S^{x,PR} \leq \widetilde{r_S^{x,PR}} = \frac{1}{\gamma_S+1}$.

Welfare in N is $W_N^{x,PR}$

$$= \underbrace{\frac{\left(\mu-\gamma_N P_N^{x,PR}\right)^2}{2\mu}}_{CS_N^{x,PR}} + \underbrace{\frac{1}{\mu}\left(\mu-\gamma_N P_N^{x,PR}\right)P_N^{x,PR} + \frac{\left(1+r_S^{x,PR}(1-\gamma_S)\right)^2}{4}}_{\pi^{x,PR}} - \underbrace{(1-\gamma_N)\frac{1}{\mu}\left(\mu-\gamma_N P_N^{x,PR}\right)P_N^{x,PR}}_{E_N^{x,PR}}$$

$$= \frac{\mu^2-\gamma_N^2\left(P_N^{x,PR}\right)^2}{2\mu} + \frac{\left(1+r_S^{x,PR}(1-\gamma_S)\right)^2}{4}.$$

The welfare maximizing price cap in N is $P_N^{x,PR*} = 0$.

Welfare in S is $W_S^{x,PR} = \underbrace{\frac{\left(1+r_S^{x,PR}(1-\gamma_S)\right)^2}{8}}_{CS_S^{x,PR}} - \underbrace{\frac{r_S(1-\gamma_S)\left(1+r_S^{x,PR}(1-\gamma_S)\right)}{2}}_{E_S^{x,PR}} = \frac{\left(1+r_S^{x,PR}(1-\gamma_S)\right)\left(1-3r_S^{x,PR}(1-\gamma_S)\right)}{8}$.

The welfare maximizing rembursement limit in S is $r_S^{x,PR*} = 0$.

For the welfare-maximizing price cap in N and the welfare-maximizing reimbursement limit in S, welfare in N is $W_N^{x,PR}\left(P_N^{x,PR*}, r_S^{x,PR*}\right) = \frac{1}{2}\mu + \frac{1}{4}$, welfare in S is $W_S^{x,PR}\left(P_N^{x,PR*}, r_S^{x,PR*}\right) = \frac{1}{8}$.

Parallel Trade and Price Cap in the North, Reimbursement Limit in the South (PR)

Under parallel trade, a price cap in the North and a reimbursement limit in the South, the firm's profit is $\pi_{N,S}^{PR} = \frac{1}{\mu}\left(\mu-\gamma_N P_N^{PR,P}\right)P_N^{PR,P} + \left(1-\left(\gamma_S r_S^{RR} + \left(p_S^{PR}-r_S^{RR}\right)\right)\right)p_S^{PR}$.

The equilibrium price in S is $p_S^{PR} = \frac{1+r_S^{PR}(1-\gamma_S)}{2}$, with $p_S^{PR} - r_S^{PR} \geq 0$, if $r_S^{PR} \leq \widetilde{r_S^{PR}} = \frac{1}{\gamma_S+1}$.

The firm's profit is $\pi_{N,S}^{PR} = \frac{1}{\mu}\left(\mu-\gamma_N P_N^{PR}\right)P_N^{PR} + \left(1-\left(\gamma_S r_S^{PR} + p_S^{PR} - r_S^{PR}\right)\right)p_S^{PR}$.

The firm exports to S ($\Delta^{PR} = \pi_{N,S}^{PR} - \pi_N^{xR} = \frac{1}{\mu}\left(\mu-\gamma_N P_N^{PR}\right)P_N^{PR} + \frac{1}{4}\left(r_S^{PR}(1-\gamma_S)+1\right)^2 - \frac{1}{\mu}\left(\mu-\gamma_N P_N^{x,PR}\right)P_N^{x,PR}$).

