Differences in Default Risks and Competition in Insurance Markets

Christina Brinkmann *

August 13, 2024

Abstract

The paper studies incentives for insurance sellers to ensure solvency when insurer default risk is a quality dimension of the insurance product. Insurance sellers choose default risks before competing for heterogeneously risk-averse clients. I show that a unique price equilibrium exists for any pair of default risks. The low-default-risk insurer has larger profits. Different from standard vertical product differentiation, market discipline in the choice of default risk emerges: the first mover chooses a low default risk, and the second mover follows with a (potentially small) default risk gap. I discuss the results in the context of over-the-counter derivatives markets.

JEL Classification Numbers: G22, G23, L13, L15

Keywords: OTC Markets, Derivatives, Insurance, Imperfect Competition, Vertical Product Differentiation

*University of Bonn. E-mail: christina-brinkmann@uni-bonn.de. I am grateful to Martin Hellwig for helpful comments and discussions. I thank anonymous referees, Jo Braithwaite, Dominik Damast, Hendrik Hakenes, Tobias Herbst, Eugen Kovac, Christian Kubitza, Stephan Lauermann, David Murphy, Martin Peitz, Farzad Saidi, Martin Schmalz, André Stenzel and Haoxiang Zhu for helpful suggestions and comments, as well as participants of the Finance Seminar at the University of Bonn and the 14th RGS Doctoral Conference in Economics. I gratefully acknowledge financial support from ECONtribute, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. Support by the German Research Foundation (DFG) through CRC TR 224 (Project C03) is gratefully acknowledged. This work was supported by a fellowship of the German Academic Exchange Service (DAAD).

1 Introduction

One characteristic of an insurance product is the default risk of its seller. Sellers can influence their own default risk through precautionary measures to ensure solvency. At the same time, sellers compete for clients, and the competition may create incentives to maintain low default risks – a relevant consideration for market stability. Consider, as a large market for risk transfer, the over-the-counter (OTC) derivatives market. Counterparty default risk is a major concern therein, especially given its role in the instabilities during the Global Financial Crisis (Duffie, 2019). This market is highly concentrated, with few large banks at the core selling derivatives to numerous heterogeneous clients. However, little is known about how oligopolistic competition in insurance markets affects insurers' choices of default risk.

This paper introduces insurer default risk as a quality dimension of the insurance product in a basic insurance model. Insurers sequentially choose their default risks while competing for risk-averse clients. Although all clients prefer lower default risk, their willingness to pay varies due to different levels of risk aversion. I investigate whether market discipline in the choice of default risk emerges in the resulting model of vertical product differentiation.

I find that this is the case when risk aversion is sufficiently relevant. The insurer with the lower default risk has larger profits, incentivizing the first mover in the choice of default risk to choose a low default risk. The second mover then follows with a (potentially small) default risk gap. I discuss implications of endogenous market discipline in the model for introducing central clearing in derivatives markets whereby sellers are shielded from competition in default risk.

In the model, two insurers offer insurance contracts to clients seeking to hedge against a common macro risk. The insurance contracts feature full coverage, except when the insurer defaults. Insurers choose their default risk by deciding on measures to ensure their solvency, e.g., setting aside capital or having balanced trading books. As a result, an insurance product is characterized by the price and its seller's default risk. Clients have CARA utility with varying levels of absolute risk aversion. Competition occurs in two stages: insurers sequentially choose their publicly observable default risks before engaging in simultaneous price competition.

The main results of the model are as follows, presented following backward induction through the stages of the model.

First, in stage three, when clients make their purchase decisions, there is self-selection. All clients prefer a low default risk to a high default risk, but their willingness to pay for low default risks varies due to differences in risk aversion. As a result, there is an indifferent client that segments the market with more risk-averse clients self-selecting to buy from the insurer with the lower default risk. This market segmentation hinges on differentiated default risks. Insurers make positive profits.

Second, in stage two, when prices are set, for every pair of default risks, a unique pair of prices exists that forms a subgame-perfect Nash equilibrium. The price equilibrium is such that the insurer with the higher default risk chooses a lower price.

Third, price equilibria and, subsequently, profits depend only on a function in default risks that is close to a function in the difference in default risks. I call the difference in default risks the *default risk gap*.

Fourth, in equilibrium, the insurer with the lower default risk (i.e., that offers the insurance product of higher quality) has larger profits than the other insurer. This renders the leadership position in quality more attractive.

This has two key implications for the first stage of the game, when default risks are sequentially chosen. First, the first mover is under pressure to choose a low default risk: Since the insurer with the lower default risk has larger profits, the first mover aims to occupy this position vis-à-vis the second mover. As a result, he chooses a sufficiently low default risk to exclude the possibility that the second mover reverses roles. In particular, the smallest optimal default risk gap approximately determines an upper bound for the default risk of the first mover. In general, the default risk of the first mover may not exceed roughly half of the worst (externally given) admissible default risk since this is an upper bound for the default risk gap.

Second, there are push-and-pull factors on the second-mover's choice of default risk. That is, if there is an optimal default risk gap for the second mover, the second mover will keep this gap (relative to the first mover). Under two conditions that are simple but probably more restrictive than necessary, the competitive situation can be summed up based on the default risk gap. If the profit of the second mover as a function of the second-mover's default risk has a unique interior maximum, and if the profit of the first mover as a function of the second-mover's default risk is increasing, one can fully characterize the equilibrium default risks. Broadly speaking, the first mover will choose a default risk below a threshold lower than the optimal default risk gap, and the default risk of the second mover will be that of the first mover plus the optimal default risk gap.

Lastly, in a numerical example, I demonstrate that the two above conditions hold for a plausible set of parameter values, and that the first-mover's default risk and the default risk gap can be small – much smaller than the admissible default risks in the model. Thus, competitive forces alone can lead to low default risks. Since the overall outcome varies somewhat smoothly with the parameter values, the conclusions drawn from the numerical example extend to a neighborhood of parameter values and are, therefore, "locally generic" for the parameters of the numerical example.

In sum, pressure to choose a low default risk for the first mover and a push-and-pull effect on the second-mover's choice of default risk can be seen as market discipline in the choice of default risks.

The model captures essential features of derivatives markets and may serve as a framework for exploring open questions about market structure. Derivatives can be seen as insurance products offered by dealers, typically large banks. Derivatives markets exhibit a hub-and-spoke structure (Abad, Aldasoro, Aymanns, D'Errico, Rousová, Hoffmann, Langfield, Neychev, and Roukny, 2016), with few dealers at the core and numerous heterogeneous clients in the periphery, which aligns with the model setup. This structure persists even when the market is centrally cleared through a central counterparty (CCP)[1], as typically only dealers are members of the CCP, and most market participants access central clearing as clients of these members (*client clearing*)[2]. However, interposing a CCP at the core of a highly concentrated market raises the question of the effect of central clearing on competition.

1 A CCP replaces a contract between two of its members with two contracts that each have the CCP on one end. It thereby insulates the contracting parties from the risk that the counterparty defaults.

2 See, e.g., Financial Stability Board (2018) and CPMI, IOSCO (2022).

The model provides a framework to conceptualize the effects of central clearing on competition. A salient feature of a centrally cleared market is that members of the CCP do not differ in their default risks from the client's perspective, primarily due to mechanisms that port clients' portfolios from one member to another in case of a default (Braithwaite and Murphy, 2020). This reduces competition in price and default risk to competition in prices alone. However, the model demonstrates that market discipline in choosing default risk emerges *as a result of* two-dimensional competition (price and default risk). Thus, the model highlights a market force that may be absent in centrally cleared markets where dealers are shielded from competition in default risks.

Related Literature. This paper contributes to three strands of the literature. First, I extend the literature on insurance markets following the seminal work by Rothschild and Stiglitz (1976). I introduce the seller's default risk as a quality dimension of the insurance product. Consequently, the focus of client heterogeneity shifts: clients differ in their risk aversion but are identical in the underlying endowment risk. A model introducing differences in insurer *service* quality has been developed by Schlesinger and Von der Schulenburg (1991), but their model centers around horizontal product differentiation and search costs.

Second, I add to the body of work on vertical product differentiation from the industrial organization literature (Gabszewicz and Thisse, 1979, 1980; Shaked and Sutton, 1982, 1983). In the standard model in Tirole (1988), which closely follows Shaked and Sutton (1982), two firms compete in product quality and price. Maximal differentiation in quality choices emerges because differentiation in quality softens price competition. I adapt this class of models to the insurance context by mapping insurers' default risks to (inverse) qualities and introducing risk aversion into consumer utility. As a result, consumer utility becomes non-linear, reversing the result of maximal differentiation: push-and-pull factors and upward pressure on both qualities emerge, which I interpret as market discipline in quality choices. To clarify which assumptions need to be removed from the standard linear model to produce analogous results, I revisit the standard model in the Online Appendix.[3]

Third, I contribute to a growing literature on the market structure of OTC derivatives

3 Differently from Moorthy (1988, 1991), who lifts the same assumptions and numerically computes and compares outcomes, I use a general convex cost function and derive the push factor directly from profit-maximizing incentives.

markets, initiated by Duffie, Gârleanu, and Pedersen (2005) and Atkeson, Eisfeldt, and Weill (2015).[4] Seminal papers on central clearing in derivatives markets have examined netting benefits (Duffie and Zhu, 2011), transparency (Acharya and Bisin, 2014), and the role of margins (Biais, Heider, and Hoerova, 2012, 2016, 2021). I add to this literature by focusing on the nature of competition between the dealers at the core of the market and by introducing the notion of differentiation in default risks. Competition between dealers is also studied in Carapella and Monnet (2020), who investigate the effect of central clearing in derivatives markets on dealers' entry decisions. The idea is that if more dealers enter as a result of the regulation, more intense competition and a resulting lower level of spreads may alter incentives to invest in efficient technologies ex-ante. Unlike their model, where all agents are risk-neutral, and the focus is on search frictions for dealers intermediating derivatives, the model in this paper emphasizes clients' risk aversion as the driving force behind dealer competition and default risk differentiation.

The rest of the paper is organized as follows. Section 2 introduces the model framework. Section 3 derives key results on self-selection and illustrates the setup. Section 4 shows uniqueness and existence of price equilibria for any pair of default risks. Section 5 analyses the choices in default risks. Section 6 presents a numerical example. Section 7 discusses an application of the model setup to derivatives markets, and Section 8 concludes.

2 Model

2.1 Setup

There is a continuum of risk-averse *clients* with a hedging need and two risk-neutral *insurers*.

Clients. Each client has an asset $\tilde{x}$ which takes the value $\underline{\theta}$ with probability p and $\overline{\theta}$ with probability $(1-p)$. Let the expected value of the asset be zero, and the bad endowment state a loss.[5] The endowment risk is the *same* across all clients, and p is commonly known.

4 See Dugast, Üslü, and Weill (2022) for recent work on the coexistence of OTC and centralized markets.
5 Otherwise $E[\tilde{x}]$ is a certain payment and consider the random variable $\tilde{x} - E[\tilde{x}]$ instead of $\tilde{x}$.

Clients are risk-averse with CARA utility[6]

$$u_a(x) = -\exp(-ax). \tag{1}$$

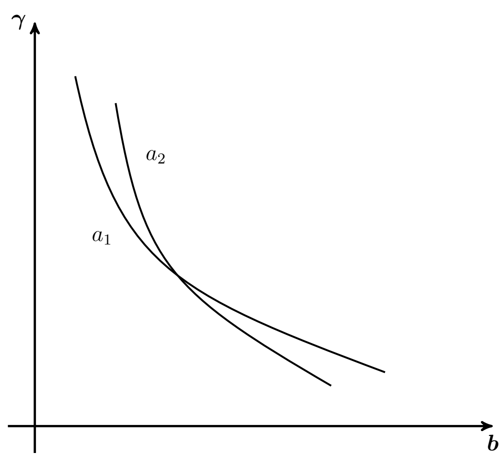
Insurance Contract and Default Risk. Each insurer offers a full-coverage insurance contract for a fixed payment of γ. However, insurers default with some probability in the bad endowment state, in which case they do not honor the contractual obligations of the insurance contract. Insurers choose their default risk $b_i, i \in \{1, 2\}$, i.e., the probability that they default in the bad endowment state.

A client with risk aversion parameter a derives the following utility from a contract (b, γ), sold by insurer with default risk b at price γ,

$$U_a(b, \gamma) := (1 - bp)u_a(-\gamma) + bpu_a(\underline{\theta}). \tag{2}$$

As illustrated in Figure 1, the marginal rate of substitution, i.e. the necessary reduction in the price γ for an increase in default risk b to keep a client indifferent, is increasing in a.[7] More risk-averse clients have a larger willingness to pay for an increase in quality.

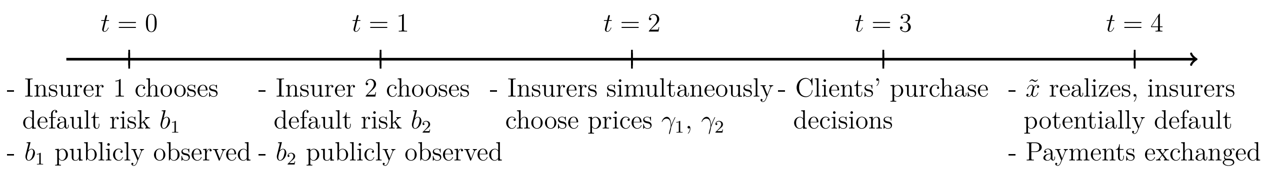
Figure 1: Illustration of Indifference Curves for Two Clients with $a_1 < a_2$



6 The model remains unchanged with the cardinally equivalent utility $v_a(x) = 1/a(1 - \exp(-ax))$.
7 One can verify that $\partial MRS(a)/\partial a = -p/((1-bp)a) \left(\exp(-a(\underline{\theta} + \gamma)) \left[1/a + \underline{\theta} + \gamma\right] - 1/a\right)$. To see that this expression is positive, note that for $(\underline{\theta} + \gamma) < -1/a$ it follows directly. For $0 > (\underline{\theta} + \gamma) > -1/a$ it follows, since for all $x \neq 0$ $\exp(x) > 1 + x$.

Timing. There are five points in time, $t \in \{0, 1, 2, 3, 4\}$. In $t = 0$, insurer 1 chooses default risk b_1. In $t = 1$, upon observing insurer 1's default risk, insurer 2 chooses his default risk b_2. In $t = 2$ they simultaneously choose prices $\gamma_i, i \in \{1, 2\}$ for a full-coverage insurance contract.[8] Upon observing the insurers' choices (b_1, γ_1) and (b_2, γ_2), clients decide from whom to buy in $t = 3$. Lastly, clients' endowments are realized, and payments are exchanged in $t = 4$ unless there is a default. Figure 2 summarizes the timing of events.

Figure 2: Timeline



We study subgame-perfect Nash equilibria in the resulting game.

2.2 Discussion

The assumption that default risks are chosen before other actions embeds a commitment assumption. Once chosen, a default risk cannot be modified at later points in time. This excludes a situation in which an insurer abandons precautionary measures after they signaled a low default risk. In the simple timing structure of the present model, commitment seems a reasonable assumption. First, precautionary measures that insurers undertake to reduce the probability of their own default, such as setting aside capital, sufficient liquidity buffers, or balancedness of the trading books, are relatively long-term strategic decisions. They are here seen as investments that are sunk costs during later phases of competition, not continued period-per-period expenses. Second, default risks need to become public information before clients' purchase decisions. The disclosure of such information and the associated build-up of reputation also takes time.

I assume that default risks are chosen sequentially, which is a simplifying assumption. With simultaneous choice, any pure-strategy equilibrium in qualities cannot be symmetric,

8 Specifically, it is assumed that γ is the upfront premium for establishing the client-insurer relationship. Afterward, the insurer offers the actuarily fair price, and clients subsequently pick trade volumes that result in full insurance. Hence, the insurer's profit per client is γ. See Appendix B1 for details.

as equal qualities yield zero profits for insurers. Thus, with simultaneous quality choices, there are multiple equilibria (with reversed roles). In Shaked and Sutton (1982), roles are thus assigned upfront. In this model, roles are instead assigned via sequential quality choice (as, e.g., in Aoki and Prusa (1997); Lehmann-Grube (1997)).

The setup implies somewhat restrictive assumptions regarding the contract terms: Insurers are limited to offering full-coverage insurance (unless they default) and only have discretion over the premia. The situation maps to a situation in which clients pay a fixed premium to establish the client-insurer relationship, after which the insurer provides insurance at fair prices and clients subsequently choose full insurance. The setup rules out a situation in which insurers offer a menu of contracts that differ in coverage and thus additionally compete in coverage. I make two points in defense of this assumption. First, the novel aspect of the model is competition in default probabilities seen as a quality dimension of the insurance products. To keep this analysis tractable, I keep other dimensions of the competition as simple as possible. Second, in the context of derivatives markets, which will be discussed later, full coverage is a typical feature. For example, a plain-vanilla interest rate swap specifies the exchange of a fixed interest rate for a floating rate without variation in coverage.

Similarly, it is assumed that insurers are unable to discriminate among clients based on their risk aversion. In other words, I assume that risk aversion is private information to clients. One may debate how much information insurers are able to acquire about the risk attitudes of their clients. In the context of derivatives markets as an over-the-counter market, clients may additionally have a hard time comparing prices. Assuming that risk aversion is private information, nonetheless, seems a natural starting point and one that facilitates an analysis with respect to vertical product differentiation. In related work on vertical product differentiation, first-degree price discrimination is ruled out.

3 Stage 3: Clients' Purchase Decisions

The model is solved by backward induction, starting with clients' purchase decisions in $t = 3$. This section establishes that the market is segmented with more risk-averse clients buying from the insurer with the lower default risk (Proposition 1) and derives some properties to graphically illustrate the setup (Figure 4).

In $t = 3$, insurers' default risks are given. To fix roles, insurer 1 defaults with a lower probability or, in other words, offers a product of higher quality. That is, let $\Delta b := b_2 - b_1 > 0$. Let $\vec{b} := (b_1, b_2)$ and $\vec{\gamma} := (\gamma_1, \gamma_2)$ denote the pairs of default risks and prices.

Lemma 1 (Characterization of the Indifferent Client). *A client with degree of risk aversion a is indifferent between two contracts (b_1, γ_1) and (b_2, γ_2) with $\Delta b > 0$ if*

$$g(a, \vec{\gamma}) := \frac{\exp(-a\Delta\gamma) - 1}{\exp(-a(\underline{\theta} + \gamma_2)) - 1} = \frac{p\Delta b}{1 - b_1 p} =: \tilde{g}(\vec{b}). \tag{3}$$

Proof. See Appendix A1. □

For any two contracts with $b_2 > b_1$, if there is a solution to (3), then $\gamma_1 > \gamma_2$.[9] That is, the insurer that offers the product of higher quality sets the higher price.

The main result of this section (Proposition 1) establishes that there is at most one client characterized by some a^* who is indifferent between contracts (b_1, γ_1) and (b_2, γ_2) and segments the market. For the existence of a unique indifferent client, the utility loss due to the payment of the price relative to the utility loss due to the bad endowment needs to diminish as clients become more risk-averse. In other words, the function g needs to decrease in the risk aversion parameter – akin to a single crossing condition. A lower bound on $-a(\underline{\theta} + \gamma_i)$ is sufficient for this, which is ensured by the following set of assumptions.[10]

Assumption A1.

$$p < \frac{1}{3}.$$

Assumption A2.

$$\text{For } i \in \{1, 2\} : b_i \in [0, b^{max}] \text{ with } b^{max} \leq \frac{1}{3}.$$

9 To see this, note that with $\Delta b > 0$, the RHS of (3) is positive. The denominator of the LHS of (3) is positive, which necessitates $\Delta\gamma < 0$.

10 From assumptions A3 and A4 we get $-a(\underline{\theta} + \gamma_i) > 2$. To see this, note that $-a(\underline{\theta} + \gamma_i) > 2 \Leftrightarrow \gamma^{max} < 2/(-\underline{a}) - \underline{\theta}$. The RHS holds, since by assumption A3 $\gamma^{max} < (-\underline{\theta})/3$ and $2/(\underline{a}) - \underline{\theta} > (-\underline{\theta})/3 \Leftrightarrow (-\underline{a})\underline{\theta} > 3$, which is ensured by assumption A4.