Welfare in N is $W_N^{PR} = \underbrace{\frac{\left(\mu-\gamma_N P_N^{PR}\right)^2}{2\mu}}_{CS_N^{PR}} + \underbrace{\frac{1}{\mu}\left(\mu-\gamma_N P_N^{PR}\right)P_N^{PR} + \frac{\left(1+r_S^{PR}(1-\gamma_S)\right)^2}{4}}_{\pi^{PR}} - \underbrace{(1-\gamma_N)\frac{1}{\mu}\left(\mu-\gamma_N P_N^{PR}\right)P_N^{PR}}_{E_N^{PR}}$

$$= \frac{\mu^2-\gamma_N^2\left(P_N^{PR}\right)^2}{2\mu} + \frac{\left(1+r_S^{PR}(1-\gamma_S)\right)^2}{4}.$$

The welfare maximizing price cap in N is $P_N^{PR*}=0$.

Welfare in S is $W_S^{PR}=\underbrace{\frac{\left(1+r_S^{PR}(1-\gamma_S)\right)^2}{8}}_{CS_S^{PR}}-\underbrace{\frac{r_S^{PR}(1-\gamma_S)\left(1+r_S^{PR}(1-\gamma_S)\right)}{2}}_{E_S^{PR}}=\frac{\left(1+r_S^{PR}(1-\gamma_S)\right)\left(1-3r_S^{PR}(1-\gamma_S)\right)}{8}$.

The welfare maximizing reimbursement limit in S is $r_S^{PR*}=0$.

For the welfare-maximizing price cap in N and the welfare-maximizing reimbursement limit in S, welfare in N is $W_N^{PR}(P_N^{PR},r_S^{PR})=\frac{2\mu+1}{4}$, welfare in S is $W_S^{PR}(P_N^{PR},r_S^{PR})=\frac{1}{8}$.

No Parallel Trade, Reimbursement Limit in the North, Price Cap in the South (RP)

Under no parallel trade, a reimbursement limit in the North and a price cap in the South, the firm's profit is $\pi_{N,S}^{x,PR}=q_N\left(p_N^{x,RP}\right)P_N^{x,RP}+q_S\left(P_S^{x,RP}\right)P_S^{x,RP}$.

The equilibrium price in N is $p_N^{x,RP}=\frac{\mu+r_N^{x,RP}(1-\gamma_N)}{2}$, with $p_N^{x,RP}-r_N^{x,RP}\geq 0$, if $r_N^{x,RP}\leq\widetilde{r_N^{x,RP}}=\frac{\mu}{\gamma_N+1}$.

Welfare in N is $W_N^{x,RP}$

$=\underbrace{\frac{\left(\mu+r_N^{x,RP}(1-\gamma_N)\right)^2}{8\mu}}_{CS_N^{x,RP}}+\underbrace{\frac{\left(\mu+r_N^{x,RP}(1-\gamma_N)\right)^2}{4\mu}+\left(1-\gamma_S P_S^{x,RP}\right)P_S^{x,RP}}_{\pi_{N,S}^{x,RP}}-\underbrace{\frac{r_N^{x,RP}(1-\gamma_N)\left(\mu+r_N^{x,RP}(1-\gamma_N)\right)}{2\mu}}_{E_N^{x,RP}}$

$=\frac{3\mu^2+2\mu r_N^{x,RP}(1-\gamma_N)-\left(r_N^{x,RP}\right)^2(1-\gamma_N)^2}{8\mu}+\left(1-\gamma_S P_S^{x,RP}\right)P_S^{x,RP}$.

The welfare-maximizing reimbursement limit in N is $r_N^{x,RP*}=\widetilde{r_N^{x,RP}}=\frac{\mu}{1+\gamma_N}$.

Welfare in S is $W_S^{x,RP}=\underbrace{\frac{\left(1-\gamma_S P_S^{x,RP}\right)^2}{2}}_{CS_S^{x,RP}}-\underbrace{(1-\gamma_S)\left(1-\gamma_S P_S^{x,RP}\right)P_S^{x,RP}}_{E_S^{x,RP}}=\frac{\left(1-\gamma_S P_S^{x,RP}\right)\left(1-P_S^{x,RP}(2-\gamma_S)\right)}{2}$.

The welfare-maximizing price cap in S is $P_S^{x,P*}=0$.

For the welfare-maximizing reimbursement limit in N and the welfare-maximizing price cap in S, welfare in N is $W_N^{x,RP}\left(r_N^{x,RP*},P_S^{x,RP*}\right)=\frac{\mu(2\gamma_N+1)}{2(\gamma_N+1)^2}$, welfare in S is $W_S^{x,RP}\left(r_N^{x,RP*},P_S^{x,RP*}\right)=\frac{1}{2}$.

Parallel Trade and Reimbursement Limit in the North, Price Cap in the South (RP)

Under parallel trade, a reimbursement limit in the North and a price cap in the South, the firm's profit is $\pi_{N,S}^{RP}=\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{RP}+\left(P_S^{RP}-r_N^{RP}\right)\right)\right)P_S^{RP}+\left(1-\left(\gamma_S P_S^{RP}\right)\right)P_S^{RP}$.

The equilibrium profit is $\pi_{N,S}^{RP}=\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{RP}+\left(P_S^{RP}-r_N^{RP}\right)\right)\right)P_S^{RP}+\left(1-\left(\gamma_S P_S^{RP}\right)\right)P_S^{RP}$.

The firm exports to S if P_S^{RP} is sufficiently high

$(\Delta^{RP}=\pi_{N,S}^{RP}-\pi_N^{xR}=\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{RP}+\left(P_S^{RP}-r_N^{RP}\right)\right)\right)P_S^{RP}+\left(1-\left(\gamma_S P_S^{RP}\right)\right)P_S^{RP}-\frac{\left(\mu+r_N^{x,RP}(1-\gamma_N)\right)^2}{4\mu}>0,$

if $P_S^{RP}>\widehat{P_S^{RP}}=\frac{r_N^{RP}+2\mu-r_N^{RP}\gamma_N-\sqrt{\left(r_N^{RP}\right)^2(1-\gamma_N)^2-4r_N^{RP}\mu(\gamma_N-1)-\mu^2(\mu\gamma_S-3)+r_N^{x,RP}(\gamma_N-1)\left(2\mu+r_N^{x,RP}-\gamma_N r_N^{x,RP}\right)(\mu\gamma_S+1)}}{2\mu\gamma_S+2}$).

For $r_N^{x,RP*}=\frac{\mu}{1+\gamma_N}$, the critical price cap is $\widehat{P_S^{RP}}=\frac{r_N^{RP}+2\mu-r_N^{RP}\gamma_N-\sqrt{\left(r_N^{RP}\right)^2(1-\gamma_N)^2-4r_N^{RP}\mu(\gamma_N-1)+4\frac{\mu^2}{(\gamma_N+1)^2}\left(\gamma_N^2+2\gamma_N-\mu\gamma_S\right)}}{2\mu\gamma_S+2}$.

Welfare in N is $W_N^{RP}=\underbrace{\frac{\left(P_S^{RP}-\mu-r_N^{RP}(1-\gamma_N)\right)^2}{2\mu}}_{CS_N^{RP}}+$

$\underbrace{\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{RP}+\left(P_S^{RP}-r_N^{RP}\right)\right)\right)P_S^{PP}+P_S^{RP}\left(1-\gamma_S P_S^{RP}\right)}_{\pi^{RP}}-\underbrace{(1-\gamma_N)\frac{1}{\mu}\left(\mu-\left(\gamma_N r_N^{RP}+\left(P_S^{RP}-r_N^{RP}\right)\right)\right)r_N^{RP}}_{E_N^{RP}}$

$=\frac{\mu^2-\left(r_N^{RP}\right)^2(1-\gamma_N)^2+2P_S^{RP}\left(\mu+r_N^{RP}(1-\gamma_N)\right)-\left(P_S^{RP}\right)^2(2\mu\gamma_S+1)}{2\mu}$.