Assumption A3.

$$\text{For } i \in \{1,2\} : \gamma_i \in [0, \gamma^{max}] \text{ with } \gamma^{max} \leq \frac{1}{3}(-\underline{\theta}).$$

Assumption A4.

$$\underline{a}(-\underline{\theta}) > \log\left[\frac{1-\frac{1}{8}}{\exp(-2)-\frac{1}{8}}\right] \approx 4.4.$$

The assumptions bound the probability of the bad endowment state (assumption A1) and the default risks (assumption A2). They, thus, focus attention on a setup of insurance against an infrequent large event as well as a setup with the default of an insurer being the exception rather than the norm.

Regarding assumption A3, note that $\gamma^{max} \leq (-\underline{\theta})$ by construction, since otherwise the price exceeds the bad endowment. A priori, there is no market for prices exceeding the price above which the most risk-averse client is unwilling to buy insurance even if offered with the lowest default risk (see Appendix B2 for details). In the numerical example, assumption A3 is non-binding in equilibrium.

Assumption A4 imposes a lower bound on the degree of risk aversion times the absolute value of the bad endowment, $a(-\underline{\theta})$ for all $a \in [\underline{a}, \overline{a}]$. It is a condition on both the range of a and $\underline{\theta}$: For any large $\underline{\theta}$, one can find a small a such that assumption A4 is violated. Intuitively, for any large payment without limitations on a, one can find clients whose utility is sufficiently close to a risk-neutral one (i.e., a close to 0) such that risk aversion barely kicks in. Assumption A4 rules out such almost risk-neutral clients – relative to the bad endowment. Hence, it demands that risk aversion is relevant for all clients.

Proposition 1 (Self-Selection). *Suppose assumptions A1 - A4. For given contracts* (b_1, γ_1) *and* (b_2, γ_2) *with* $\Delta b > 0$*, there is at most one indifferent client* $a^*(\vec{\gamma})$ *satisfying*

$$g(a^*(\vec{\gamma}), \vec{\gamma}) = \tilde{g}(\vec{b}) = \frac{p\Delta b}{1 - b_1 p}. \tag{4}$$

Such an indifferent client $a^*(\vec{\gamma}) \in [\underline{a}, \overline{a}]$ indeed exists, if

$$g(\overline{a}, \vec{\gamma}) \leq \frac{p\Delta b}{1 - b_1 p} \leq g(\underline{a}, \vec{\gamma}). \tag{5}$$

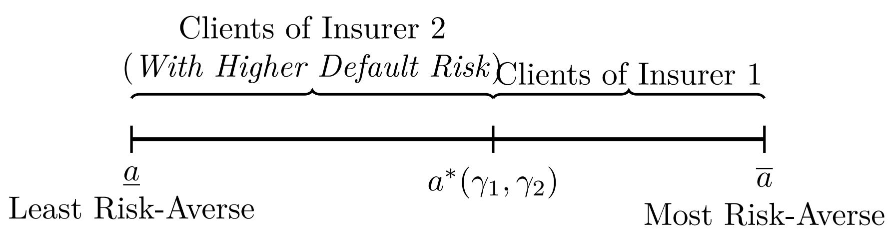
In this case, client a will choose insurer 1 iff

$$a \geq a^*(\vec{\gamma}). \tag{6}$$

Proof. See Appendix A2. □

The result implies market segmentation. Clients with a risk aversion larger $a^*(\vec{\gamma})$ buy from insurer 1, while insurer 2 receives clients with a level of risk aversion below the threshold $a^*(\vec{\gamma})$, as depicted in Figure 3.

Figure 3: Market Segmentation for Two Contracts, (b_1, γ_1) and (b_2, γ_2) with $b_2 > b_1$



Formally, for given default risks $\vec{b}$, a^* is defined via $g(a^*(\vec{b}, \vec{\gamma}), \vec{\gamma}) = \tilde{g}(\vec{b})$ on the set

$$\mathcal{G}_{[\underline{a}, \overline{a}]} := \left\{ \vec{\gamma} \,|\, 0 \leq \gamma_2 < \gamma_1 \leq \gamma^{max} \text{ and } g(\underline{a}, \vec{\gamma}) \leq \tilde{g}(\vec{b}) \leq g(\overline{a}, \vec{\gamma}) \right\}. \tag{7}$$

Let $\mathcal{G}_0 := \left\{ 0 \leq \gamma_2 < \gamma_1 \leq \gamma^{max} \right\}$. Then the insurers' profits are

$$\Pi_1(\gamma_1, \gamma_2) = \begin{cases} (\overline{a} - a^*(\gamma_1, \gamma_2))\,\gamma_1 & \text{on } \mathcal{G}_{[\underline{a}, \overline{a}]} \\ (\overline{a} - \underline{a})\,\gamma_1 & \text{on } \mathcal{G}_0 \setminus \mathcal{G}_{[\underline{a}, \overline{a}]} \text{ if } \tilde{g}(\vec{b}) \leq g(\underline{a}, \vec{\gamma}) \\ 0 & \text{on } \mathcal{G}_0 \setminus \mathcal{G}_{[\underline{a}, \overline{a}]} \text{ if } g(\overline{a}, \vec{\gamma}) \leq \tilde{g}(\vec{b}) \end{cases} \tag{8}$$

$$\Pi_2(\gamma_1, \gamma_2) = \begin{cases} (a^*(\gamma_1, \gamma_2) - \underline{a})\,\gamma_2 & \text{on } \mathcal{G}_{[\underline{a}, \overline{a}]} \\ 0 & \text{on } \mathcal{G}_0 \setminus \mathcal{G}_{[\underline{a}, \overline{a}]} \text{ if } \tilde{g}(\vec{b}) \leq g(\underline{a}, \vec{\gamma}) \\ (\overline{a} - \underline{a})\,\gamma_2 & \text{on } \mathcal{G}_0 \setminus \mathcal{G}_{[\underline{a}, \overline{a}]} \text{ if } g(\overline{a}, \vec{\gamma}) \leq \tilde{g}(\vec{b}) \end{cases} \tag{9}$$

In the following we restrict attention to the set $\mathcal{G}_{[\underline{a},\overline{a}]}$.

The setup admits no closed-form solutions. Instead, in the remainder of this section, we characterize market share elasticities and graphically illustrate the setup. The following notation is introduced for an explicit characterization in the next Lemma of how the indifferent client changes as either insurer increases prices, but not needed for the subsequent text. Define

$$\tilde{A} : [\underline{a}, \overline{a}] \times [0, -\underline{\theta})^2 \to \mathbb{R}, \quad (a, \vec{\gamma}) \mapsto \exp(-a\Delta\gamma) \tag{10}$$

$$\text{and} \quad \tilde{B}_i : [\underline{a}, \overline{a}] \times [0, -\underline{\theta}) \to \mathbb{R}, \quad (a, \gamma_i) \mapsto \exp(-a(\underline{\theta} + \gamma_i)) \tag{11}$$

and let

$$A(\vec{\gamma}) := \tilde{A}(a^*(\vec{\gamma}), \vec{\gamma}), \quad \text{and} \quad B_i(\vec{\gamma}) := \tilde{B}_i(a^*(\vec{\gamma}), \gamma_i) \tag{12}$$

be the two functions, defined on $[0, -\underline{\theta})^2$, one obtains when inserting the indifferent client $a^*(\vec{\gamma})$ into (10) and (11). Since, for given $\vec{b}$, the RHSs of (3) and (A32) are constant, we infer that the respective LHSs, i.e.

$$g(a^*(\vec{\gamma}), \vec{\gamma}) = \frac{A(\vec{\gamma}) - 1}{B_2(\vec{\gamma}) - 1} \quad \text{and} \quad h(a^*(\vec{\gamma}), \vec{\gamma}) = \frac{1 - \frac{1}{A(\vec{\gamma})}}{B_1(\vec{\gamma}) - 1}, \tag{13}$$

are constants, and call them g and h, respectively. Finally, define

$$\xi_2 := (\underline{\theta} + \gamma_2), \quad \varphi_1 := \xi_2 B_1 \quad \text{and} \quad \tau_1 := (\Delta\gamma - g\varphi_1), \tag{14}$$

$$\text{as well as} \quad \xi_1 := (\underline{\theta} + \gamma_1), \quad \varphi_2 := \xi_1 B_2, \quad \text{and} \quad \tau_2 := (\Delta\gamma - h\varphi_2). \tag{15}$$

The following Lemma shows that both insurers indeed lose market share when increasing prices.

Lemma 2 (**Market Shares**). *Suppose assumptions A1 - A4. The indifferent client is increasing in γ_1 and decreasing in γ_2, namely*

$$\partial_1 a^* = \frac{a^*}{\tau_1} > 0 \tag{16}$$

$$\partial_2 a^* = \frac{-a^*}{\tau_2} < 0. \tag{17}$$

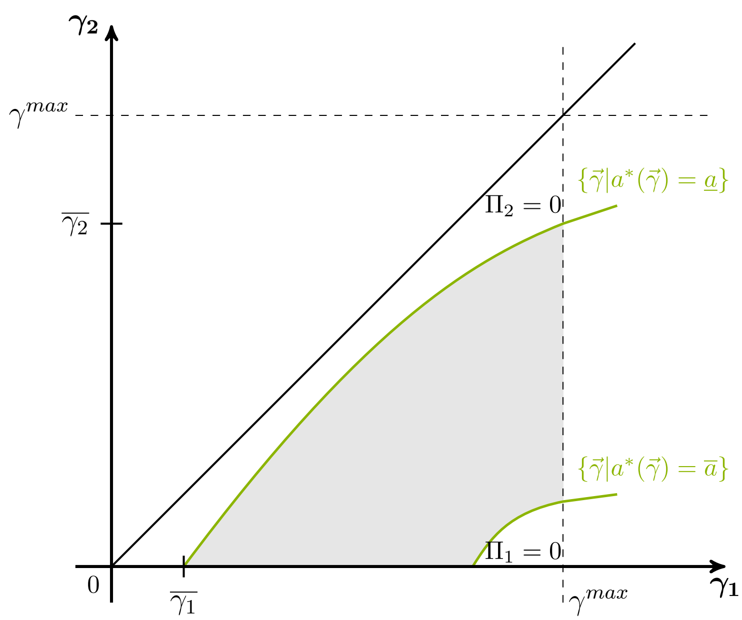
For the slope of a contour line $\{(\gamma_1, \gamma_2) | \, a^*(\gamma_1, \gamma_2)$ *constant*$\}$ *we have*

$$\frac{-\partial_2 a^*}{\partial_1 a^*} =: \alpha < 1. \tag{18}$$

Proof. See Appendix A4. □

Figure 4 visualizes the setup with prices set by insurers 1 and 2 on the x- and y-axis, respectively. With insurer 1 the insurer with the lower default risk offering insurance at a higher

Figure 4: Illustration of the Setup



Notes: The figure depicts, for given default risks, insurer 1's and 2's prices on the x- and y-axis, respectively. See the text for a detailed explanation.

price, pairs of prices lie below the diagonal. The green line just below the diagonal depicts the pairs of prices above which insurer 2 has no market share and, subsequently, no profits. For $\gamma_2 \in [0, \gamma^{max}]$, we parameterize these pairs by defining $\gamma_1^{\underline{a}}(\gamma_2)$ such that $a^*(\gamma_1^{\underline{a}}(\gamma_2), \gamma_2) = \underline{a}$. From Lemma 2 we know that contour lines of a^* have a slope below one.

The visualization in Figure 4 offers an alternative justification for assumption A4. Denote by $\overline{\gamma_1}$ and $\overline{\gamma_2}$ the intercepts of upper green line, $\gamma_1^{\underline{a}}(\cdot)$, with the x- and y-axis, respectively. Assumption A4 is equivalent to demanding $\overline{\gamma_2} > 0$ (see Appendix B3 for details). In other words, assumption A4 demands that insurer 1 does not a priori get the entire market – making the setup interesting to begin with.

4 Stage 2: Price Setting

This section establishes existence and uniqueness of a Nash equilibrium in prices in $t = 2$ for a given pair of default risks under the following two additional assumptions.

Assumption A5.

$$\overline{a} \leq \frac{3}{2}\,\underline{a}.$$

Assumption A6.

$$\partial_1 \Pi_1 \left(\gamma_1^{\underline{a}}(\gamma_2^*(\gamma^{max})), \gamma_2^*(\gamma^{max})\right) \geq 0.$$

Assumption A6 is a technical assumption involving insurer 2's reaction function γ_2^*. It demands that at a point at which insurer 1 "owns" the entire market, insurer 1 has no incentive to decrease prices. The assumption is required because a negative market share at negative prices also leads to positive turnover – a case certainly not of interest.

Proposition 2 (Existence and Uniqueness). *Suppose assumptions A1 - A6. Consider a pair of default risks (b_1, b_2) with $\Delta b \geq 0$. Then,*

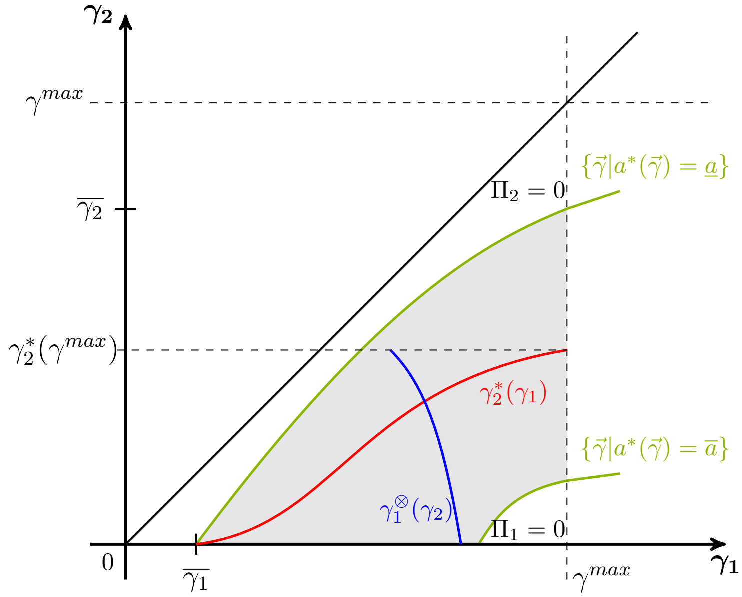
i) If $b_1 < b_2$, there exists a unique Nash equilibrium in prices (γ_1, γ_2).

ii) If $b_1 = b_2$, a client can be indifferent only if $\gamma_2 = \gamma_1$. That is, if insurers' default risks coincide, pure price competition drives prices to marginal costs (which are set to zero here).

Proof. See Appendix A11. The proof builds on the existence of insurer 1's and insurer 2's reaction functions (Propositions 3 and 4). □

The intuition of the proof is as follows: Insurer 2's reaction function, $\gamma_2^*(\gamma_1)$ in red, is strictly increasing. Thus, there exists an inverse function. From the boundary values of the

Figure 5: Price Equilibrium



Notes: The figure depicts, for given default risks, insurer 1's and 2's prices on the x- and y-axis, respectively. See the text for a detailed explanation.

inverse function, there must be an intersection with insurer 1's reaction function, $\gamma_1^{\otimes}$ (as depicted in Figure 5). Formally, we apply Brouwer's Fixed Point Theorem for existence. From the bounds on $\partial_2\gamma_1^{\otimes}$ and $\partial_1\gamma_2^*$ in Propositions 4 and 3, respectively, it follows that there can be at most one intersection.

We now formally show existence and properties of insurer 1's and insurer 2's reaction functions.

Proposition 3 (Insurer 2's Reaction Function). *Suppose assumptions A1 - A4. Suppose some fixed default risks* (b_1, b_2) *with* $\Delta b > 0$*. Then,*

i) for any $\gamma_1 \in [0, \gamma^{max}]$*, there is a unique best response in prices for insurer 2,* $\gamma_2^*(\gamma_1)$*. For* $\gamma_1 \in (\overline{\gamma}_1, \gamma^{max})$*,* γ_2^* *is in the interior of* $\mathcal{G}_{[\underline{a},\overline{a}]}$ *and uniquely characterized via* $\partial_2\Pi_2 = 0$*.*

ii) for $\gamma_1 \in [\overline{\gamma}_1, \gamma^{max}]$*,* γ_2^* *is a smooth function and strictly increasing in* γ_1*.*

iii) $\partial_1\gamma_2^* < 1/\alpha^*$ *with* $\alpha^* := \alpha(\gamma_1, \gamma_2^*(\gamma_1))$*, i.e.,* α *evaluated on insurer 2's reaction function.*

Proof. See Appendix A6. □

The strategy of the proof is standard: Uniqueness follows from $\partial_2^2\Pi_2 < 0$, and existence follows since profits are a continuous function that is zero at the boundaries of the interval.

For the other insurer, existence of a reaction function is not straightforward since insurer 1's profit function is not necessarily concave. In fact, parameter restrictions ensuring concavity are not compatible with the existing set of assumptions that require risk aversion to have enough bite. Without concavity of insurer 1's profit function, points that satisfy the first-order condition need not correspond to best responses. Instead, we prove an auxiliary Lemma (Lemma 4 in the Appendix) for a smooth real-valued function f on some interval $[a, b]$ with $\partial f(a) > 0$: If there exists a point in the interval below which local extrema may only be local minima and above which local extrema may only be local maxima, then f has a global maximum. Assumptions A5 and A6 ensure that we can use this Lemma to obtain insurer 1's best responses for the relevant interval, that is, for $\gamma_2 \in [0, \gamma_2^*(\gamma^{max})]$.

Proposition 4 (Insurer 1's Reaction Function). *Suppose assumptions A1 - A6. Suppose some fixed default risks (b_1, b_2) with $\Delta b > 0$. Then,*

i) for any $\gamma_2 \in [0, \gamma_2^*(\gamma^{max})]$, there is a unique best response in prices for insurer 1, $\gamma_1^{\otimes}(\gamma_2)$. $\gamma_1^{\otimes}$ is uniquely characterized via

$$\partial_1\Pi_1(\gamma_1^{\otimes}(\gamma_2), \gamma_2) = 0 \qquad or \qquad \Big(\gamma_1^{\otimes}(\gamma_2) = \gamma^{max}\, and\ \forall \mu \geq \gamma_1^{a}(\gamma_2) : \partial_1\Pi_1(\mu, \gamma_2) > 0\Big).$$

ii) $\gamma_1^{\otimes}$ is a continuous function, smooth except at finitely many points.

iii) $\partial_2\gamma_1^{\otimes} < \alpha^{\otimes}$ with $\alpha^{\otimes} := \alpha(\gamma_1^{\otimes}(\gamma_2), \gamma_2)$, i.e., α evaluated on insurer 1's reaction function.

Proof. See Appendix A7. □

5 Stage 1: Choices of Default Risks

This section analyses subgame-perfect equilibria in default risks and the competitive mechanism at play.

The model features vertical product differentiation: Ceteris paribus, all clients prefer the insurer with the lower default risk, but clients differ in their valuation for low default risks.

Compared to the standard model of vertical product differentiation as in Shaked and Sutton (1982) or Tirole (1988), the inclusion of risk aversion leads to non-linear utility which does not admit closed-form solutions. Yet, an analog to the result in Shaked and Sutton (1982) that the high-quality firm has larger profits still holds.

Proposition 5 (Lower Default Risk More Attractive). *Suppose assumptions A1 - A6. At any Nash equilibrium in prices,*

i) the insurer with the lower default risk (quality leader) has larger profits, $\Pi_1 > \Pi_2$,

ii) the insurer with the lower default risk has a larger market share, $(\overline{a} - a^*) > (a^* - \underline{a})$.

Proof. See Appendix A12. □

There is a simple characterization of default risks that lead to the same price equilibrium.