Welfare in S is $W_S^{RP} = \underbrace{\frac{\left(1-\gamma_S P_S^{RP}\right)^2}{2}}_{CS_S^{RP}} - \underbrace{(1-\gamma_S)\left(1-\gamma_S P_S^{RP}\right)P_S^{RP}}_{E_S^{RP}} = \frac{\left(1-\gamma_S P_S^{RP}\right)\left(1-P_S^{RP}(2-\gamma_S)\right)}{2}$.

The best response function is $r_N^{RP}\left(P_S^{RP}\right) = P_S^{RP}$.

The welfare-maximizing reimbursement limit in N is $r_N^{RP*} = \frac{\mu(1+\gamma_N)-\mu\sqrt{\gamma_N(1+\gamma_N)-\mu\gamma_S+1}}{\mu\gamma_S+\mu\gamma_N\gamma_S+\gamma_N(1+\gamma_N)}$. The welfare maximizing price cap in S is $P_S^{RP*} = \frac{\mu(1+\gamma_N)-\mu\sqrt{\gamma_N(1+\gamma_N)-\mu\gamma_S+1}}{\mu\gamma_S+\mu\gamma_N\gamma_S+\gamma_N(1+\gamma_N)}$.

Welfare in N is $W_N^{RP}\left(r_N^{RP*},\, P_S^{RP*}\right)$
$= \frac{2\mu\sqrt{1+\gamma_N(1+\gamma_N)-\mu\gamma_S}(\gamma_N+1)(\mu\gamma_S-\gamma_N(1-\gamma_N))+\gamma_N\mu\left(3\gamma_N+\gamma_N^2-\gamma_N^3+2\right)+\mu^2\gamma_S\left(\gamma_N^2(2\gamma_N+3)+3\mu\gamma_S+\mu\gamma_N\gamma_S(\gamma_N+2)-2\right)}{2(\gamma_N+1)^2(\gamma_N+\mu\gamma_S)^2}$.

Welfare in S is $W_S^{RP}\left(r_N^{RP*},\, P_S^{RP*}\right)$
$= \frac{\mu^2\gamma_S\left(\gamma_N+\gamma_N^2+1\right)(2-\gamma_S)-2\mu\gamma_N(\gamma_N+1)^2(1-\gamma_S)+\gamma_N^2(\gamma_N+1)^2-\mu^3\gamma_S^2(2-\gamma_S)+2\sqrt{1-\mu\gamma_S+\gamma_N(\gamma_N+1)}\mu(\gamma_N+1)\left(\gamma_N-\mu\gamma_S+\mu\gamma_S^2\right)}{2(\gamma_N+1)^2(\gamma_N+\mu\gamma_S)^2}$.

A.3. Choice of Regulatory Instruments

No Parallel Trade

For a price cap in S, welfare in N is higher under a price cap
$(W_N^{xP}\left(P_N^{xP*}, P_S^{xP*}\right) - W_N^{xP}\left(r_N^{xR*}, P_S^{xP*}\right) = \frac{\mu\gamma_N(\gamma_N+2)+2\left(1-\gamma_N^2\right)}{2(\gamma_N+1)^2} > 0)$. For a reimbursement limit in S, welfare in N is higher under a price cap $(W_N^{xP}\left(P_N^{xP*}, r_S^{xR*}\right) - W_N^{xR}\left(r_N^{xR*}, r_S^{xR*}\right) = \frac{\mu\gamma_N^2}{2(\gamma_N+1)^2} > 0)$.

Independent of the instrument in N, welfare in S is higher under a price cap
$(W_S^{PP}(., P_S^* = 0) - W_S^R(., r_S^* = 0) = \frac{3(2-\gamma_S)}{8} > 0)$.