Proposition 6 (Price Equilibrium Depends on the Default Risk Gap). *Suppose assumptions A1 - A6.*

i) Default risks (b_1^0, b_2^0) *and* (b_1, b_2) *with* $\tilde{g}(b_1, b_2) = \tilde{g}(b_1^0, b_2^0)$ *lead to the same price equilibrium.*

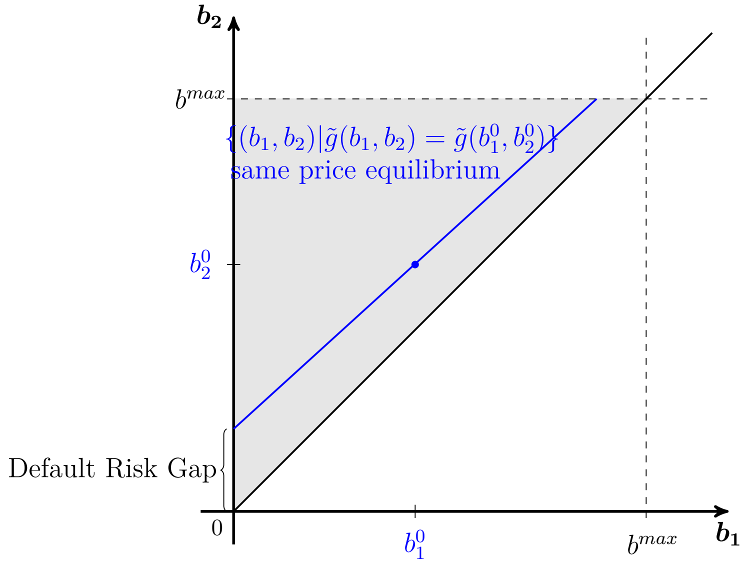
ii) On the set of default risks $\{b \in [0, b^{max}]^2 | b_1 < b_2\}$*, price equilibria (and subsequently profits) are constant on straight lines with slope* $(1 - \tilde{g})$.

Proof. See Appendix A13. □

Figure 6 illustrates the result, with default risks of insurers 1 and 2 on the x- and y-axis, respectively. Since insurer 1 has the lower default risk, the default risks lie above the diagonal (shaded). For (b_1^0, b_2^0), the blue line depicts all pairs of default risks that lead to the same value of $\tilde{g}$ and, consequently, the same price equilibrium. $\tilde{g}$ is small, so the slope of the blue line is nearly parallel to the 45-degree line. Therefore, the gap between pairs of default risks that lead to the same price equilibrium (*default risk gap*) changes only slightly as b_1 changes.

For default risks $\vec{b} = (b_1, b_2)$, let $\vec{\gamma}^{\square}(\vec{b})$ be the corresponding price equilibrium. As shown in Proposition 2, the price equilibrium exists and is unique; hence $\vec{\gamma}^{\square}(\vec{b})$ is well-defined. In

Figure 6: Default Risks Leading to the Same Price Equilibrium Lie on Straight Lines



Notes: The figure depicts insurer 1's and 2's default risks on the x- and y-axis, respectively. See the text for a detailed explanation.

Appendix B4, we show that price equilibria are smooth functions in qualities. By $\Pi_i^{\square}(\vec{b})$, we denote profits associated with a pair of default risks under optimal price setting in the subsequent period, $\Pi_i^{\square}(\vec{b}) := \Pi_i\left(\vec{\gamma}^{\square}(\vec{b}), \vec{b}\right)$.

Proposition 7 (First Mover Chooses a Low Default Risk, Second Mover Follows). *Suppose assumptions A1 - A6. Let Π_2^* be the global maximum of $\Pi_2^{\square}(0, b_2)$ as a function of b_2. Let b_2^s and b_2^l be the smallest and largest b_2 for which this maximum is assumed, i.e., b_2^s is the smallest default risk gap of a subgame-perfect equilibrium of the form $(0, b_2)$. Let (b_1^*, b_2^*) be a subgame-perfect Nash equilibrium in default risks. Then*

i) $b_1^* < b_2^s$.

ii) $b_2^* \leq \left(2 - \tilde{g}(0, b_2^l)\right) b_2^l$.

A general upper bound for b_1^* is

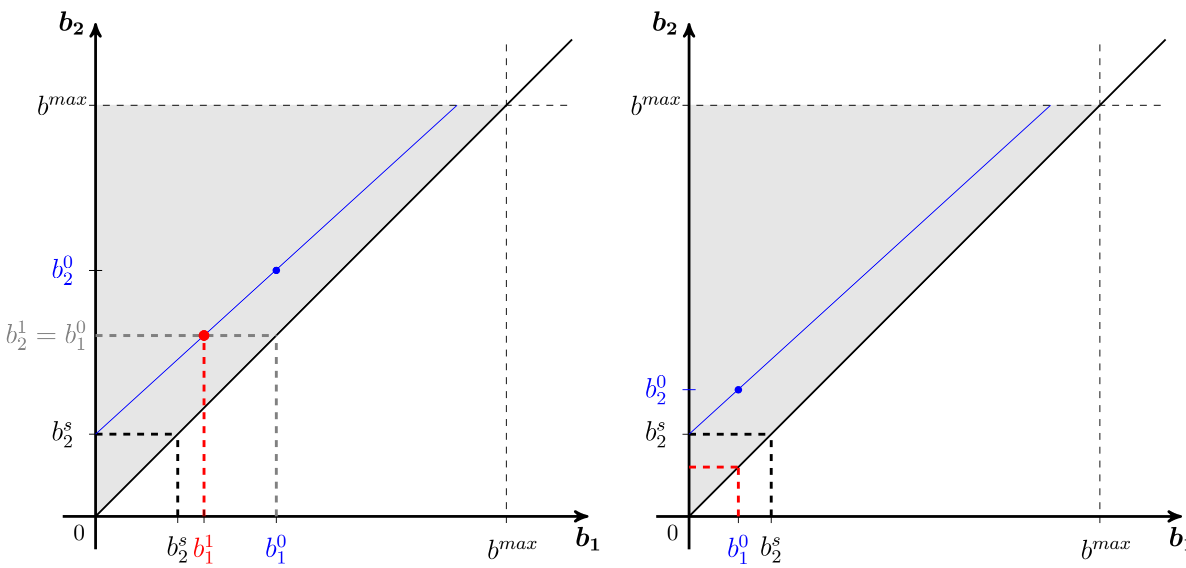
$$b_1^* < \left(\frac{8}{15}\right) b^{max}. \tag{19}$$

Proof. See Appendix A14. □

The intuition is illustrated in Figure 7, where Panel (a) shows a feasible and profitable deviation for the second mover, which is infeasible in Panel (b). The default risk of the lower-risk and higher-risk insurer is depicted on the x- and y-axis, respectively. In Panel (a), consider the default risk pair $b_1^0 < b_2^0$ as indicated by the blue dot. Other default risk pairs leading to the same price equilibrium lie on straight lines, as shown by Proposition 6 and indicated by the blue line. A profitable deviation for the second mover is to choose a default risk that leads to the same price equilibrium but with reversed roles (the reversal of roles is indicated in Panel (a) by the grey dotted lines, while the profitable deviation for the second mover is indicated by the red dotted line). Such profitable deviations are infeasible for default risk choices of the first mover below b_2^s — as illustrated in Panel (b). The risk level that rules out this profitable deviation is characterized by i) or (19) in Proposition 7.

Figure 7: Competitive Mechanism

(a) (b_1^1, b_2^1) Leads to Reversed Roles **(b)** Profitable Deviation in (a) Infeasible



Notes: Each panel depicts insurer 1's and 2's default risks on the x- and y-axis, respectively. See the text for a detailed explanation.

Proposition 7 implies that the maximal default risk to ensure the position of quality leader is smaller than the smallest optimal default risk gap (at $b_1 = 0$.) This is approximately equal to the smallest optimal default risk gap at other b_1, since the blue line is almost parallel to the 45-degree line. Hence, the smallest optimal default risk gap is approximately an upper bound for the default risk of the first mover, and therefore, twice the default risk gap is approximately an upper bound for the second-mover's default risk.

Under two conditions that are simple but probably more restrictive than necessary, the competitive situation can be summed up based on the default risk gap.

Proposition 8 (Push-and-Pull Effect for Second Mover). *Suppose assumptions A1 - A6 and that the following two conditions hold*

$$\Pi_2^{\square}(0, b_2) \text{ as a function of } b_2 \text{ has a unique maximum for a } b_2 < b^{max}, \tag{N1}$$

$$\Pi_1^{\square}(0, b_2) \text{ as a function of } b_2 \text{ is increasing in } b_2. \tag{N2}$$

Let Π_2^* *be the global maximum of* $\Pi_2^{\square}(0, b_2)$*, assumed at* b_2^s*. Let* $\bar{b_1}$ *be the minimum of all* b_1 *such that* $\Pi_1^{\square}(0, b_1) = \Pi_2^*$*. Then,* (b_1^*, b_2^*) *is a subgame-perfect equilibrium iff*

$$b_1^* \in [0, \bar{b}_1] \tag{20}$$

$$b_2^* = (1 - \tilde{g}(0, b_2^s))b_1^* + b_2^s. \tag{21}$$

The second-mover's choice of default risk is pinned down by the first-mover's choice plus the optimal default risk gap. The first-mover's choice of default risk thus exerts a push-and-pull effect on the second-mover's choice of default risk.

Proof. See Appendix A15. □

Proposition 8 suggests that the first mover chooses a default risk pinned down by the default risk gap, and the second mover follows at an optimal distance. This is in contrast to the result of maximal product differentiation as in the standard model in Tirole (1988), which closely follows Shaked and Sutton (1982). In Tirole (1988), two firms compete in quality (chosen first) and price (chosen second) for clients that differ in their valuation of quality. The key mechanism is that for any two pairs of quality choices, firms choose prices in such a way that the resulting market shares remain unchanged. This eliminates a quantity effect, and with only a price effect left, firms soften price competition as much as possible by choosing maximally differentiated qualities. The result of maximal differentiation in qualities in the standard model hinges on three assumptions: first, clients' utility is linear; second, it is assumed that the market is always fully covered; and third, costs are quality-invariant. In the present model, the main departure from the standard model is the non-linearity of the utility function stemming from risk aversion in the insurance context. As a result, market shares are no longer invariant for varying quality pairs, and we obtain market discipline in quality choices. In the Online Appendix, I revisit the standard model and show that one can obtain a similar result in the standard model with linear utility when removing the assumptions of full market coverage and introducing (general) convex costs for quality provision.

While conditions (N1) and (N2) cannot hold in general, e.g., a parameter value of b^{max} sufficiently small may violate (N1), I conjecture that they hold for a wide range of parameters. They hold in a numerical example for plausible parameter values, as shown in the following section.

6 Numerical Example

In a numerical example with plausible parameter values, I explicitly characterize the subgame-perfect Nash equilibria and demonstrate that the default risk gap can indeed be small.

Parameter Values. Consider the model for a specific set of parameters, namely

$$\underline{\theta} = -100 \cdot 10^6 \tag{22}$$

$$p = 0.03 \tag{23}$$

$$\underline{a}(-\underline{\theta}) = 4.5 \tag{24}$$

$$\overline{a} = \frac{3}{2}\underline{a} \tag{25}$$

$$\gamma^{max} = 33 \cdot 10^6 \tag{26}$$

$$b^{max} = \frac{1}{3} \tag{27}$$

(22) and (23) correspond to a scenario with a large rare loss, e.g., a 100 million loss from a sudden movement in exchange rates that occurs every 33 years on average. (24), (25), (26) and (27) are chosen in the simplest way such that assumptions A4, A5, A3 and A2, respectively, are satisfied.

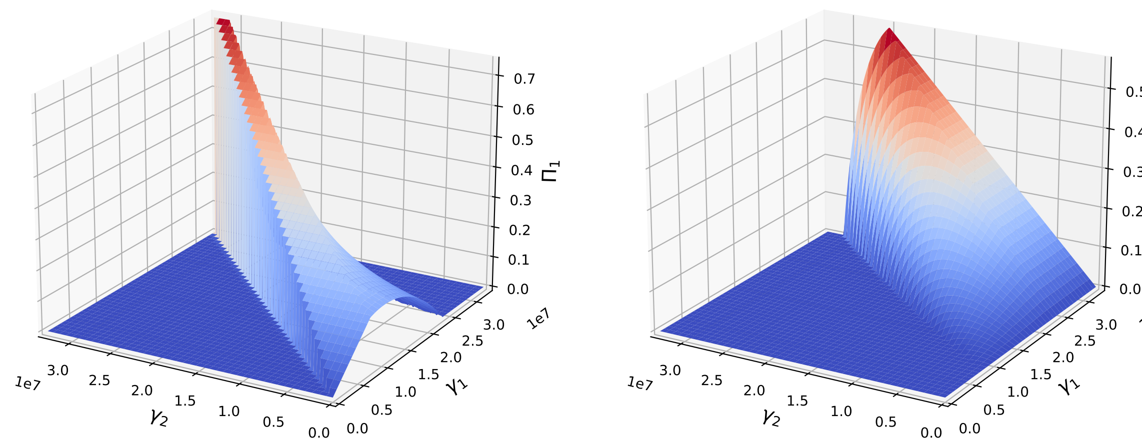
Based on Proposition 6, we first consider $b_1 = 0$.

Solving for the price equilibrium for $(0, b_2)$ *for some fixed* b_2*.* For $(0, b_2)$, we numerically solve for the indifferent client as a function of prices (γ_1, γ_2). As an illustration, for $b_2 = 0.15$,

Figure 8 shows the resulting profit functions for both insurers.

Figure 8: Illustration of Insurers' Profit Functions at $b_2 = 0.15$

(a) Insurer 1 **(b)** Insurer 2



Notes: The figure shows insurer 1's (Panel (a)) and insurer 2's (Panel (b)) profits (z-axis) as functions of prices (γ_1 on the x-axis, γ_2 on the y-axis) in a numerical example. Profit functions are drawn for the following vector of default risks: $b_0 = 0, b_1 = 0.15$. The parameter values used in the numerical example are listed in the text.

Equilibrium profits for $(0, b_2)$ *as a function of* b_2*.* We then solve for price equilibria (and subsequently profits) for a range of b_2. Figure 9 shows the resulting equilibrium profits for both insurers as a function of b_2. In particular, the second mover's profit as a function of b_2 has a unique interior maximum, while the first mover's profit as a function of b_2 is increasing. That is, conditions (N1) and (N2) hold. Additionally, insurer 1's are an order of magnitude larger than insurer 2's.

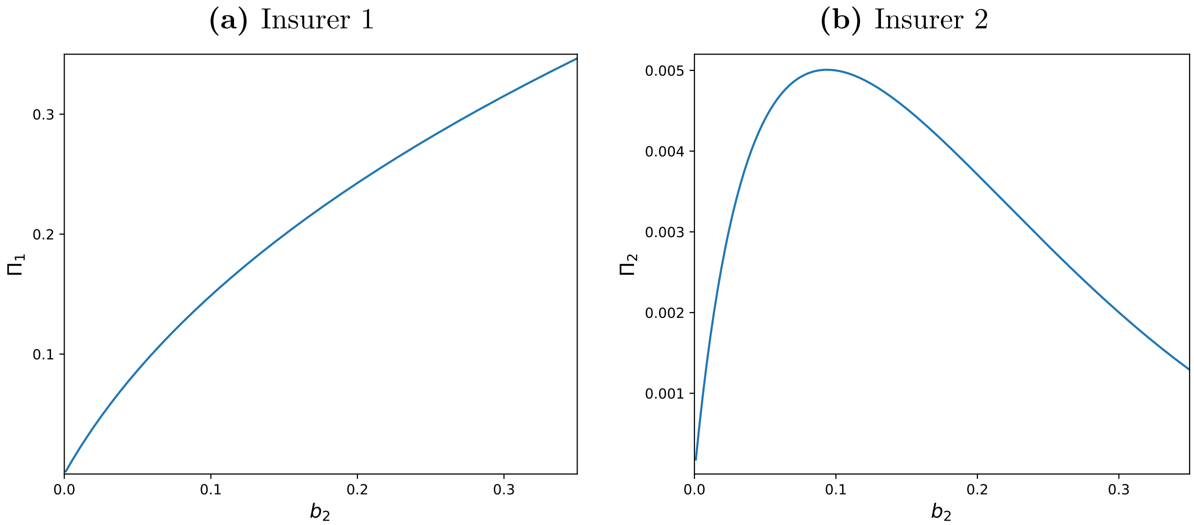
Equilibrium qualities. We then calculate $\bar{b}_1 \approx 0.0023$, hence the resulting equilibrium default risks are

$$b_1^* \in [0, 0.0023] \tag{28}$$

$$b_2^* = 0.9972\, b_1^* + 0.0937. \tag{29}$$

In particular, the first mover chooses a default risk close zero, and the second mover follows at a distance that equals the optimal default risk gap. This default risk gap is lower than a third of the largest admissible default risk in the model setup.

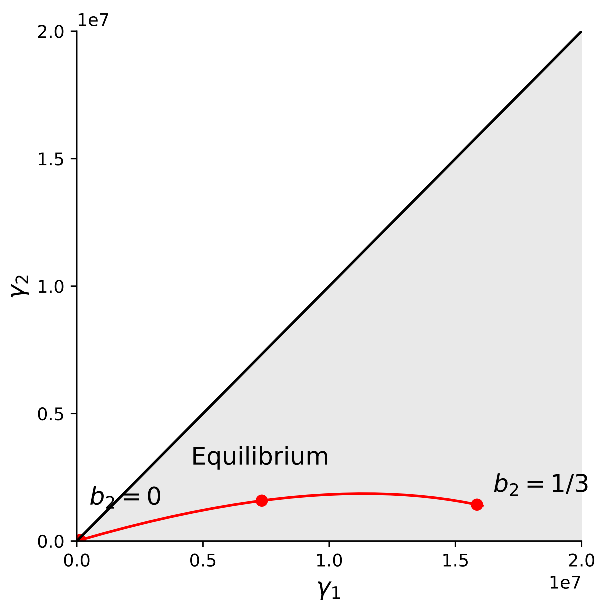
Figure 9: Equilibrium Profits as a Function of b_2



Notes: The figure shows insurer 1's (Panel (a)) and insurer 2's (Panel (b)) profits (y-axis) as functions of the default risk gap in a numerical example. As motivated in the text, b_1 is fixed at zero and insurer 2's default risk, b_2, is depicted on the x-axis. The parameter values used in the numerical example are listed in the text.

Equilibrium prices are depicted in Figure 10.

Figure 10: Equilibrium Prices



Notes: The figure shows optimal prices at varying default risk gaps in a numerical example. Insurer 1's and insurer 2's prices are depicted on the x- and y-axis, respectively. The red line depicts pairs of optimal prices for $(0, b_2)$ with b_2 ranging in $[0, 1/3]$, with the prices at the equilibrium in default risks marked. The parameter values used in the numerical example are listed in the text.

7 Application to Derivatives Markets

An application of the model setup is the derivatives market, as it naturally maps key features of these markets. First, the stylized insurance contract considered in the model is typical for derivatives markets: the contract features full coverage against a macro risk (as does, e.g., a plain-vanilla interest rate swap) but comes with the risk of counterparty default. Thus, both the price of the derivative and the counterparty risk may influence the purchase decision. Second, derivatives markets have a hub-and-spoke structure with numerous clients with differing risk attitudes seeking insurance from a small set of large banks (called *dealers*) – which aligns with the model structure. Third, dealers choose their own default risk, e.g., by setting aside capital, choosing liquidity buffers, or maintaining balanced trading books. Gregory (2014, p.135), for example, details how an institution's creditworthiness as assessed by ratings plays a role, as well as its capital base, liquidity, and operational requirements for processing trades.

Introducing a central counterparty (CCP) in derivatives markets raises questions about its impact on competition among dealers. In a centrally cleared market, a CCP interposes itself between a buyer and a seller, replacing the existing contract between them with two contracts that each have the CCP on one end. It thereby insulates the contracting parties from the risk that the counterparty defaults. CCPs can support financial stability through netting, enforcing margining and improving transparency for better regulatory oversight. However, the effects on competition in a highly concentrated market are little understood.

The starting point of this project was the observation that in a centrally cleared market unless there is a default, the client-dealer relationship remains largely unchanged because clients do not directly interact with the CCP. Only members of the CCP can directly clear with the CCP, while most market participants access clearing services through members (CPMI, IOSCO, 2022). Consider a trade where a market participant buys a derivative from a dealer. Suppose the common situation in which the dealer is a member of the CCP and not only the executing broker of the trade but also the client's clearing service provider. Then, the resulting flow is: client — client account at the dealer – CCP — house account dealer. Thus, from the dealer's perspective, client clearing changes little as the CCP protects the dealer from its own default while the client still buys the derivative from the dealer.

However, central clearing alters the nature of competition between dealers, and my model is able to clarify a market force that may be absent in a centrally cleared market. A CCP facilitates porting arrangements, meaning that if a clearing member defaults, clients' portfolios are transferred to another solvent member (Braithwaite, 2016; Braithwaite and Murphy, 2020). The success of the London Clearing House (LCH) during the Global Financial Crisis can largely be attributed to such porting arrangements. From the client's perspective, there is no longer differentiation in contract continuity between dealers, eliminating this quality dimension of competition. Viewed through the lens of the model, the market force that incentivizes dealers to choose a low default risk – beyond requirements mandated by regulation – may then be absent: This paper shows that with two-dimensional competition in price and default risk, market discipline in the choice of default risks emerges. Without perceived differences in default risks, pure price competition prevails.

8 Conclusion

I study precautionary measures insurance sellers undertake to ensure their solvency within a model of vertical product differentiation. To that end, I introduce the seller's default risk as a quality dimension of the insurance product. Analogous to standard analyses of vertical product differentiation, I show that more risk-averse clients self-select to buy from the dealer with the lower default risk, leading to market segmentation and higher profits for the dealer with the lower default risk. The key insight from the model is that competition in two dimensions (price, default risk) gives rise to market discipline in insurers' default risk choices: the first over in the choice of default risk chooses a low default risk, and the second mover follows suit.

I discuss the model implications for competition in derivatives markets. The result highlights a market force that may be absent in a centrally cleared market where dealers compete for clients but are insulated from competition in default risk.

A central counterparty in the model framework is conceptualized ad-hoc and not formally introduced, leaving many aspects of central clearing (e.g., loss-sharing mechanisms, margins, CCP's default probability) beyond the scope of the current model. Retaining the simple framework that maps the market structure with client clearing and incorporates risk aversion while modeling a CCP in more detail is left for future research.

APPENDIX

A Appendix: Proofs

Remark 1. I use the following notation. For functions $G : \mathbb{R}^3 \to \mathbb{R}$ and $H : \mathbb{R}^2 \to \mathbb{R}^3$,

$$D(G \circ H) = (\partial_1(G \circ H), \partial_2(G \circ H)) =: (d_1 G, d_2 G)\,.$$

This is to indicate that the chain rule on the composite function is considered although we write G instead of the composite function $(G \circ H)$ for brevity. Here, typically $H : (\gamma_1, \gamma_2) \mapsto (a^*(\gamma_1, \gamma_2), \gamma_1, \gamma_2)$.

A1 Proof of Lemma 1

For the indifferent client we have

$$
\begin{aligned}
U_a(b_1, \gamma_1) &= U_a(b_2, \gamma_2) && \text{(A3)}\\
\Leftrightarrow (1 - b_1 p) u_a(-\gamma_1) + b_1 p u_a(\underline{\theta}) &= (1 - b_2 p) u_a(-\gamma_2) + b_2 p u_a(\underline{\theta}) && \text{(A4)}\\
\Leftrightarrow u_a(-\gamma_1) - u_a(-\gamma_2) + p\,[b_2 u_a(-\gamma_2) - b_1 u_a(-\gamma_1)] &= p \Delta b u_a(\underline{\theta}) && \text{(A5)}\\
\Leftrightarrow [u_a(-\gamma_1) - u_a(-\gamma_2)]\,(1 - b_1 p) &= p \Delta b\,[u_a(\underline{\theta}) - u_a(-\gamma_2)] && \text{(A6)}\\
\Leftrightarrow \frac{u_a(-\gamma_1) - u_a(-\gamma_2)}{u_a(\underline{\theta}) - u_a(-\gamma_2)} &= \frac{p \Delta b}{1 - b_1 p} && \text{(A7)}\\
\Leftrightarrow \frac{\exp(-a\Delta\gamma) - 1}{\exp(-a(\underline{\theta} + \gamma_2)) - 1} &= \frac{p \Delta b}{1 - b_1 p}. && \text{(A8)}
\end{aligned}
$$

□

A2 Proof of Proposition 1

ad i). The proof proceeds by showing that $\partial_a g < 0$. Suppose this was true. Then the LHS of (3) is monotonically decreasing, while the RHS of (3) is fixed, yielding at most one solution.

Claim. $\partial_a g < 0$.

Proof of claim. For the derivative of the function g with respect to a we get

$$
\begin{aligned}
\frac{\partial g(a)}{\partial a} &= \frac{-\Delta\gamma \exp(-a\Delta\gamma)\,(\exp(-a(\underline{\theta} + \gamma_2)) - 1)}{(\exp(-a(\underline{\theta} + \gamma_2)) - 1)^2} && \text{(A9)}\\
&\quad + \frac{(\exp(-a\Delta\gamma) - 1)\,(\underline{\theta} + \gamma_2) \exp(-a(\underline{\theta} + \gamma_2))}{(\exp(-a(\underline{\theta} + \gamma_2)) - 1)^2}\\
&= \frac{1}{(\exp(-a(\underline{\theta} + \gamma_2)) - 1)^2} \Bigg[\exp(-a\Delta\gamma) \Bigg(- \Delta\gamma\,(\exp(-a(\underline{\theta} + \gamma_2)) - 1) && \text{(A10)}
\end{aligned}
$$

$$+(\underline{\theta}+\gamma_2)\exp(-a(\underline{\theta}+\gamma_2))\Big) - (\underline{\theta}+\gamma_2)\exp\left(-a(\underline{\theta}+\gamma_2)\right)\Big]$$

$$= \frac{1}{\left(\exp(-a(\underline{\theta}+\gamma_2))-1\right)^2} \tag{A11}$$
$$\left[\exp(-a\Delta\gamma)\left(\exp\left(-a(\underline{\theta}+\gamma_2)\right)(\underline{\theta}+\gamma_1)+\Delta\gamma\right) - (\underline{\theta}+\gamma_2)\exp(-a(\underline{\theta}+\gamma_2))\right]$$

$$= \underbrace{\frac{\exp(-a\Delta\gamma)}{\left(\exp(-a(\underline{\theta}+\gamma_2))-1\right)^2}}_{>0} \tag{A12}$$
$$\left[\underbrace{\Delta\gamma}_{<0} + \underbrace{\exp\left(-a(\underline{\theta}+\gamma_1)\right)}_{>0}\underbrace{\left(\exp(-a\Delta\gamma)(\underline{\theta}+\gamma_1)-(\underline{\theta}+\gamma_2)\right)}_{:=f(a)}\right]$$

using that

$$\exp(-a(\underline{\theta}+\gamma_2)) = \exp(-a(\underline{\theta}+\gamma_1))\exp(-a\Delta\gamma). \tag{A13}$$

Then

$$f(a) < 0 \Rightarrow \frac{\partial g(a)}{\partial a} < 0. \tag{A14}$$

We have

$$f(a) = \exp(-a\Delta\gamma)(\underline{\theta}+\gamma_1) - (\underline{\theta}+\gamma_2) < 0 \tag{A15}$$
$$\Leftrightarrow \exp(-a\Delta\gamma)(\underline{\theta}+\gamma_1) < (\underline{\theta}+\gamma_2) \tag{A16}$$
$$\Leftrightarrow \frac{\exp(-a(\underline{\theta}+\gamma_2))}{\exp(-a(\underline{\theta}+\gamma_1))}(\underline{\theta}+\gamma_1) < (\underline{\theta}+\gamma_2) \tag{A17}$$
$$\Leftrightarrow \frac{\exp(-a(\underline{\theta}+\gamma_2))}{(\underline{\theta}+\gamma_2)} < \frac{\exp(-a(\underline{\theta}+\gamma_1))}{(\underline{\theta}+\gamma_1)}. \tag{A18}$$

For $x < 0$ the function

$$h(x) := \frac{\exp(-ax)}{x} \tag{A19}$$

is negative and

$$h'(x) = h(x)\left[-a - \frac{1}{x}\right] > 0 \;\;\Leftrightarrow\;\; a + \frac{1}{x} > 0 \;\;\Leftrightarrow\;\; a(-x) > 1. \tag{A20}$$

For $x = \underline{\theta} + \gamma$ this is true from assumption A3. Since $\underline{\theta}+\gamma_2 < \underline{\theta}+\gamma_1$, (A18) indeed holds and proves the claim.

ad ii). With $g(\cdot, \vec{\gamma})$ strictly decreasing, existence under (5) follows immediately.

ad iii). A client with risk aversion parameter a chooses insurer 1 if

$$
\begin{aligned}
U_a(b_1,\gamma_1) &> U_a(b_2,\gamma_2) && \text{(A21)}\\
\Leftrightarrow (1-b_1p)u_a(-\gamma_1)+b_1pu_a(\underline{\theta}) &> (1-b_2p)u_a(-\gamma_2)+b_2pu_a(\underline{\theta}) && \text{(A22)}\\
\Leftrightarrow \left[u_a(-\gamma_1)-u_a(-\gamma_2)\right](1-b_1p) &> p\Delta b\underbrace{\left(u_a(\underline{\theta})-u_a(-\gamma_2)\right)}_{<0} && \text{(A23)}\\
\Leftrightarrow \frac{u_a(-\gamma_1)-u_a(-\gamma_2)}{u_a(\underline{\theta})-u_a(-\gamma_2)} &< \frac{p\Delta b}{1-b_1p} && \text{(A24)}\\
\Leftrightarrow g(a) &< g(a^*) && \text{(A25)}\\
\Leftrightarrow a &> a^*(\gamma_1,\gamma_2). && \text{(A26)}
\end{aligned}
$$

□

A3 Proof of Lemma 3

The idea is to proceed analogously to the proof of Proposition 1, but add and subtract $b_2u_a(-\gamma_1)$ instead of $b_1u_a(-\gamma_2)$. Namely, for the indifferent client we have

$$
\begin{aligned}
U_a(b_1,\gamma_1) &= U_a(b_2,\gamma_2) && \text{(A27)}\\
\Leftrightarrow u_a(-\gamma_1)-u_a(-\gamma_2)+p\left[b_2u_a(-\gamma_2)-b_1u_a(-\gamma_1)\right] &= p\Delta b u_a(\underline{\theta}) && \text{(A28)}\\
\Leftrightarrow \left[u_a(-\gamma_1)-u_a(-\gamma_2)\right](1-b_2p) &= p\Delta b\left[u_a(\underline{\theta})-u_a(-\gamma_1)\right] && \text{(A29)}\\
\Leftrightarrow \frac{u_a(-\gamma_1)-u_a(-\gamma_2)}{u_a(\underline{\theta})-u_a(-\gamma_1)} &= \frac{p\Delta b}{1-b_2p} && \text{(A30)}\\
\Leftrightarrow \frac{1-\exp(-(-a\Delta\gamma))}{\exp(-a(\underline{\theta}+\gamma_1))-1} &= \frac{p\Delta b}{1-b_2p}. && \text{(A31)}
\end{aligned}
$$

□

A4 Proof of Lemma 2

An auxiliary lemma offers a second characterization of the indifferent client, symmetric to the one in Lemma 1, and exploiting this symmetry will be key in the sequel.

Lemma 3. *The client who is indifferent between two contracts* (b_1,γ_1) *and* (b_2,γ_2) *with* $\Delta b>0$*, has a second characterization*

$$
h(a,\vec{\gamma}) := \frac{1-\exp(-(-a\Delta\gamma))}{\exp(-a(\underline{\theta}+\gamma_1))-1} = \frac{p\Delta b}{1-b_2p}. \tag{A32}
$$

Proof.
The idea is to proceed analogously to the proof of Proposition 1, but add and subtract $b_2u_a(-\gamma_1)$ instead of $b_1u_a(-\gamma_2)$. Namely, for the indifferent client we have

$$
U_a(b_1,\gamma_1) = U_a(b_2,\gamma_2) \tag{A33}
$$

$$\Leftrightarrow u_a(-\gamma_1) - u_a(-\gamma_2) + p\left[b_2 u_a(-\gamma_2) - b_1 u_a(-\gamma_1)\right] = p\Delta b u_a(\underline{\theta}) \tag{A34}$$

$$\Leftrightarrow \left[u_a(-\gamma_1) - u_a(-\gamma_2)\right](1 - b_2 p) = p\Delta b\left[u_a(\underline{\theta}) - u_a(-\gamma_1)\right] \tag{A35}$$

$$\Leftrightarrow \frac{u_a(-\gamma_1) - u_a(-\gamma_2)}{u_a(\underline{\theta}) - u_a(-\gamma_1)} = \frac{p\Delta b}{1 - b_2 p} \tag{A36}$$

$$\Leftrightarrow \frac{1 - \exp(-(-a\Delta\gamma))}{\exp(-a(\underline{\theta} + \gamma_1)) - 1} = \frac{p\Delta b}{1 - b_2 p}. \tag{A37}$$

□

Following the notation as noted in the remark at the beginning of the appendix, $\partial_i a^* = d_i a^*$. We show three claims from which Lemma 2 directly follows.

Claim 1. $a^* = \tau_1(d_1 a^*)$ with $\tau_1, (d_1 a^*) > 0$.

Claim 2. $a^* = \tau_2(-d_2 a^*)$ with $\tau_2, (-d_2 a^*) > 0$.

Claim 3. $(-d_2 a^*)/(d_1 a^*) =: \alpha = (1 - gB_1) = 1/(1 + hB_2) = \tau_1/\tau_2 < 1$.

Proof of claim 1. For the function $g(a^*(\vec{\gamma}), \vec{\gamma})$, as defined in (3), we have from Proposition 1

$$0 = d_1 g = \partial_1 g|_{a=a^*} + \partial_a g|_{a=a^*} \cdot d_1 a^* \tag{A38}$$

$$\Leftrightarrow d_1 a^* = \frac{-\partial_1 g|_{a=a^*}}{\partial_a g|_{a=a^*}}. \tag{A39}$$

In the following write $\partial_i g$ shorthand for $\partial_i g|_{a=a^*}$. We have

$$\partial_1 g = a^* \frac{A}{B_2 - 1} > 0. \tag{A40}$$

and from Proposition 1 we know that $\partial_a g < 0$. Hence, in light of (A39), we have $d_1 a^* > 0$.

Further, note that the expression for $\partial_a g$, derived in the proof of Proposition 1, can be written in short-hand notation as follows

$$\partial_a g = \frac{A}{(B_2 - 1)}\left[-\Delta\gamma + (\underline{\theta} + \gamma_2)\underbrace{\frac{(A-1)}{(B_2 - 1)}}_{=g}\underbrace{\frac{B_2}{A}}_{=B_1}\right] \overset{\text{(A40)}}{=} \frac{\partial_1 g}{a^*}\left[-\Delta\gamma + g\varphi_1\right]. \tag{A41}$$

Inserted into (A38) this yields

$$0 = \partial_1 g + \frac{\partial_1 g}{a^*}(-\Delta\gamma + g\varphi_1) d_1 a^* \tag{A42}$$

$$= \underbrace{\frac{\partial_1 g}{a^*}}_{>0}\left[a^* + (-\Delta\gamma + g\varphi_1) d_1 a^*\right]. \tag{A43}$$

Hence

$$a^* = (\Delta\gamma - g\varphi_1)\underbrace{d_1 a^*}_{>0}, \tag{A44}$$

and subsequently

$$\tau_1 = (\Delta\gamma - g\varphi_1) > 0. \tag{A45}$$

Proof of claim 2. Analogously, for the function $h(a^*(\vec{\gamma}), \vec{\gamma})$, as defined in (A32), we have

$$0 = d_2 h = \partial_2 h|_{a=a^*} + \partial_a h|_{a=a^*} \cdot d_2 a^* \tag{A46}$$

$$\Leftrightarrow d_2 a^* = \frac{-\partial_2 h|_{a=a^*}}{\partial_a h|_{a=a^*}}. \tag{A47}$$

Similar to before we write $\partial_i h$ shorthand for $\partial_i h|_{a=a^*}$. Then we have

$$\partial_2 h = (-a^*)\frac{1}{A(B_1 - 1)} < 0, \tag{A48}$$

and

$$\partial_a h = (-\Delta\gamma)\frac{1}{A(B_1 - 1)} + (\underline{\theta} + \gamma_1)\frac{(1 - \frac{1}{A})B_1}{(B_1 - 1)^2} \tag{A49}$$

$$= \frac{1}{A(B_1 - 1)^2}\left[\Delta\gamma - \Delta\gamma B_1 - (\underline{\theta} + \gamma_1)B_1 + (\underline{\theta} + \gamma_1)AB_1\right] \tag{A50}$$

$$= \underbrace{\frac{1}{A(B_1 - 1)^2}}_{>0}\left[\underbrace{\Delta\gamma}_{<0} + \underbrace{B_1}_{>0}\Big(A(\underline{\theta} + \gamma_1) - (\underline{\theta} + \gamma_2)\Big)\right]. \tag{A51}$$

From the proof of Proposition 1 we know that $A(\underline{\theta} + \gamma_1) - (\underline{\theta} + \gamma_2)$ is negative, hence $\partial_a h < 0$. Then from (A47) we get $d_2 a^* < 0$.

For the remaining part, note that $AB_1 = B_2$ and hence (A49) can also be written as

$$\partial_a h = \frac{1}{A(B_1 - 1)}\left[-\Delta\gamma + (\underline{\theta} + \gamma_1)AB_1\frac{(1 - \frac{1}{A})B_1}{(B_1 - 1)}\right] \tag{A52}$$

$$= \frac{\partial_a h}{a^*}\left[\Delta\gamma - \varphi_2 h\right]. \tag{A53}$$

Inserted into (A46) this yields

$$0 = \underbrace{\frac{\partial_2 h}{a^*}}_{<0}\left[a^* + (\Delta\gamma - \varphi_2 h)d_2 a^*\right]. \tag{A54}$$

Hence,

$$a^* = -(\Delta\gamma - \varphi_2 h)\underbrace{d_2 a^*}_{<0}, \tag{A55}$$

and subsequently

$$\tau_2 = (\Delta\gamma - \varphi_2 h) > 0. \tag{A56}$$

Proof of claim 3. We first establish that

$$(1 - gB_1) = \frac{B_1 - 1}{B_2 - 1} = \frac{1}{(1 + hB_2)}. \tag{A57}$$

This follows, since from the definition

$$1 - gB_1 = 1 - \frac{A-1}{B_2 - 1}\frac{B_2}{A} = \frac{B_2 - A}{A(B_2 - 1)} = \frac{B_1 - 1}{B_2 - 1} \tag{A58}$$

$$1 + hB_2 = 1 + \frac{1 - \frac{1}{A}}{B_1 - 1}B_2 = \frac{B_1 - 1 + B_2 - \frac{B_2}{A}}{B_1 - 1} = \frac{B_2 - 1}{B_1 - 1}. \tag{A59}$$

In light of (16) and (17) we have

$$\begin{aligned}
\alpha &= \frac{\Delta\gamma - g\varphi_1}{\Delta\gamma - h\varphi_2} && \text{(A60)}\\
&= \frac{(\underline{\theta} + \gamma_2) - (\underline{\theta} + \gamma_1) - g\varphi_1}{(\underline{\theta} + \gamma_2) - (\underline{\theta} + \gamma_1) - h\varphi_2} && \text{(A61)}\\
&= \frac{(\underline{\theta} + \gamma_2)(1 - gB_1) - (\underline{\theta} + \gamma_1)}{-(\underline{\theta} + \gamma_1)(1 + hB_2) + (\underline{\theta} + \gamma_2)} && \text{(A62)}\\
&= \frac{(1 - gB_1)\left((\underline{\theta} + \gamma_2) - \frac{1}{(1 - gB_1)}(\underline{\theta} + \gamma_1)\right)}{(\underline{\theta} + \gamma_2) - (1 + hB_2)(\underline{\theta} + \gamma_1)} && \text{(A63)}\\
&= (1 - gB_1), && \text{(A64)}
\end{aligned}$$

which concludes the proof.