Parallel Trade

N,S	Price cap	Reimbursement Limit
Price cap	W_N^{PP}, W_S^{PP}	W_N^{PR}, W_S^{PR}
Reimbursement Limit	W_N^{RP}, W_S^{RP}	W_N^{RR}, W_S^{RR}

1) For price cap in S, welfare in N is higher under a reimbursement limit
$(W_N^{PP} - W_N^{RP} = -\frac{\Omega_{W_N^{PP}-W_N^{RP}}}{2(\gamma_N+\mu\gamma_S)^2(\gamma_N+1)^2},$
with $\Omega_{W_N^{PP}-W_N^{RP}} = 2\mu\sqrt{\gamma_N+\gamma_N^2-\mu\gamma_S+1}\left(\gamma_N+1\right)\left(-\gamma_N+\gamma_N^2+\mu\gamma_S\right)+\gamma_N\mu\left(2\gamma_N-\gamma_N^2-2\gamma_N^3+2\right)$
$-2\mu^2\gamma_S\left(1-\mu\gamma_S\right)-\mu^2\gamma_N\gamma_S\left(\gamma_N+2\right), W_N^{PP} - W_N^{RP} < 0)$.

2. i) For $\mu < \widetilde{\mu^{RR}}$: For reimbursement limit in S, welfare in N is higher under a price cap
$(W_N^{PR} - W_N^{RR} = \frac{\Omega_{W_N^{PR}-W_N^{RR},1}}{4(\gamma_N+\mu\gamma_S)^2\left(4\mu^2+6\mu+1\right)} > 0,$
with $\Omega_{W_N^{PR}-W_N^{RR},1} = 8\mu^5\gamma_S^2 + 16\mu^4\gamma_S\left(\gamma_N+\gamma_S\right) + 2\mu^3\left(8\gamma_N-\gamma_S^2+16\gamma_N\gamma_S\right)$
$+\mu^2\left(24\gamma_N+4\gamma_N^2+\gamma_S^2-4\gamma_N\gamma_S\right) - 2\mu\gamma_N\left(2\gamma_N-\gamma_S-2\right)+\gamma_N^2,$
$W_N^{PR} - W_N^{RR} > 0)$.

ii) For $\mu > \widetilde{\mu^{RR}}$: For reimbursement limit in S, welfare in N is higher under a price cap
$(W_N^{PR} - W_N^{RR} = \frac{\Omega_{W_N^{PR}-W_N^{RR},2}}{8(\gamma_N+\mu\gamma_S)^2(\mu+1)^3(\gamma_N+1)^2} > 0$
$\Omega_{W_N^{PR}-W_N^{RR},2} = 2\mu\sqrt{\mu+1}\left(2\mu+3\right)\left(4\mu+3\right)\left(\gamma_N+1\right)\left(\gamma_N+\mu\gamma_S\right)^2$
$-8\mu\left(\sqrt{\mu+1}\right)^3\left(\gamma_N+1\right)\left(\mu+2\right)\left(\gamma_N+\mu\gamma_S\right)^2 + 4\mu^6\gamma_S^2\left(2\gamma_N+\gamma_N^2-1\right)$
$+2\mu^5\gamma_S\left(4\gamma_N\left(2\gamma_N+\gamma_N^2-1\right)-5\gamma_S+7\gamma_N\gamma_S\left(2+\gamma_N\right)\right)$
$+\mu^4\left(8\gamma_N\left(\gamma_N+\gamma_N^2+1\right)-\gamma_S^2+\gamma_N\gamma_S\left(56\gamma_N+38\gamma_S+28\gamma_N^2+19\gamma_N\gamma_S-20\right)\right)$
$+\mu^3\left(2\gamma_N\left(13\gamma_N+14\gamma_N^2+\gamma_N^3+12\right)+8\gamma_S^2+\gamma_N\gamma_S\left(76\gamma_N+22\gamma_S+38\gamma_N^2+11\gamma_N\gamma_S-2\right)\right)$

$+\mu^2\left(\gamma_N\left(35\gamma_N+38\gamma_N^2+7\gamma_N^3+24\right)+2\gamma_S^2+2\gamma_N\gamma_S\left(22\gamma_N+2\gamma_S+11\gamma_N^2+\gamma_N\gamma_S+8\right)\right)$
$+\mu\gamma_N\left(20\gamma_N+4\gamma_S+22\gamma_N^2+7\gamma_N^3+8\gamma_N\gamma_S+4\gamma_N^2\gamma_S+8\right)+2\gamma_N^2\left(\gamma_N+1\right)^2,$
$W_N^{PR}-W_N^{RR}>0$).