□

A5 Auxiliary Properties

Proposition 9. *As always, we consider the set $\mathcal{G}_{[\underline{a}, \overline{a}]}$. Then the following properties hold*

$$\begin{aligned}
d_2^2 a^* &= \frac{(-d_2 a^*)}{\tau_2}\left[2 + h\varphi_2 a^*\Big(1 - \frac{\xi_2}{\tau_2}\Big)\right] < 0 && \text{(A65)}\\
d_1 d_2 a^* &= (d_1 a^*)^2 \frac{\alpha}{a^*}\left[a^* \xi_2 \frac{h\varphi_2}{\tau_2} - 2\right] > 0 && \text{(A66)}
\end{aligned}$$

$$
\begin{aligned}
d_1^2 a^* &= \left[\frac{2}{a^*} - g\varphi_1\xi_2\frac{\alpha}{\tau_1}\right] &\text{(A67)}\\
d_2^2 \Pi_2 &= (d_2 a^*)\left[2 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_1)h\varphi_2}{\tau_1} - 2\right)\right] < 0 &\text{(A68)}\\
d_1 d_2 \Pi_2 &= (d_1 a^*)\left[1 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2\right)\right] > 0 &\text{(A69)}\\
d_1^2 \Pi_1 &= (-d_1 a^*)\left[2 + \frac{\gamma_1}{\tau_1}\left(2 - \frac{a^*\xi_2 g\varphi_1}{\tau_2}\right)\right] &\text{(A70)}\\
d_1 d_2 \Pi_1 &= (-d_2 a^*)\left[1 + \frac{\gamma_1}{\tau_1}\left(2 - \frac{a^*\xi_2 h\varphi_2}{\tau_2}\right)\right] &\text{(A71)}\\
d_1^2\Pi_1 + \frac{1}{\alpha}d_1 d_2 \Pi_1 &< 0,\ \textit{hence } d_1^2\Pi_1 \neq 0 \vee d_1 d_2\Pi_1 \neq 0. &\text{(A72)}
\end{aligned}
$$

Proof.
ad $d_2^2 a^*$. Note that

$$\frac{\xi_1}{\tau_1} - \frac{\xi_2}{\tau_2} = \frac{1}{\tau_1}\underbrace{[\xi_1 - \alpha\xi_2]}_{=-\tau_1} = -1 \tag{A73}$$

We have,

$$
\begin{aligned}
d_2^2 a^* &= d_2\left(-\frac{a^*}{\tau_2}\right) &\text{(A74)}\\
&= -\frac{d_2 a^*}{\tau_2} + a^*\frac{1}{\tau_2^2}d_2\tau_2 &\text{(A75)}\\
&= -\frac{d_2 a^*}{\tau_2}[1 + d_2\tau_2] &\text{(A76)}\\
&= -\frac{d_2 a^*}{\tau_2}[1 + 1 + h\xi_1 B_2(a^* + (d_2 a^*)\xi_2)] &\text{(A77)}\\
&= -\frac{d_2 a^*}{\tau_2}\left[2 + h\varphi_2 a^*\left(1 - \frac{\xi_1}{\tau_2}\right)\right] &\text{(A78)}\\
&= -\frac{d_2 a^*}{\tau_2}\left[2 + h\varphi_2 a^*\frac{\xi_1}{\tau_1}\right] &\text{(A79)}\\
&= \underbrace{(d_2 a^*)}_{<0}\frac{1}{\tau_2^2}\left[-2\tau_2 + h\varphi_2(-a^*\xi_1)\frac{1}{\alpha}\right] &\text{(A80)}
\end{aligned}
$$

$d_2^2 a^*$ is negative iff

$$-a^*\xi_1 > 2\underbrace{\frac{\tau_2}{-h\varphi_2}}_{\in(0,1)}\underbrace{\alpha}_{<1}, \tag{A81}$$

which is ensured by $-\underline{a}\xi_1 > 2$ from assumption A3.

ad $d_1 d_2 a^*$. We have

$$\begin{aligned} d_1\varphi_2 &= B_2 + \xi_1 d_1 B_2 && \text{(A82)}\\ &= B_2 + \xi_1 B_2(-d_2 a^*)\xi_2 && \text{(A83)}\\ &= B_2\left((1 - \xi_1\xi_2 d_1 a^*\right) && \text{(A84)} \end{aligned}$$

Then

$$\begin{aligned} d_1 d_2 a^* &= -d_1\left(\frac{a^*}{\tau_2}\right) && \text{(A85)}\\ &= -\frac{d_1 a^*\tau_2 - a^*(-1 - hB_2(1 - \xi_1\xi_2 d_1 a^*)}{\tau_2^2} && \text{(A86)}\\ &= \frac{-d_1 a^*\left[\tau_2 - a^*\xi_1\xi_2 hB_2\right] + a^*(1 + hB_2)}{\tau_2^2} && \text{(A87)}\\ &= \underbrace{\frac{-d_1 a^*}{\tau_2^2}}_{<0}\underbrace{\Big[\tau_2 - a^*\xi_1\xi_2 hB_2 + \tau_1(1 + hB_2)\Big]}_{=:W} && \text{(A88)} \end{aligned}$$

Hence, $d_1 d_2 a^* > 0$ if the expression in brackets is negative. This is indeed the case, since

$$\begin{aligned} W &= 2\Delta\gamma + hB_2\left[\tau_1 - \xi_1 - a^*\xi_1\xi_2\right] - g\varphi_1 && \text{(A89)}\\ &= \Delta\gamma(2 + hB_2) - \xi_1 hB_2 \underbrace{(1 + a^*\xi_2)}_{=-1+(2+a^*(\underline{\theta}+\gamma_2))} - g\varphi_1(1 + hB_2) && \text{(A90)}\\ &= hB_2\left[\xi_1 + \Delta\gamma\right] - g\xi_2 B_1(1 + hB_2) - \xi_1 hB_2(2 + a^*\xi_2) + 2\Delta\gamma && \text{(A91)}\\ &= \xi_2\underbrace{(hB_2 - gB_1(1 + hB_2))}_{=\frac{\Delta B}{B_1-1}-\frac{\Delta B}{B_2-1}\frac{B_1-1}{B_2-1}=0} - \xi_1 hB_2(2 + a^*\xi_2) + 2\Delta\gamma && \text{(A92)}\\ &= \underbrace{-\xi_1}_{>0} hB_2\underbrace{(2 + a^*\xi_2)}_{<0} + \underbrace{2\Delta\gamma}_{<0} && \text{(A93)}\\ &< 0, && \text{(A94)} \end{aligned}$$

which together yields

$$\begin{aligned} d_1 d_2 a^* &= \frac{(d_1 a^*)h\varphi_2}{\tau_2^2}\left(a^*\xi_2 + 2\frac{\tau_2}{(-h\varphi_2)}\right) && \text{(A95)}\\ &= \frac{(d_1 a^*)h\varphi_2}{\tau_2^2}\left[a^*\xi_2 + 2\frac{\tau_2}{(-h\varphi_2)}\right] && \text{(A96)}\\ &= (d_1 a^*)^2\frac{\alpha}{a^*}\left[a^*\xi_2\frac{h\varphi_2}{\tau_2} - 2\right]. && \text{(A97)} \end{aligned}$$

ad $d_1^2 a^*$. We know

$$d_1\varphi_1 = \varphi_1\xi_2\alpha(d_1 a^*), \qquad \text{(A98)}$$

hence

$$\begin{aligned}
d_1^2 a^* &= d_1\left(\frac{a^*}{\tau_1}\right) && \text{(A99)}\\
&= \frac{d_1 a^*}{\tau_1} + a^* d_1\left(\frac{1}{\tau_1}\right) && \text{(A100)}\\
&\overset{\text{(A98)}}{=} \frac{d_1 a^*}{\tau_1} - a^* \frac{1}{(\tau_1)^2}\Big[-1 + g\varphi_1\xi_2\alpha(d_1 a^*)\Big] && \text{(A101)}\\
&= \frac{d_1 a^*}{\tau_1}\Big[2 - g\varphi_1\xi_2\alpha(d_1 a^*)\Big] && \text{(A102)}\\
&= (d_1 a^*)\frac{a^*}{\tau_1}\left[\frac{2}{a^*} - g\varphi_1\xi_2\alpha\frac{d_1 a^*}{a^*}\right] && \text{(A103)}\\
&= (d_1 a^*)^2\left[\frac{2}{a^*} - g\varphi_1\xi_2\frac{\alpha}{\tau_1}\right]. && \text{(A104)}
\end{aligned}$$

ad $d_2^2\Pi_2$. Using (A66) and (A65),

$$\begin{aligned}
d_2^2\Pi_2 &= 2d_2 a^* + \gamma_2 d_2^2 a^* && \text{(A105)}\\
&= (d_2 a^*)\left[2 + \frac{\gamma_2}{\tau_2^2}\Big((-h\varphi_2)a^*(\tau_2 - \xi_2) - 2\tau_2\Big)\right] < 0 && \text{(A106)}
\end{aligned}$$

We use $\xi_1/\tau_1 - \xi_2/\tau_2 = (-1)$ to simplify to

$$d_2^2\Pi_2 = (d_2 a^*)\left[2 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_1)h\varphi_2}{\tau_1} - 2\right)\right]. \tag{A107}$$

ad $d_1 d_2\Pi_2$. Using (A66) and (A65),

$$\begin{aligned}
d_1 d_2\Pi_2 &= (d_1 a^*) + (d_1 d_2 a^*)\gamma_2 && \text{(A108)}\\
&= (d_1 a^*)\left[1 + \frac{\gamma_2}{\tau_2^2}\underbrace{\Big((-a^*\xi_2)(-h\varphi_2) - 2\tau_2\Big)}_{:=E}\right] > 0, && \text{(A109)}
\end{aligned}$$

since $E > 0$ by assumption (A4). Again this further simplifies to

$$d_1 d_2\Pi_2 = (d_1 a^*)\left[1 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2\right)\right]. \tag{A110}$$

ad $d_1^2\Pi_1$. Using (A67),

$$\begin{aligned}
d_1^2\Pi_1 &= d_1\left[(\overline{a} - a^*) - (d_1 a^*)\gamma_1\right] && \text{(A111)}\\
&= -2(d_1 a^*) - \gamma_1(d_1^2 a^*) && \text{(A112)}
\end{aligned}$$

$$= -d_1 a^* \left[2 + \gamma_1 (d_1 a^*) \left(\frac{2}{a^*} - g\varphi_1 \xi_2 \frac{\alpha}{\tau_1}\right)\right] \tag{A113}$$

$$= -d_1 a^* \left[2 + \frac{\gamma_1}{\tau_1} \left(2 - \frac{a^* \xi_2 g \varphi_1}{\tau_2}\right)\right]. \tag{A114}$$

ad $d_1 d_2 \Pi_1$. Using (A66),

$$d_2 d_1 \Pi_1 = d_2 \left[(\overline{a} - a^*) - (d_1 a^*)\gamma_1\right] \tag{A115}$$

$$= (-d_2 a^*) - \gamma_1 (d_1 a^*)^2 \frac{\alpha}{a^*} \left[a^* \xi_2 \frac{h\varphi_2}{\tau_2} - 2\right] \tag{A116}$$

$$= (-d_2 a^*) \left[1 + \frac{\gamma_1}{\tau_1} \left(2 - \frac{a^* \xi_2 h \varphi_2}{\tau_2}\right)\right]. \tag{A117}$$

ad $d_1^2 \Pi_1 + \frac{1}{\alpha} d_1 d_2 \Pi_1$. Using (A66) and (A68),

$$d_1^2 \Pi_1 + \frac{1}{\alpha}(d_1 d_2 \Pi_1) = (-d_1 a^*) \left[2 + \frac{\gamma_1}{\tau_1} \left(2 - \frac{(a^* \xi_2) g \varphi_1}{\tau_2}\right)\right] + (d_1 a^*) \left[1 + \frac{\gamma_1}{\tau_1} \left(2 - \frac{(a^* \xi_2) h \varphi_2}{\tau_2}\right)\right] \tag{A118}$$

$$= (-d_1 a^*) + (d_1 a^*) \frac{\gamma_1}{\tau_1} \left[\frac{(a^* \xi_2) g \varphi_1}{\tau_2} - \frac{(a^* \xi_2) h \varphi_2}{\tau_2}\right] \tag{A119}$$

$$= \underbrace{(-d_1 a^*)}_{<0} + (d_1 a^*) \frac{\gamma_1}{\tau_1} \underbrace{\frac{(-a^* \xi_2)}{\tau_2}}_{2/\tau_2 > 0} \underbrace{(h\varphi_2 - g\varphi_1)}_{<0} \tag{A120}$$

$$< 0, \tag{A121}$$

with $(h\varphi_2 - g\varphi_1) < 0$, since

$$g\varphi_1 - h\varphi_2 = hB_2 \left[\underbrace{\frac{gB_1}{hB_2}}_{=\alpha} \xi_2 - \xi_1\right] = hB_2 \tau_2 > 0 \tag{A122}$$

where the last equality follows, since

$$\tau_2 - \xi_2 = \Delta\gamma - h\varphi_2 - (\xi_1 + \Delta\gamma) = (-\xi_1)(1 + hB_2) = \frac{-\xi_1}{\alpha}. \tag{A123}$$

□

A6 Proof of Proposition 3

Following the notation as noted in the remark at the beginning of the appendix, $\partial_1 \gamma_2^* = d_1 \gamma_2^*$ and $\partial_i \Pi_2 = d_i \Pi_2$. Auxiliary properties are proven in Appendix A5. We first prove the following central claim.

Claim. The following notation is used: For a function $f(\vec{\gamma})$ let $f^*(\gamma_1) := f(\gamma_1, \gamma_2^*(\gamma_1))$. Then,

$$d_1\gamma_2^* = \frac{(d_1 d_2 \Pi_2)^*}{(-d_2^2 \Pi_2)^*}. \tag{A124}$$

Proof of claim. By definition, $0 \equiv (d_2\Pi_2)^*$ and thus

$$0 = d_1((d_2\Pi_2)^*) = (d_1 d_2 \Pi_2)^* + (d_2^2\Pi_2)^* d_1\gamma_2^* \tag{A125}$$

$$\Leftrightarrow d_1\gamma_2^* = \frac{(d_1 d_2 \Pi_2)^*}{(-d_2^2 \Pi_2)^*}. \tag{A126}$$

ad i). From (A68) we have concavity of Π_2, which ensures uniqueness of a solution. For existence, note that $\gamma_2 \mapsto \Pi_2(\gamma_1, \gamma_2)$ as continuous function on a compact interval, assumes its maximum. But $\Pi_2(\gamma_1, 0) = \Pi_2(\gamma_1, \gamma^{max}) = 0$, hence the maximum is assumed in the interior.

ad ii). For $\gamma_2^* \in \mathcal{C}^\infty$, we make use of the implicit function theorem. We know $d_2\Pi_2 \in \mathcal{C}^\infty$ and $d_2^2\Pi_2 < 0$. Hence, from the implicit function theorem the mapping

$$\gamma_1 \mapsto \gamma_2^*(\gamma_1) = \arg_{\gamma_2}\{d_2\Pi_2(\gamma_1, \gamma_2) = 0\} \tag{A127}$$

is smooth. Monotonicity of γ_2^* follows from the claim together with (A69) and (A68).

ad iii). Using (A68) and (A69),

$$\alpha^* d_1\gamma_2^* = \alpha^* \frac{(d_1 d_2 \Pi_2)^*}{(-d_2^2 \Pi_2)^*} \tag{A128}$$

$$= \alpha^* \frac{(d_1 a^*)\left[1 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2\right)\right]}{(-d_2 a^*)\left[2 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_1)h\varphi_2}{\tau_1} - 2\right)\right]} \tag{A129}$$

$$= \frac{1 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2\right)}{2 + \frac{\gamma_2}{\tau_2}\left(\frac{(a^*\xi_1)h\varphi_2}{\tau_1} - 2\right)}. \tag{A130}$$

In the numerator

$$\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2 = (-a^*\xi_2)\frac{(-h\varphi_2)}{\tau_2} - 2 > 0, \tag{A131}$$

since $a^*(-\xi_2) > 2$ from assumption A3, and in the denominator

$$\frac{(a^*\xi_1)h\varphi_2}{\tau_1} - 2 = (a^*h\varphi_2)\left(1 - \frac{\xi_2}{\tau_2}\right) - 2 = \underbrace{\frac{(a^*\xi_2)h\varphi_2}{\tau_2} - 2}_{>0} + \underbrace{(-h\varphi_2)a^*}_{>0} > 0, \tag{A132}$$

hence, we get $\alpha^* d_1 \gamma_2^* < 1$.

$\square$

A7 Proof of Proposition 4

Following the notation as noted in the remark at the beginning of the appendix, $\partial_2\gamma_1^{\otimes} = d_2\gamma_1^{\otimes}$ and $\partial_i\Pi_1 = d_i\Pi_1$. The proof proceeds by showing a basic lemma from real analysis (Lemma 4) and then proving its applicability in the present context (Lemma 5 and Lemma 6). The lemmata are presented upfront and proven in the subsequent appendices.

Notation. $\gamma_1^{\overline{a}}(\gamma_2)$ is defined similar to $\gamma_1^{\underline{a}}(\gamma_2)$. In particular, $\gamma_1^{\overline{a}}(\gamma_2)$ is defined as $a^*(\gamma_1^{\overline{a}}(\gamma_2), \gamma_2) = \overline{a}$ if there is a solution in $\mathcal{G}_{[\underline{a},\overline{a}]}$, and as $\gamma_1^{\overline{a}}(\gamma_2) = \gamma^{max}$ otherwise.

Lemma 4. *Let f be a smooth function on some interval $[a,b] \subset \mathbb{R}$. If there exists a $\mu \in [a,b]$ such that*

$$\forall x < \mu: \quad df(x) = 0 \Rightarrow d^2f(x) > 0 \tag{S1}$$

$$\forall x > \mu: \quad df(x) = 0 \Rightarrow d^2f(x) < 0 \tag{S2}$$

$$df(a) > 0, \tag{S3}$$

then f has a global maximum τ and $\forall x < \tau : df(x) > 0$ and $\forall x > \tau : df(x) < 0$.

Lemma 5. *Consider a fixed γ_2 for which*

$$d_1\Pi_1(\gamma_1^{\underline{a}}(\gamma_2), \gamma_2) > 0. \tag{T3}$$

If assumption A5 holds, there exists a $\mu \in [\gamma_1^{\underline{a}}(\gamma_2), \gamma_1^{\overline{a}}(\gamma_2)]$ such that for all $\gamma_1 \in [\gamma_1^{\underline{a}}(\gamma_2), \gamma_1^{\overline{a}}(\gamma_2)]$

$$\gamma_1 < \mu \Rightarrow \quad \Big(d_1\Pi_1 = 0 \Rightarrow d_1^2\Pi_1(\gamma_1) > 0\Big) \tag{T1}$$

$$\gamma_1 > \mu \Rightarrow \quad \Big(d_1\Pi_1 = 0 \Rightarrow d_1^2\Pi_1(\gamma_1) < 0\Big). \tag{T2}$$

If $\mu = \gamma_1^{\overline{a}}(\gamma_2)$, then $\gamma_1^{\overline{a}}(\gamma_2) = \gamma^{max}$.

Lemma 6. *Assumption A6 implies that, for all $\gamma_2 \in [0, \gamma_2^*(\gamma^{max})]$,*

$$d_1\Pi_1(\gamma_1^{\underline{a}}(\gamma_2), \gamma_2) > 0. \tag{A133}$$

ad i). Lemma 5 and Lemma 6 show that for any $\gamma_2 \in [0, \gamma_2^*(\gamma^{max})]$ we can make use of Lemma 4. Then we know from Lemma 4 that for all $\gamma_2 \in [0, \gamma_2^*(\gamma^{max})]$, a) $\Pi_1(\cdot, \gamma_2)$ has a unique global maximum $\tau \in [\gamma_1^{\underline{a}}(\gamma_2), \gamma_1^{\overline{a}}(\gamma_2)]$, b) $\tau = \text{argmin}_\mu\{d_1\Pi_1 = 0 \vee \mu = \gamma^{max}\}$, i.e., τ is either the unique solution to $d_1\Pi_1(\tau) = 0$, or $\tau = \gamma^{max}$, and, c) for all $\gamma_1 < \tau$: $d_1\Pi_1(\gamma_1) > 0$

and for all $\gamma_1 > \tau$: $d_1\Pi_1(\gamma_1) < 0$.

ad ii). We consider $\gamma_1^{\otimes}$ separately on

$$\mathcal{N}_1 := \{\gamma_2 \in [0, \gamma_2^*(\gamma^{max})] | \gamma_1^{\otimes} < \gamma^{max}\} \tag{A134}$$
$$\mathcal{N}_2 := \{\gamma_2 \in [0, \gamma_2^*(\gamma^{max})] | \gamma_1^{\otimes} = \gamma^{max}\} \tag{A135}$$

and first show continuity on $[0, \gamma_2^*(\gamma^{max})] = \mathcal{N}_1 \dot\cup \mathcal{N}_2$.

On $\mathcal{N}_1$, we already know from (A71) that $d_1^2\Pi_1 \neq 0 \vee d_1 d_2 \Pi_1 \neq 0$. Hence $\{d_1\Pi_1 = 0\}$ is a smooth curve. Then we make a case distinction.

1) If $d_1^2\Pi_1 \neq 0$, we know from the implicit function theorem that one can parameterize $\{d_1\Pi_1 = 0\}$ via $\gamma_1^{\otimes}(\gamma_2)$. In particular, such a parameterization is smooth.

2) At a point $q = (\gamma_1^{\otimes}(\gamma_2), \gamma_2)$ with $d_1^2\Pi_1(q) = 0$, we have $d_1 d_2 \Pi_1(q) \neq 0$, hence one can parameterize $\{d_1\Pi_1 = 0\}$ locally via $\gamma_2^{\otimes}(\gamma_1)$. $\gamma_2^{\otimes}(\gamma_1)$ has to strictly increase in some neighborhood around q or strictly decrease in some neighborhood around q, since otherwise the inverse couldn't exist. Hence, $\gamma_2^{\otimes}$ is bijective on some neighborhood U of $\gamma_1^{\otimes}(\gamma_2)$ and V of $\gamma_2^{\otimes}$. Then $\gamma_1^{\otimes}$ is monotone on U and continuous.

Hence, $\gamma_1^{\otimes}$ is continuous on $\mathcal{N}_1$, $\mathcal{N}_1$ is open and $\gamma_1^{\otimes}$ is also continuous on the closure of $\mathcal{N}_1$, $\overline{\mathcal{N}}_1$. Since the complement of $\mathcal{N}_1$ is $\mathcal{N}_2$, $\mathcal{N}_2$ is closed. On $\mathcal{N}_2$, $\gamma_1^{\otimes}$ is a constant function and as such continuous on $\mathcal{N}_2$. Thus, since $\gamma_1^{\otimes}$ is continuous on $\overline{\mathcal{N}}_1$ and on $\mathcal{N}_2$, it is continuous on $[0, \gamma_2^*(\gamma^{max})]$.

It remains to show smoothness except at isolated points.

Again, we consider $\mathcal{N}_1$ first. If $d_1^2\Pi_1 \neq 0$, the above argument has already shown smoothness. At a point $q = (\gamma_1^{\otimes}(\gamma_2), \gamma_2)$ with $d_1^2\Pi_1(q) = 0$, we have $d_1 d_2 \Pi_1(q) \neq 0$, hence one can parameterize $\{d_1\Pi_1 = 0\}$ locally via $\gamma_2^{\otimes}(\gamma_1)$. $d_1^2\Pi_1(\gamma_1, \gamma_2^{\otimes}(\gamma_1))$ is an analytic function, i.e., the Taylor expansion converges at every point with positive radius of convergence. From complex analysis (see e.g. Theorem 4.8 in Shakarchi and Stein (2003)) we know that, if the zeros of the function accumulate, then $d_1^2\Pi_1(\gamma_1, \gamma_2^{\otimes}(\gamma_1)) \equiv 0$ on some open neighborhood U of γ_1. But this is a contradiction: Consider the image $V := \{(\gamma_1, \gamma_2^{\otimes}(\gamma_1)) | \gamma_1 \in U\} \subset \{d_1\Pi_1 = 0\}$. There, $d_1^2\Pi_1 = 0 \wedge d_1 d_2 \Pi_1 \neq 0$ everywhere. So the tangent to $\{d_1\Pi_1 = 0\}$ may not have a component in d_2-direction. But this means that $\gamma_2^{\otimes}(\gamma_1)$ is constant on U – a contradiction to the argument in the proof of i). This proves the claim on an open neighborhood of $\overline{\mathcal{N}}_1$.

On $\mathcal{N}_2$, $\gamma_1^{\otimes}$ is constant and thus smooth on all interior points of $\mathcal{N}_2$, i.e., except possibly on points on $\mathcal{N}_2 \cap \overline{\mathcal{N}}_1$. But these points are isolated by the proof for $\mathcal{N}_1$.

In addition, exception points are well-behaved:
Claim. Let γ_2^0 be a point at which $\gamma_1^{\otimes}$ is non-differentiable, i.e. $d_1^2\Pi_1(\gamma_1^{\otimes}(\gamma_2^0), \gamma_2^0) = 0$. Then,

i) $d_2\gamma_1^{\otimes}$ converges to minus infinity in γ_2^0.

ii) $\gamma_1^{\otimes}$ decreases in a neighborhood of γ_2^0.

Proof of claim.
Consider a point $q = (\gamma_1^{\otimes}(\gamma_2^0), \gamma_2^0)$ with $d_1^2\Pi_1(q) = 0$, and a locally inverse function $\gamma_2^{\otimes}$ of the parameterization $\gamma_1^{\otimes}$ in a neighborhood V. Since $d_1\Pi_1(q) = 0$, $d_1\gamma_2^{\otimes}(\gamma_1^{\otimes}) = 0$ and $\lim_{\gamma_1 \to \gamma_1^{\otimes}} d_1\gamma_2^{\otimes}(\gamma_1) = 0$. From monotonicity of $\gamma_2^{\otimes}(\gamma_1)$, either for all $\gamma_2^0 \neq \gamma_2 \in V$, $d_1\gamma_2^{\otimes} > 0$ or for all $\gamma_2^0 \neq \gamma_2 \in V$, $d_1\gamma_2^{\otimes} < 0$. Since $d_2\gamma_1^{\otimes} = 1/d_1\gamma_2^{\otimes}$, either $\lim_{\gamma_2 \to \gamma_2^0} d_2\gamma_1^{\otimes} = \infty$ or $\lim_{\gamma_2 \to \gamma_2^0} d_2\gamma_1^{\otimes} = -\infty$, but the first case is ruled out by part iii).

ad iii). $d_2\gamma_1^{\otimes}$, if defined, is either equal to 0 or $\gamma_1^{\otimes} < \gamma^{max}$. and $d_1^2\Pi_1 \neq 0$. But $0 < \alpha$, so we may assume $\gamma_1^{\otimes} < \gamma^{max}$. We first prove a preliminary claim.

Claim. Consider a continuously differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$ and vectors $(1, a)$ and $(1, b)$ with $0 < a < b \in \mathbb{R}$. Consider a point $p \in \mathbb{R}^2$ with $Df(p) \neq 0$. If in p the directional derivatives $D_{(1,a)}f$ and $D_{(1,b)}f$ have the same sign, then all directional derivatives $D_{(1,x)}f$ with $a \leq x \leq b$ have the same sign in p. If in p one of the directional derivatives, $D_{(1,a)}f, D_{(1,b)}f$, is equal and the other unequal to zero, then for all $x \in (a, b)$ $D_{(1,x)}f \neq 0$ and has the same sign.

Proof of claim. We have $Df(p) \neq 0$, hence the gradient $\mathrm{grad}(f) = (d_1 f, d_2 f)$ does not vanish at p. Hence,

$$D_{(1,x)}f = \langle (1, x), \mathrm{grad}(f) \rangle = d_1 f + x\, d_2 f \tag{A136}$$

is a linear function in x. Subsequently, if $d_1 f + x\, d_2 f$ (as a function in x) has the same sign for $x = a$ and $x = b$, it has the same sign for all $x \in (a, b)$. This also holds in case one of the two directional derivatives $D_{(1,a)}f, D_{(1,b)}f$ are zero. This proves the claim.

$d_2\gamma_1^{\otimes}$ is defined by $D_{(d_2\gamma_1^{\otimes},1)}(d_1\Pi_1) = 0$. Thus, it is to show that for $\kappa > \alpha$, $D_{(\kappa,1)}(d_1\Pi_1) \neq 0$. Since $D_{(\kappa,1)}f = \kappa D_{(1,1/\kappa)}f$, this is equivalent to showing that for $1/\kappa \in (0, 1/\alpha)$, $D_{(1,1/\kappa)}(d_1\Pi_1) \neq 0$. With the above claim it thus remains to show that

$$d_1^2\Pi_1(\gamma_1^{\otimes}(\gamma_2), \gamma_2) = D_{(1,0)}(d_1\Pi_1) \leq 0 \tag{A137}$$

$$\text{and} \qquad D_{(1,1/\alpha)}(d_1\Pi_1) < 0. \tag{A138}$$

(A138) holds, since $D_{(1,1/\alpha)}(d_1\Pi_1) = d_1^2\Pi_1 + \frac{1}{\alpha}(d_1 d_2 \Pi_1) < 0$ from (A72). (A137) holds since $\gamma_1^{\otimes}(\gamma_2)$ is a local maximum..

□

A8 Proof of Lemma 4

Condition (S1) requires that for $x < \mu$, f only has local minima. Condition (S2), on the other hand, requires that for $x > \mu$, f only has local maxima. Hence, f is increasing on the interval $[a, \mu]$, since otherwise from condition (S3) there was a local maxima below μ. On the interval $[b, \mu]$ there can be at most one local maximum τ, since otherwise there would be another local minima in between - contradiction.

Subsequently, f is increasing on $[a, \mu]$ and decreasing on $[\mu, b]$. Hence, τ is a global maximum and from monotonicity we have $\forall x < \tau : df(x) \geq 0$ and $\forall x > \tau : df(x) \leq 0$. But $df(x)$ must not be zero for $x \neq \tau$, since otherwise from (S1) and (S2) at that point there would be another local extremum, which would entail another extremum in between - contradiction.

□

A9 Proof of Lemma 5

If $d_1\Pi_1 = 0$,

$$d_1^2\Pi_1 \geq 0 \tag{A139}$$

$$\Leftrightarrow -(2d_1a^* + \gamma_1 d_1^2 a^*) \geq 0 \tag{A140}$$

$$\Leftrightarrow 2d_1a^* \leq -\gamma_1(d_1a^*)^2\left[\frac{2}{a^*} - g\varphi_1\frac{\xi_2\alpha}{\tau_1}\right] \tag{A141}$$

$$\overset{d_1\Pi_1=0}{\Leftrightarrow} 2 \leq -\gamma_1\frac{(\overline{a}-a^*)}{\gamma_1}\left[\frac{2}{a^*} - g\varphi_1\frac{\xi_2\alpha}{\tau_1}\right] \tag{A142}$$

$$\Leftrightarrow 2a^* \leq (\overline{a}-a^*)\left[-2 + g\varphi_1\xi_2 \underbrace{a^*\frac{\alpha}{\tau_1}}_{=(-d_2a^*)}\right] \tag{A143}$$

$$\Leftrightarrow 2\overline{a} \leq (\overline{a}-a^*)g\varphi_1\xi_2(-d_2a^*), \tag{A144}$$

where we used $d_1\Pi_1 = 0 \Leftrightarrow d_1a^* = \frac{(\overline{a}-a^*)}{\gamma_1}$ as well as (A67).

Define the RHS of (A144) as

$$R(\gamma_1, \gamma_2) := (\overline{a}-a^*)g\varphi_1\xi_2(-d_2a^*). \tag{A145}$$

Claim. It suffices to show $d_1R < 0$.

Proof of claim. If $d_1R < 0$, there can be at most one μ with $2\overline{a} = R(\mu, \gamma_2)$ and for this μ (T1) and (T2) hold. In case there is no μ with $2\overline{a} = R(\mu, \gamma_2)$, we distinguish the following cases:

1) If there is an interior local maximum, at this interior local maximum we must have $d_1\Pi_1^2 < 0$. Hence $2\overline{a} > R$ on the entire interval and $\mu = \gamma_1^{\underline{a}}(\gamma_2)$ satisfies the condition.

2) If there is no interior local maximum, Π_1 increases on the entire interval by Assumption A6 and Lemma 6, and $\mu = \gamma^{max}$ satisfies the condition. In that case also $\gamma_1^{\overline{a}}(\gamma_2) = \gamma^{max}$, because otherwise $\Pi_1(\gamma_1^{\overline{a}}(\gamma_2), \gamma_2) = 0$ would contradict monotonicity of Π_1.

3) By Assumption A6 and Lemma 6 there can be no interior local minima.

Claim. $d_1R < 0$.

Proof of claim. Using (A70) in Proposition 4,

$$\begin{aligned}
d_1 R &= (-d_1 a^*)\frac{R}{(\overline{a}-a^*)} + d_1\varphi_1 \frac{R}{\varphi_1} - (d_1 d_2 a^*)\frac{R}{(-d_2 a^*)} && \text{(A146)}\\
&= R(d_1 a^*)\cdot\left[-\frac{1}{(\overline{a}-a^*)} + \alpha\xi_2 - \frac{1}{\alpha}\frac{(d_1 d_2 a^*)}{(d_1 a^*)^2}\right] && \text{(A147)}\\
&= R(d_1 a^*)\cdot\left[-\frac{1}{(\overline{a}-a^*)} + \alpha\xi_2 - \frac{1}{\alpha}\frac{\alpha}{a^*}\left(a^*\xi_2\frac{h\varphi_2}{\tau_2} - 2\right)\right] && \text{(A148)}\\
&= \underbrace{R(d_1 a^*)}_{>0}\cdot\left[-\frac{1}{(\overline{a}-a^*)} + \frac{2}{a^*} - \xi_2(\alpha + \frac{h\varphi_2}{\tau_2})\right] && \text{(A149)}
\end{aligned}$$

Subsequently

$$\begin{aligned}
d_1 R < 0 &\Leftrightarrow \frac{2\overline{a}-3a^*}{(\overline{a}-a^*)a^*} < \xi_2\left(\alpha + \frac{h\varphi_2}{\tau_2}\right) && \text{(A150)}\\
&\Leftrightarrow \underbrace{\frac{2\overline{a}-3a^*}{(\overline{a}-a^*)}}_{<2} < \underbrace{a^*\xi_2}_{<(-2)}\left(\underbrace{\alpha}_{\in(0,1)} + \underbrace{\frac{h\varphi_2}{\tau_2}}_{<(-1)}\right). && \text{(A151)}
\end{aligned}$$

Assumption A5 ensures that the LHS of (A151) is negative, and, thus, under assumption A5 (A151) holds.

□

A10 Proof of Lemma 6

From assumption A6 we have $d_1\Pi_1(\gamma_1^{\underline{a}}(\gamma_2),\gamma_2) > 0$ for $\gamma_2 = \gamma_2^*(\gamma^{max})$. From (A71), we know

$$D_{(1,1/\alpha)}(d_1\Pi_1) = d_1^2\Pi_1 + \frac{1}{\alpha}(d_1 d_2\Pi_1) < 0. \tag{A152}$$

Hence, $d_1\Pi_1(\gamma_1^{\underline{a}}(\gamma_2),\gamma_2)$ increases along $\{a^* = \underline{a}\}$ as γ_2 decreases, and, thus, $d_1\Pi_1(\gamma_1^{\underline{a}}(\gamma_2),\gamma_2) > 0$ for all $\gamma_2 \in [0,\gamma_2^*(\gamma^{max})]$.

□

A11 Proof of Proposition 2

ad Existence. We consider insurer 2's reaction function

$$\begin{aligned}
\gamma_2^* : [\gamma_1^{\underline{a}}(0),\gamma^{max}] &\to [0,\gamma_2^*(\gamma^{max})] && \text{(A153)}\\
\gamma_1 &\mapsto \gamma_2^*(\gamma_1) && \text{(A154)}
\end{aligned}$$

and insurer 1's reaction function

$$\gamma_1^{\otimes} : [0, \gamma_2^*(\gamma^{max})] \to [\gamma_1^{a}(0), \gamma^{max}] \tag{A155}$$
$$\gamma_2 \mapsto \gamma_1^{\otimes}(\gamma_2). \tag{A156}$$

From Propositions 4 and 3 we know that γ_2^* and $\gamma_1^{\otimes}$ are continuous functions. Subsequently,

$$(\gamma_2^* \circ \gamma_1^{\otimes}) : [0, \gamma_2^*(\gamma^{max})] \to [0, \gamma_2^*(\gamma^{max})] \tag{A157}$$

is a continuous self-mapping on a nonempty, compact and convex set and, hence, by Brouwer's fixed point theorem (rf Mas-Colell, Whinston, and Green (1995, p. 952)) there exists a fixed point. By construction a fixed point either satisfies both FOCs or lies at the boundary.

ad Uniqueness. Since insurer 2's reaction function γ_2^* is strictly increasing, there exists an inverse function, denoted by γ_1^{*-1}. From part iii) of Proposition 4 we have for insurer 2's reaction function $d_1\gamma_2^* < 1/\alpha^*$, hence, for its inverse function

$$d_2\gamma_1^{*-1} > \alpha^*. \tag{A158}$$

At the same time, we know from Proposition 3 that for insurer 1's reaction function

$$d_2\gamma_1^{\otimes} < \alpha^{\otimes}. \tag{A159}$$

Consider the mapping

$$\gamma_2 \mapsto \gamma_1^{*-1}(\gamma_2) \mapsto a^*(\gamma_1^{*-1}(\gamma_2), \gamma_2). \tag{A160}$$

Then $a^*(\gamma_1^{*-1}(\cdot), \cdot)$ as a function of γ_2 is increasing in γ_2, since

$$0 < d_2a^*(\gamma_1^{*-1}(\gamma_2), \gamma_2) = (d_1a^*)(d_2\gamma_1^*) + d_2a^* \Leftrightarrow d_2(\gamma_1^*) > \frac{(-d_2a^*)}{(d_1a^*)} = \alpha, \tag{A161}$$
$$\Leftrightarrow \frac{1}{d_1(\gamma_2^*)} > \alpha, \tag{A162}$$

which holds by Proposition 3 part iii).

Likewise, one can consider the analogous mapping using insurer 1's reaction function

$$\gamma_2 \mapsto \gamma_1^{\otimes}(\gamma_2) \mapsto a^*(\gamma_1^{\otimes}(\gamma_2), \gamma_2). \tag{A163}$$

Then $a^*(\gamma_1^{\otimes}(\cdot), \cdot)$ as a function of γ_2 is decreasing in γ_2, since

$$0 > d_2a^*(\gamma_1^{\otimes}(\gamma_2), \gamma_2) = (d_1a^*)(d_2\gamma_1^{\otimes}) + d_2a^* \Leftrightarrow d_2(\gamma_1^{\otimes}) < \frac{(-d_2a^*)}{(d_1a^*)} = \alpha, \tag{A164}$$

which holds by Proposition 4 part iii).

Since a^* values at a point must coincide at a point at which the two function intersect, there can be at most one intersection.

□

A12 Proof of Proposition 5

First, at a Nash equilibrium $\vec{\gamma}$ one has $d_1\Pi_1(\vec{\gamma}) \geq 0 = d_2\Pi_2(\vec{\gamma})$ with $d_1\Pi_1(\vec{\gamma}) > 0$ only if $\gamma_1 = \gamma^{max}$. Note furthermore that

$$d_1\Pi_1 \geq 0 \Leftrightarrow (\overline{a} - a^*) - \gamma_1 d_1 a^* \geq 0 \tag{A165}$$

$$d_2\Pi_2 = 0 \Leftrightarrow (a^* - \underline{a}) + \gamma_2 d_2 a^* = 0. \tag{A166}$$

Using Lemma 2 part iii), it thus follows that at a point $\vec{\gamma}$ with $d_1\Pi_1(\vec{\gamma}) \geq 0 = d_2\Pi_2(\vec{\gamma})$ we have

$$1 > \alpha = \frac{-d_2 a^*}{d_1 a^*} \geq \frac{-d_2 a^*}{(\overline{a} - a^*)}\gamma_1 = \frac{(a^* - \underline{a})}{(\overline{a} - a^*)}\frac{\gamma_1}{\gamma_2} = \frac{\Pi_2}{\Pi_1}\frac{\gamma_1^2}{\gamma_2^2} > \frac{\Pi_2}{\Pi_1}, \tag{A167}$$

where the last inequality follows since $\Delta\gamma < 0 \Leftrightarrow \gamma_1/\gamma_2 > 1$. Hence, (A167) yields $(a^* - \underline{a}) < (\overline{a} - a^*)$ and $\Pi_2 < \Pi_1$.