3. For a price cap in N, welfare in S is higher under a price cap ($W_S^{PP}-W_S^{PR}=\frac{3}{8}>0$).

4. i) For $\mu<\widetilde{\mu^{RR}}$: For a reimbursement limit in N, welfare in S is higher under a price cap
($W_S^{RP}-W_S^{RR}=\frac{\Omega_{W_S^{RP}-W_S^{RR},1}}{2(\gamma_N+\mu\gamma_S)^2(\gamma_N+1)^2\left(4\mu^2+6\mu+1\right)^2}$,
with $\Omega_{W_S^{RP}-W_S^{RR},1}=\left(4\mu^2+6\mu+1\right)^2\mu^2\gamma_S\left(\gamma_N+\gamma_N^2+1\right)(2-\gamma_S)-2\mu\gamma_N\left(\gamma_N+1\right)^2(1-\gamma_S)\left(4\mu^2+6\mu+1\right)^2$
$+\gamma_N^2\left(\gamma_N+1\right)^2\left(4\mu^2+6\mu+1\right)^2-\mu^3\gamma_S^2\left(2-\gamma_S\right)\left(4\mu^2+6\mu+1\right)^2$-$(3\mu+1)^2\left(\gamma_N+1\right)^2\left(\gamma_N+\mu\gamma_S\right)^2$
$+2\sqrt{1-\mu\gamma_S+\gamma_N\left(\gamma_N+1\right)}\mu\left(\gamma_N+1\right)\left(\gamma_N-\mu\gamma_S+\mu\gamma_S^2\right)\left(4\mu^2+6\mu+1\right)^2,$
$W_S^{RP}-W_S^{RR}>0$).

ii) For $\mu>\widetilde{\mu^{RR}}$: For a reimbursement limit in N, welfare in S is higher under a price cap
($W_S^{RP}-W_S^{RR}=\frac{\Omega_{W_S^{RP}-W_S^{RR},2}}{16(\gamma_N+\mu\gamma_S)^2(\gamma_N+1)^2(\mu+1)^3}$,
with $\Omega_{W_S^{RP}-W_S^{RR},2}=8\left(\mu+1\right)^3\mu^2\gamma_S\left(\gamma_N+\gamma_N^2+1\right)(2-\gamma_S)-16\left(\mu+1\right)^3\mu\gamma_N\left(\gamma_N+1\right)^2(1-\gamma_S)$
$+8\left(\mu+1\right)^3\gamma_N^2\left(\gamma_N+1\right)^2-8\left(\mu+1\right)^3\mu^3\gamma_S^2\left(2-\gamma_S\right)+2\left(\gamma_N+\mu\gamma_S\right)^2\left(10\gamma_N+5\gamma_N^2+6\right)$
$+16\left(\mu+1\right)^3\sqrt{1-\mu\gamma_S+\gamma_N\left(\gamma_N+1\right)}\mu\left(\gamma_N+1\right)\left(\gamma_N-\mu\gamma_S+\mu\gamma_S^2\right)$
$-4\left(\gamma_N+\mu\gamma_S\right)^2\sqrt{\mu+1}\left(22\mu+32\mu^2+12\mu^3+3\right)\left(\gamma_N+1\right)+16\mu\left(\sqrt{\mu+1}\right)^3\left(\gamma_N+\mu\gamma_S\right)^2\left(\gamma_N+1\right)$
$+6\mu\left(\gamma_N+\mu\gamma_S\right)^2\left(14\gamma_N+7\gamma_N^2+11\right)+24\mu^4\left(\gamma_N+\mu\gamma_S\right)^2$
$+8\mu^2\left(\gamma_N+\mu\gamma_S\right)^2\left(14\gamma_N+7\gamma_N^2+17\right)+8\mu^3\left(\gamma_N+\mu\gamma_S\right)^2\left(6\gamma_N+3\gamma_N^2+13\right).$
$W_S^{RP}-W_S^{RR}>0$).