□

A13 Proof of Proposition 6

The optimization problem for a given vector of default risks $\vec{b}^0$ depends only on $\tilde{g}(\vec{b}^0) = p(b_2^0 - b_1^0)/(1 - b_1^0 p)$. Hence, vectors of default risks with the same $\tilde{g}$ yield the same Nash equilibria.

Claim. For a given pair of default risks $(b_1^0, b_2^0) = \vec{b}^0$ with $\tilde{g}(\vec{b}^0)$, the set of default risks $\vec{b}$ with the same $\tilde{g}$ is

$$\left\{ \left(b_1^0 - \alpha, b_2^0 - (1 - \tilde{g}(b_1^0, b_2^0))\alpha\right) \middle| \alpha \in \left[b_1^0 - \frac{1}{3}, b_1^0\right] \right\}. \tag{A168}$$

Proof of claim. We have

$$\partial_{b_2}\tilde{g}\Big|_{\vec{b}^0} = \frac{p}{1 - b_1^0 p} \tag{A169}$$

$$\partial_{b_1}\tilde{g}\Big|_{\vec{b}^0} = \frac{-p(1 - b_1^0 p) + p(b_2^0 - b_1^0)p}{(1 - b_1^0 p)^2} \tag{A170}$$

$$= -p\frac{(1 - b_2^0 p)}{(1 - b_1^0 p)^2} \tag{A171}$$

$$= -\frac{p}{(1 - b_1^0 p)}\Big[1 - \underbrace{\frac{p\Delta b}{(1 - b_1^0 p)}}_{=\tilde{g}(\vec{b}^0)}\Big] \tag{A172}$$

$$-\frac{\partial_{b_1}\tilde{g}}{\partial_{b_2}\tilde{g}}\bigg|_{\vec{b}^0} = (1-\tilde{g}(\vec{b}^0)) \in (0,1) \tag{A173}$$

Hence, from the implicit function theorem we know that sets $\{\vec{\gamma}|\tilde{g}(\vec{\gamma}) = c\}$ are submanifolds that have (for a given c) the same slope $(1-\tilde{g})$ at each point. Hence they are straight lines.

$\square$

A14 Proof of Proposition 7

ad i). Let (b_1^*, b_2^*) be a subgame-perfect equilibrium for prescribed roles. If Π_2^* has multiple maxima, b_2^s and b_2^l are defined as the smallest and largest b_2 at which the maximum is assumed. The following goes through for b_2^s and b_2^l, in particular for b_2^s. From the definition of b_2^s

$$b_2^s = b_2^* - (1-\tilde{g}(b_1^*, b_2^*))\, b_1^*. \tag{A174}$$

For all $\lambda \in [0, b_1^*]$, all $b^\lambda = (\lambda, b_2^s + (1-\tilde{g}(b_1^*, b_2^*))\lambda)$ are also subgame-perfect equilibria with

$$\Pi_i^\square(b_1^*, b_2^*) = \Pi_i^\square(b_1^\lambda, b_2^\lambda) \tag{A175}$$

for $i \in \{1, 2\}$ from from Proposition 6 ii), since

$$\frac{b_2^\lambda - b_2^*}{b_1^\lambda - b_1^*} = \frac{-(1-\tilde{g}(b_1^*, b_2^*))(b_1^* - \lambda)}{(\lambda - b_1^*)} = 1 - \tilde{g}(b_1^*, b_2^*). \tag{A176}$$

Now, suppose $b_1^* > b_2^s$. Then, for $\lambda = (b_1^* - b_2^s)/(1-\tilde{g}(b_1^*, b_2^*)) > 0$, $b_2^\lambda = b_1^*$ and

$$\lambda - b_1^* = \frac{1}{1-\tilde{g}(b_1^*, b_2^*)}\,(b_1^* - (b_2^s + (1-\tilde{g}(b_1^*, b_2^*))b_1^*)) = \frac{1}{1-\tilde{g}(b_1^*, b_2^*)}(b_1^* - b_2^*) < 0, \tag{A177}$$

hence $\lambda < b_1^*$. So in this case, $(b_1^\lambda, b_2^\lambda)$ offers a profitable deviation for insurer 2 by choosing λ and thereby reserving roles and capturing profit $\Pi_1^\square(b_1^\lambda, b_2^\lambda) > \Pi_2^\square(b_1^\lambda, b_2^\lambda) = \Pi_2^\square(b_1^*, b_2^*)$. Contradiction.

ad ii). In any subgame-perfect equilibrium, we have

$$b_2^* \leq b_2^l + \left(1 - \tilde{g}(0, b_2^l)\right) b_1^* \leq \left(2 - \tilde{g}(0, b_2^l)\right) b_2^l. \tag{A178}$$

ad Upper Bound. In general, from Proposition 6 we know that pairs of default risks (b_1, b_2) with

$$(b_1, b_2) = (b_1^* - z, b_2^* - (1-\tilde{g}(b_1^*, b_2^*))z), \tag{A179}$$

$z \in [b_1^* - \frac{1}{3}, b_1^*]$, lead to the same Nash equilibria in prices. Hence, the second mover has the option to choose a quality $b_2^1 < b_1^*$ with

$$(b_2^1, b_1^*) = \left(\frac{1}{(1-\tilde{g})} \left[(1-\tilde{g})b_1^* - (b_2^* - b_1^*) \right], b_1^* \right) \tag{A180}$$

that leads to the same Nash equilibrium in prices, but with reversed roles. By Proposition 5 this is a profitable deviation. This deviation is infeasible if

$$b_2^* - b_1^* > (1 - \tilde{g}(b_1^*, b_2^*))b_1^* \tag{A181}$$

$$\Leftrightarrow b_2^* > (2 - \tilde{g}(b_1^*, b_2^*))b_1^* \tag{A182}$$

$$\Leftrightarrow b_1^* < \underbrace{\frac{1}{(2-\tilde{g})}}_{<2-1/8 \text{ from Lemma 7}} \underbrace{b_2^*}_{<b^{max}} . \tag{A183}$$

□

A15 Proof of Proposition 8

"$\Leftarrow$" Similar to the proof of Proposition 7, insurer 1 must choose b_1 in such a way that it is not profitable for insurer 2 to become quality leader. This is the case if Π_2^* exceeds any profit insurer 2 can capture with reversed roles. Since from condition (N2) the profit of the quality leader is increasing in the lower quality, consider the smallest $\bar{b_1}$ such that $\Pi_1^\square(0, \bar{b_1}) = \Pi_2^*$. Then, $b_1 < \bar{b_1}$ leaves no profitable deviation for the follower.

"$\Rightarrow$" Let (b_1^*, b_2^*) be a subgame-perfect equilibrium with prescribed roles. Then for $i \in \{1, 2\}$ and $\lambda \leq b_1^*$

$$\Pi_i^\square(b_1^* - \lambda, b_1^*) = \Pi_i^\square \left(0, b_1^* - (1 - \tilde{g}(b_1^* - \lambda, b_1^*))(b_1^* - \lambda)\right) \tag{A184}$$

$$= \Pi_i^\square \left(0, \lambda + \tilde{g}(b_1^* - \lambda, b_1^*)(b_1^* - \lambda)\right). \tag{A185}$$

Suppose $b_1^* > \bar{b_1}$. Then let $\lambda = \bar{b_1}$. Hence

$$\Pi_1^\square(b_1^* - \bar{b_1}, b_1^*) = \Pi_1^\square \left(0, \bar{b_1} + \tilde{g}(b_1^* - \bar{b_1}, b_1^*)(b_1^* - \bar{b_1})\right) > \Pi_2^* \tag{A186}$$

by definition of $\bar{b_1}$. Since $\tilde{g}(b_1^* - \bar{b_1}, b_1^*)(b_1^* - \bar{b_1}) > 0$, this is a feasible profitable deviation. Contradiction.

□

B Appendix: Additional Results

B1 Optimal Choice of State-Contingent Payments

This section clarifies the contracting problem that has the specified insurance contract as outcome. Suppose an insurer with default risk b offers a contract that involves a fixed rate γ for establishing the client-insurer relationship, after which the insurer offers the actuarially fair price and the coverage is determined endogenously. All payments are due in $t = 4$. This includes γ, which, although set ex-ante, is also exchanged in $t = 4$ and hence only due if the insurer survives.

Clients chooses payments (y, z) with

$$y \text{ due if } \tilde{x} = \overline{\theta} \text{ and the insurer survives} \tag{B3}$$

$$z \text{ due if } \tilde{x} = \underline{\theta} \text{ and the insurer survives} \tag{B4}$$

to maximize expected utility

$$(1-p)u\Big(\overline{\theta} - y\Big) + p(1-b)u\Big(\underline{\theta} - z\Big) + bpu(\underline{\theta}) \tag{B5}$$

subject to the constraint

$$(1-p)y + p(1-b)z - \left[\gamma - \frac{bp\underline{\theta}}{(1-bp)}\right](1-bp) \geq 0 \tag{B6}$$

$$\Leftrightarrow (1-p)y + p(1-b)z \geq \gamma(1-bp) - bp\underline{\theta}. \tag{B7}$$

(B6) and (B7) offer two views on the constraint. (B7) demands that the expected cash flows to the insurer (LHS) must be at least as high as the expected fee already agreed upon minus the expected endowment if the insurer survives. To see the latter part note that

$$E\left[\tilde{x}|\text{insurer survives}\right] P[\text{insurer survives}] = (1-p)\overline{\theta} + p(1-b)\underline{\theta} \overset{E[\tilde{x}]=0}{=} -bp\underline{\theta} \tag{B8}$$

$$\Leftrightarrow E\left[\tilde{x}|\text{insurer survives}\right] = \frac{-bp\underline{\theta}}{(1-bp)} > 0. \tag{B9}$$

The risk-averse clients passes the risky endowment to the insurer unless the insurer defaults.

(B6) offers an alternative explanation. Let γ^{nom} be the expression in brackets, i.e.

$$\gamma^{nom} := \gamma - \frac{bp\underline{\theta}}{(1-bp)}. \tag{B10}$$

Then the third term on the LHS of (B6) is the "nominal" fee per client-insurer relationship, γ^{nom}, times the survival probability of the insurer, since only in that case the payment is actually exchanged. It is subtracted because this fee for establishing the client-insurer relationship has already been agreed upon, so the insurer already "mentally set it aside" and subsequently wants to break even in $t = 3$. Compared to γ, from the definition we have $\gamma = \gamma^{nom} + bp\underline{\theta}/(1-bp) < \gamma^{nom}$. In view of (B9) the adjustment term, $bp\underline{\theta}/(1-bp)$, is precisely the expected endowment conditional on the survival of the insurer. Since it is positive, the client claims this extra revenue for himself, rendering γ the "true" fees for the insurer. In the formulation of the insurer's constraint in (B6) one assumes that the insurer chooses "true" fees γ instead of "nominal" ones γ^{nom}. This reparametrization will make subsequent calculations tractable as we will see, while simplifying the intuition.

The following proposition then is a direct result from solving a client's optimization problem

$$\max_{y,z} \left\{ (1-p)u(\overline{\theta}-y) + p(1-b)u(\underline{\theta}-z) + bpu(\underline{\theta}) \middle| (1-p)y + p(1-b)z = \gamma(1-bp) - bp\underline{\theta} \right\} \tag{B11}$$

.

Proposition 10. *For a given (b,γ), the client optimally chooses*

$$y^*(b,\gamma) = \gamma + \frac{p(1-b)\overline{\theta} - p\underline{\theta}}{(1-bp)} \tag{B12}$$

$$z^*(b,\gamma) = \gamma - \frac{(1-p)}{(1-bp)}\overline{\theta} - \frac{b(1-bp)-(1-p)}{(1-b)(1-bp)}\underline{\theta}. \tag{B13}$$

Let $r^*(b,\gamma)$ be the payoff a client is left with in an optimal insurance contract unless the counterparty defaults (residual endowment), i.e. $r^*(b,\gamma) := \overline{\theta} - y^*(b,\gamma) = \underline{\theta} + z^*(b,\gamma)$. Then, as one would expect from risk aversion, $r^*(b,\gamma)$ does not depend on the endowment state, namely

$$r^*(b,\gamma) = -\gamma. \tag{B14}$$

B2 Market Coverage

An insurance contract (b,γ) is called *feasible for a* if client a prefers the contract to none. This translates into the following condition

$$pu_a(\underline{\theta}) + (1-p)u_a(\overline{\theta}) \leq (1-bp)u_a(-\gamma) + bpu_a(\underline{\theta}) \tag{B15}$$

$$\Leftrightarrow bp\left[u_a(-\gamma) - u_a(\underline{\theta})\right] + u_a(\overline{\theta}) - u_a(-\gamma) \leq p\left[u_a(\overline{\theta}) - u_a(-\gamma) + u_a(-\gamma) - u_a(\underline{\theta})\right] \tag{B16}$$

$$\Leftrightarrow (1-p)\left[u_a(\overline{\theta}) - u_a(-\gamma)\right] \leq p(1-b)\left[u_a(-\gamma) - u_a(\underline{\theta})\right] \tag{B17}$$

(B17) admits an intuitive interpretation: Client a prefers the contract to no insurance, if the expected utility gain from avoiding the bad endowment in case the seller does not default (RHS) outweighs the expected utility loss from the fee if the good endowment materializes (LHS).[11]

11 Note that from (B17) we also know that for any feasible contract $(\underline{\theta}+\gamma) < 0$. (Since $-\gamma < 0 < \overline{\theta}$, the LHS of (B17) is positive, hence, the RHS needs to be positive as well.) Indeed, we already restricted attention to $\gamma < (-\underline{\theta})$ by assumption A3.

The following proposition characterizes the client that is indifferent between insurance contract (b, γ) and no insurance.

Proposition 11. *Client a is indifferent between (b, γ) and no insurance, if*

$$\gamma = \gamma_a^{exit}(b) := (-\underline{\theta}) - \frac{1}{a} \ln\left(\frac{K(b)+1}{K(b)+\exp(-a(\overline{\theta}-\underline{\theta}))}\right) \tag{B18}$$

with $K(b) = (1-b)p/(1-p)$. $\gamma_a^{exit}(b)$ is strictly increasing in a and decreasing in b.

Proof. In light of (B17), a client a is indifferent between buying contract (b, γ) and no insurance, if

$$\frac{u_a(\overline{\theta}) - u_a(-\gamma)}{u_a(-\gamma) - u_a(\underline{\theta})} = \frac{p}{1-p}(1-b) \tag{B19}$$

$$\Leftrightarrow \frac{\exp(-a\overline{\theta}) - \exp(a\gamma)}{\exp(a\gamma) - \exp(-a\underline{\theta})} = K(b) \tag{B20}$$

$$\Leftrightarrow \frac{\exp(-a(\overline{\theta}+\gamma)) - 1}{1 - \exp(-a(\underline{\theta}+\gamma))} = K(b) \tag{B21}$$

$$\Leftrightarrow \frac{\exp(-a\Delta\theta)\exp(-a(\underline{\theta}+\gamma)) - 1}{1 - \exp(-a(\underline{\theta}+\gamma))} = K(b) \tag{B22}$$

$$\Leftrightarrow \exp(-a(\underline{\theta}+\gamma)) = \frac{K(b)+1}{K(b)+\exp(-a\Delta\theta)} \tag{B23}$$

$$\Leftrightarrow \gamma = \gamma_a^{exit}(b) := (-\underline{\theta}) - \frac{1}{a}\ln\left(\frac{K(b)+1}{K(b)+\exp(-a\Delta\theta)}\right), \tag{B24}$$

with $K(b) := (1-b)p/(1-p)$ and $\Delta\theta := (\overline{\theta} - \underline{\theta})$.

ad $\gamma_a^{exit}(b)$ increasing in a. We have

$$\frac{\partial \gamma_a^{exit}}{\partial a} = \frac{1}{a}\left[\frac{1}{a}\log\left(\frac{K(b)+1}{K(b)+\exp(-a\Delta\theta)}\right) - \frac{\exp(-a\Delta\theta)}{K(b)+\exp(-a\Delta\theta)}\Delta\theta\right]. \tag{B25}$$

With

$$y := \frac{1-\exp(-a\Delta\theta)}{K(b)+\exp(-a\Delta\theta)} \tag{B26}$$

this reads

$$\frac{\partial \gamma_a^{exit}}{\partial a} = \frac{1}{a}\left[\frac{1}{a}\ln(1+y) + \left(y - \frac{1}{K(b)+\exp(-a\Delta\theta)}\right)\Delta\theta\right] \tag{B27}$$

$$= \frac{1}{a}\left[\frac{1}{a}y\left(\frac{\log(1+y)}{y} + a\Delta\theta\right) - \frac{1}{K(b)+\exp(-a\Delta\theta)}\Delta\theta\right] \tag{B28}$$

$$= \left(\frac{1}{a}\right)^2 \frac{1}{K(b)+\exp(-a\Delta\theta)}\left[(1-\exp(-a\Delta\theta))\left(\frac{\log(1+y)}{y} + a\Delta\theta\right) - a\Delta\theta\right] \tag{B29}$$

$$= \left(\frac{1}{a}\right)^2 \frac{1}{K(b)+\exp(-a\Delta\theta)} \left[(1-\exp(-a\Delta\theta))\frac{\log(1+y)}{y} - \exp(-a\Delta\theta)a\Delta\theta\right] \tag{B30}$$

With $x := a\Delta\theta$ this expression is positive if and only if

$$\frac{\exp(x)-1}{x} > \frac{y}{\log(1+y)} \tag{B31}$$

$$\Leftrightarrow \log(1+y) > y\frac{x}{\exp(x)-1} \tag{B32}$$

$$\Leftrightarrow \log\left(\frac{K(b)+1}{K(b)+\exp(-x)}\right) > \frac{x}{\exp(x)}\frac{1}{K(b)+exp(-x)}. \tag{B33}$$

For $x = 0$ the LHS and RHS are 0. For $x > 0$ the derivative w.r.t. x of the LHS reads

$$\frac{\partial LHS}{\partial x} = \frac{\exp(-x)}{K(b)+\exp(-x))}, \tag{B34}$$

while the derivative of the RHS reads

$$\frac{\partial RHS}{\partial x} = \frac{\exp(-x)}{K(b)+\exp(-x)} \underbrace{\left[(1-x) + \frac{1}{K(b)+\exp(-x)}\frac{x}{\exp(x)}\right]}_{<1}. \tag{B35}$$

To see why the expression in brackets is smaller one, note that

$$(1-x) + \frac{1}{K(b)+\exp(-x)}\frac{x}{\exp(x)} < 1 \Leftrightarrow \frac{1}{K(b)+\exp(-x)} < \exp(x) \Leftrightarrow 0 < K(b)\exp(x),$$

which always holds and proves the claim.

ad $\gamma_a^{exit}(b)$ increasing in b. Follows directly, since

$$\frac{\partial\gamma_{\underline{a}}^{exit}}{\partial K(b)} = \frac{1-\exp(-a\Delta\theta)}{(1+K(b))(K(b)+\exp(-a\Delta\theta))} > 0 \tag{B36}$$

$$\text{and } \partial_b K(b) < 0. \tag{B37}$$

□

The result is intuitive: the fee at which a client is indifferent between the contract and no insurance is higher the more risk-averse he is. The next corollary follows as a direct consequence.

Corollary 1. *i) For fixed default probability b_i, an insurance contract (b_i, γ_i) is feasible for client a if $\gamma_i < \gamma_a^{exit}(b_i)$.*

ii) Let $a^{exit}(b_i, \gamma_i)$ be the client that is indifferent between contract (b_i, γ_i) and no insurance. For γ_i outside of $[\gamma_{\underline{a}}^{exit}(b_i), \gamma_{\overline{a}}^{exit}(b_i)]$, a^{exit} lies outside of the interval $[\underline{a}, \overline{a}]$ and is set to the respective boundary. Then clients with $a < a^{exit}(b_i, \gamma_i)$ prefer no insurance.

iii) If the fee set by the unsafer insurer, γ_2, is smaller than $\gamma_{\underline{a}}^{exit}(b_2)$, then $a^{exit} < \underline{a}$ and there is full market coverage.

In the analysis, I restrict attention to the case in which the market is fully covered.[12]

B3 Formal Results on the Illustration

Lemma 7. *i) For $\gamma_2 \in [0, \gamma^{max}]$ define*

$$\gamma_1^{\underline{a}}(\gamma_2) \text{ such that } a^*(\gamma_1^{\underline{a}}(\gamma_2), \gamma_2) = \underline{a} \tag{B38}$$
$$\gamma_1^{\overline{a}}(\gamma_2) \text{ such that } a^*(\gamma_1^{\overline{a}}(\gamma_2), \gamma_2) = \overline{a}. \tag{B39}$$

Then $\gamma_1^{\underline{a}} < \gamma_1^{\overline{a}}$ and

$$\gamma_1^{\underline{a}} \leq \gamma^{max} \quad \Leftrightarrow \quad \gamma_2 \leq \overline{\gamma_2} \tag{B40}$$
$$\text{with } \overline{\gamma_2} := \arg_\gamma\{a^*(\gamma^{max}, \gamma) = \underline{a}\} = (-\underline{\theta}) - \frac{1}{\underline{a}} \log\left[\frac{1 - \tilde{g}(\vec{b})}{\exp(-2) - \tilde{g}(\vec{b})}\right]. \tag{B41}$$

ii) Analogously, for $\gamma_1 \in [0, \gamma^{max}]$ define

$$\gamma_2^{\underline{a}}(\gamma_1) \text{ such that } a^*(\gamma_1, \gamma_2^{\underline{a}}(\gamma_1)) = \underline{a} \tag{B42}$$
$$\gamma_2^{\overline{a}}(\gamma_1) \text{ such that } a^*(\gamma_1, \gamma_2^{\overline{a}}(\gamma_1)) = \overline{a} \tag{B43}$$

Then $\gamma_2^{\underline{a}} > \gamma_2^{\overline{a}}$ and

$$\gamma_2^{\underline{a}} \geq 0 \quad \Leftrightarrow \quad \gamma_1 \geq \overline{\gamma_1} \tag{B44}$$
$$\text{with } \overline{\gamma_1} := \arg_\gamma\{a^*(\gamma, 0) = \underline{a}\} = \frac{1}{\underline{a}} \log\left[1 + \tilde{g}(\vec{b})\left(\exp(\underline{a}(-\underline{\theta})) - 1\right)\right]. \tag{B45}$$

iii) As one would expect from the picture $\gamma_2 \leq \overline{\gamma_2}$ iff $\gamma_1 \geq \overline{\gamma_1}$.

iv) insurer 1 gets the entire market if

$$\overline{\gamma_2} \leq 0 \quad \Leftrightarrow \quad \underline{a}(-\underline{\theta}) \leq \log\left[\frac{1 - \tilde{g}(\vec{b})}{\exp(-2) - \tilde{g}(\vec{b})}\right]. \tag{B46}$$

12 Later we will introduce $\gamma_2^*(\gamma^{max})$, that is, insurer 2's best response to the largest possible fee set by insurer 1. insurer 2's reaction function is increasing. Hence $\gamma_2^*(\gamma^{max})$ is the largest fee possibly set by insurer in equilibrium, and if $\gamma_2^*(\gamma^{max}) \leq \gamma_{\underline{a}}^{exit}(b_2)$ there is full market coverage anyways. Otherwise, insurer 2's reaction function remains unaltered until $\gamma_{\underline{a}}^{exit}(b_2)$. Above that point, insurer 2 potentially looses market share "from below" when increasing fees, which may induce him to set fees as best responses. Hence, we expect the reaction function to change above $\gamma_{\underline{a}}^{exit}(b_2)$, but it should leave the core of the analysis unchanged.

With $\tilde{g}(\vec{b}) < 1/8$ from assumption A1 and A2,

$$\underline{a}(-\underline{\theta}) > \log\left[\frac{1-\frac{1}{8}}{\exp(-2)-\frac{1}{8}}\right] \approx 4.4, \tag{B47}$$

ensures that the setup is interesting. This is exactly assumption A4.

Proof. ad i). First of all, we show that for a fixed $\gamma_2 \in [0, \gamma^{max}]$ such $\gamma_1^{\underline{a}}(\gamma_2), \gamma_1^{\overline{a}}(\gamma_2)$ indeed exist. Whenever clear form the context we suppress the dependence on γ_2. Note that for $a \in [\underline{a}, \overline{a}]$ $g(a, \gamma_2, \gamma_2) = 0$, while $\lim_{\gamma\to\infty} g(a, \gamma, \gamma_2) = \lim_{\gamma\to\infty} \frac{1}{c_1}(\exp(a\gamma)c_2 - 1) = \infty$ with $c_1 := \exp(-a(\underline{\theta} + \gamma_2))$ and $c_2 := \exp(-a\gamma_2)$ independent of γ. Hence, from continuity such $\gamma_1^{\underline{a}}, \gamma_1^{\overline{a}}$ exist and, since $\partial_1 g > 0$, they are also unique.

Claim. $\gamma_1^{\underline{a}} < \gamma_1^{\overline{a}}$

Proof of claim. Since $\partial_a g < 0$ we have $\tilde{g}(\vec{b}) = g(\overline{a}, \gamma_1^{\overline{a}}, \gamma_2) = g(\underline{a}, \gamma_1^{\underline{a}}, \gamma_2) > g(\overline{a}, \gamma_1^{\underline{a}}, \gamma_2)$. With $\partial_1 g > 0$ this implies $\gamma_1^{\underline{a}} < \gamma_1^{\overline{a}}$.

For the last part of the statement we have

$$\gamma_1^{\underline{a}} \leq \gamma^{max} \tag{B48}$$

$$\Leftrightarrow g(\underline{a}, \gamma^{max}, \gamma_2) \geq \tilde{g}(\vec{b}) \tag{B49}$$

$$\frac{\exp(-\underline{a}(\underline{\theta}+\gamma_2))\exp(-2) - 1}{\exp(-\underline{a}(\underline{\theta}+\gamma_2)) - 1} \geq \tilde{g}(\vec{b}) \tag{B50}$$

$$\exp(-\underline{a}(\underline{\theta}+\gamma_2))\left(\exp(-2) - \tilde{g}(\vec{b})\right) \geq 1 - \tilde{g}(\vec{b}) \tag{B51}$$

$$-\underline{a}(\underline{\theta}+\gamma_2) \geq \log\left[\frac{1-\tilde{g}(\vec{b})}{\exp(-2)-\tilde{g}(\vec{b})}\right] \tag{B52}$$

$$\gamma_2 \leq (-\underline{\theta}) - \frac{1}{\underline{a}}\log\left[\frac{1-\tilde{g}(\vec{b})}{\exp(-2)-\tilde{g}(\vec{b})}\right]. \tag{B53}$$

Note that we use $\tilde{g}(\vec{b}) < \exp(-2)$ here, which is ensured by assumptions A1 and A2.

ad ii). The argument for existence is analogous to before, so is the argument for $\gamma_2^{\underline{a}} > \gamma_2^{\overline{a}}$ except that now $\partial_2 g < 0$. For the last part we have

$$\gamma_2^{\underline{a}} \geq 0 \tag{B54}$$

$$\Leftrightarrow g(\underline{a}, \gamma_1, 0) \geq \tilde{g}(\vec{b}) \tag{B55}$$

$$\Leftrightarrow \frac{\exp(\underline{a}\gamma_1) - 1}{\exp(\underline{a}(-\underline{\theta})) - 1} \geq \tilde{g}(\vec{b}) \tag{B56}$$

$$\Leftrightarrow \exp(\underline{a}\gamma_1) \geq 1 + \tilde{g}(\vec{b})\left(\exp(\underline{a}(-\underline{\theta})) - 1\right) \tag{B57}$$

$$\Leftrightarrow \gamma_1 \geq \overline{\gamma_1} =: \frac{1}{\underline{a}}\log\left[1 + \tilde{g}(\vec{b})\left(\exp(\underline{a}(-\underline{\theta}) - 1\right)\right]. \tag{B58}$$

ad iii). We have

$$\overline{\gamma_2} \geq 0 \tag{B59}$$

$$\Leftrightarrow (-\underline{\theta}) - \frac{1}{\underline{a}} \log \left[\frac{1 - \tilde{g}(\vec{b})}{\exp(-2) - \tilde{g}(\vec{b})} \right] \geq 0 \tag{B60}$$

$$\Leftrightarrow \log \left[\frac{1 - \tilde{g}(\vec{b})}{\exp(-2) - \tilde{g}(\vec{b})} \right] \leq \underline{a}(-\underline{\theta}). \tag{B61}$$

At the same time

$$\overline{\gamma_1} \leq \gamma^{max} \tag{B62}$$

$$\Leftrightarrow 2 + \log \left[1 + \tilde{g}(\vec{b}) \left(\exp(\underline{a}(-\underline{\theta})) - 1 \right) \right] \leq \underline{a}(-\underline{\theta}) \tag{B63}$$

$$\Leftrightarrow \log \left[\exp(2) \left(1 + \tilde{g}(\vec{b}) \left(\exp(\underline{a}(-\underline{\theta})) - 1 \right) \right) \right] \leq \underline{a}(-\underline{\theta}) \tag{B64}$$

$$\Leftrightarrow \exp(2) \left(1 + \tilde{g}(\vec{b}) \left(\exp(\underline{a}(-\underline{\theta})) - 1 \right) \right) \leq \exp(\underline{a}(-\underline{\theta})) \tag{B65}$$

$$\Leftrightarrow \exp(2) \left(1 - \tilde{g}(\vec{b}) \right) \leq \exp(\underline{a}(-\underline{\theta})) \left(1 - \exp(2)\tilde{g}(\vec{b}) \right) \tag{B66}$$

$$\Leftrightarrow \frac{1 - \tilde{g}(\vec{b})}{\exp(-2) - \tilde{g}(\vec{b})} \leq \exp(\underline{a}(-\underline{\theta})) \tag{B67}$$

$$\Leftrightarrow \overline{\gamma_2} \geq 0. \tag{B68}$$

ad iv). Since the LHS of (B61) is increasing in $\tilde{g}(\vec{b})$ and under Assumption A2 $\tilde{g}(\vec{b}) < 1/8$,

$$\underline{a}(-\underline{\theta}) > \log \left[\frac{1 - \frac{1}{8}}{\exp(-2) - \frac{1}{8}} \right] \approx 4.4 \tag{B69}$$

ensures $\overline{\gamma_2} \geq 0$ for all admissible parameters and hence renders the setup interesting. □

B4 Price Equilibria are Smooth Functions of Qualities

Proposition 12 (Price Equilibrium Smooth Function in Qualities). *Without loss of generality let $b_1 = 0$. Let*

$$\mathcal{D}_1 := \left\{ b_2 \middle| \gamma_1^{\square}(b_2) \lneq \gamma^{max} \right\} \tag{B70}$$

be the set of b_2 that lead to price equilibria in the interior. Let

$$\mathcal{D}_2 := \left\{ b_2 \middle| \gamma_1^{\square}(b_2) = \gamma^{max}, (d_1\Pi_1)^{\square} \gneq 0 \right\} \tag{B71}$$

be the set of b_2 that lead to price equilibria in which insurer 1 chooses the highest admissible price. The price equilibrium $\vec{\gamma}^{\square}(\vec{b})$ as a function of quality choices $\vec{b}$ is a smooth function on $\mathcal{D}_1$ and $\mathcal{D}_2$.

Proof. Let

$$\mathcal{M} := \{(b_2, \vec{\gamma}) | b_2 \in (0, b^{max}], 0 \le \gamma_2 < \gamma_1 \le \gamma^{max}\} \tag{B72}$$

and

$$\mathcal{L}_1 := \{d_2\Pi_2 = 0\} \cap \{d_1\Pi_1 = 0\} \subset \mathcal{M} \tag{B73}$$
$$\mathcal{L}_2 := \{d_2\Pi_2 = 0\} \cap \{\gamma_1 = \gamma^{max}\} \subset \mathcal{M}. \tag{B74}$$

Then we know that price equilibria are a subset of $\mathcal{L} := \mathcal{L}_1 \cup \mathcal{L}_2$, and, that $\mathcal{L}_1$ consists of price equilibria.

Claim 1. $\mathcal{L}_1$ is a smooth submanifold of $\mathcal{M}$ and $\vec{\gamma}^{\square}$ is smooth on $\mathcal{D}_1$.
Proof of claim 1. $\mathcal{L}_1$ is the intersection of nullsets of smooth functions

$$\vec{f} := \begin{pmatrix} d_2\Pi_2 \\ d_1\Pi_1 \end{pmatrix}. \tag{B75}$$

The intersection of two nullsets $\{\vec{f} = 0\}$ is smooth if $rank(Df) = 2$. If $d_1^2\Pi_1 \neq 0$,

$$\det(D_{\vec{\gamma}}f) = \det \begin{pmatrix} d_1d_2\Pi_2 & d_2^2\Pi_2 \\ d_1^2\Pi_1 & d_2d_1\Pi_1 \end{pmatrix} \tag{B76}$$
$$= -(d_2^2\Pi_2)(d_1^2\Pi_1)\left[1 - \frac{(d_1d_2\Pi_2)}{(d_2^2\Pi_2)}\frac{(d_2d_1\Pi_1)}{(d_1^2\Pi_1)}\right] \tag{B77}$$
$$= -(d_2^2\Pi_2)(d_1^2\Pi_1)\left[1 - \underbrace{(d_2\gamma_1^{\otimes})}_{<1/\alpha}\underbrace{(d_1\gamma_2^{*})}_{<\alpha}\right] \tag{B78}$$
$$\neq 0. \tag{B79}$$

If $d_1^2\Pi_1 = 0$, then $d_1d_2\Pi_1 \neq 0$, and thus

$$\det(D_{\vec{\gamma}}f) = (d_1d_2\Pi_2)(d_2d_1\Pi_1) \neq 0. \tag{B80}$$

As shown in Proposition 2, for any $\vec{b}$ there is exactly one price equilibrium $\vec{\gamma}^{\square}(\vec{b})$ such that $(\vec{b}, \vec{\gamma}^{\square}(\vec{b})) \in \mathcal{L}$. This defines a function

$$\vec{\gamma}^{\square} : (0, b^{max}) \to \mathcal{L} \tag{B81}$$
$$b_2 \mapsto \vec{\gamma}^{\square}(0, b_2)) \tag{B82}$$

with $\vec{\gamma}^{\square} : \mathcal{D}_i \to \mathcal{L}_i$ for $i \in \{1, 2\}$. Hence, from the Implicit Function Theorem, $\vec{\gamma}^{\square}|_{\mathcal{D}_1}$ is the smooth parameterization of the submanifold $\mathcal{L}_1$.

Claim 2. $\mathcal{L}_2$ is a smooth submanifold of $\mathcal{M}$ and $\vec{\gamma}^{\square}$ is smooth on $\mathcal{D}_2$.
Proof of claim 2. The proof proceeds analogously, but now $\mathcal{L}_2$ is the intersection of nullsets

of

$$\vec{g} := \begin{pmatrix} d_2\Pi_2 \\ \gamma_1 - \gamma^{max} \end{pmatrix}, \tag{B83}$$

with

$$\det(D_{\vec{\gamma}}g) = \det\begin{pmatrix} d_1 d_2 \Pi_2 & 1 \\ d_2^2\Pi_2 & 0 \end{pmatrix} = d_2^2\Pi_2 < 0. \tag{B84}$$

□

References

Abad, J., I. Aldasoro, C. Aymanns, M. D'Errico, L. F. Rousová, P. Hoffmann, S. Langfield, M. Neychev, and T. Roukny (2016): "Shedding Light on Dark Markets: First Insights From the new EU-wide OTC Derivatives Dataset," *ESRB: Occasional Paper Series*, (2016/11).

Acharya, V., and A. Bisin (2014): "Counterparty Risk Externality: Centralized Versus Over-The-Counter Markets," *Journal of Economic Theory*, 149, 153–182.

Aoki, R., and T. J. Prusa (1997): "Sequential Versus Simultaneous Choice with Endogenous Quality," *International Journal of Industrial Organization*, 15(1), 103–121.

Atkeson, A. G., A. L. Eisfeldt, and P.-O. Weill (2015): "Entry and Exit in OTC Derivatives Markets," *Econometrica*, 83(6), 2231–2292.

Biais, B., F. Heider, and M. Hoerova (2012): "Clearing, Counterparty Risk, and Aggregate Risk," *IMF Economic Review*, 60(2), 193–222.

——— (2016): "Risk-Sharing or Risk-Taking? Counterparty Risk, Incentives, and Margins," *The Journal of Finance*, 71(4), 1669–1698.

——— (2021): "Variation Margins, Fire Sales, and Information-Constrained Optimality," *The Review of Economic Studies*, 88(6), 2654–2686.

Braithwaite, J. (2016): "The Dilemma of Client Clearing in the OTC Derivatives Markets," *European Business Organization Law Review*, 17(3), 355–378.

Braithwaite, J. P., and D. Murphy (2020): "Take on Me: OTC Derivatives Client Clearing in the European Union," *LSE Legal Studies Working Paper*.

Carapella, F., and C. Monnet (2020): "Dealers' Insurance, Market Structure, and Liquidity," *Journal of Financial Economics*, 138(3), 725–753.

CPMI, IOSCO (2022): "Client Clearing: Access and Portability," *Discussion Paper*.

Duffie, D. (2019): "Prone to Fail: The Pre-Crisis Financial System," *Journal of Economic Perspectives*, 33(1), 81–106.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005): "Over-The-Counter Markets," *Econometrica*, 73(6), 1815–1847.

Duffie, D., and H. Zhu (2011): "Does a Central Clearing Counterparty Reduce Counterparty Risk?," *The Review of Asset Pricing Studies*, 1(1), 74–95.

Dugast, J., S. Üslü, and P.-O. Weill (2022): "A Theory of Participation in OTC and Centralized Markets," *The Review of Economic Studies*, 89(6), 3223–3266.

Financial Stability Board (2018): "Incentives to Centrally Clear Over-the-Counter (OTC) Derivatives," *FSB and Standard-Setting Bodies Report*. `https://www.bis.org/publ/othp29.pdf`.

Gabszewicz, J. J., and J.-F. Thisse (1979): "Price Competition, Quality and Income Disparities," *Journal of Economic Theory*, 20(3), 340–359.

——— (1980): "Entry (and Exit) in a Differentiated Industry," *Journal of Economic Theory*, 22(2), 327–338.

Gregory, J. (2014): *Central Counterparties: Mandatory Central Clearing and Initial Margin Requirements for OTC Derivatives.* John Wiley & Sons.

Lehmann-Grube, U. (1997): "Strategic Choice of Quality when Quality is Costly: The Persistence of the High-Quality Advantage," *The RAND Journal of Economics*, pp. 372–384.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory.* Oxford university press New York, 1st edn.

Moorthy, K. S. (1988): "Product and Price Competition in a Duopoly," *Marketing Science*, 7(2), 141–168.

——— (1991): "Erratum to: Product and Price Competition in a Duopoly," *Marketing Science*, 10(3), 270.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*," *The Quarterly Journal of Economics*, 90(4), 629–649.

Schlesinger, H., and J.-M. G. Von der Schulenburg (1991): "Search Costs, Switching Costs and Product Heterogeneity in an Insurance Market," *Journal of Risk and Insurance*, 58(1), 109–119.

Shakarchi, R., and E. Stein (2003): *Complex Analysis.* Princeton University Press Princeton, NJ.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition Through Product Differentiation," *The Review of Economic Studies*, 49(1), 3–13.

——— (1983): "Natural Oligopolies," *Econometrica*, 51(5), 1469–1483.

Tirole, J. (1988): *The Theory of Industrial Organization.* MIT press, 1st edn.

ONLINE APPENDIX

C Online Appendix: Standard Model of Vertical Product Differentiation Revisited

This section clarifies which assumption in the standard model of vertical product differentiation need to be relaxed to yield endogenous market discipline. I revisit the standard model (see e.g. Tirole (1988, section 7.5.1)) and lift the assumptions of full market coverage and quality-invariant costs. The section then shows a refined principle of product differentiation and in how far upward pressure on qualities emerges.

C1 Setup

Agents. There are two firms that produce the same good, but of different qualities $s_i, i \in 1, 2$ taken from some interval $[\underline{s}, \overline{s}]$, $\underline{s} \geq 0$. There is a continuum of consumers who each demand one unit of the good. Consumers differ in their preference for quality captured by a taste parameter θ. Specifically, a consumer with taste parameter θ derives linear utility $U(p, s) = \theta s - p$ from a good of quality s sold at price p. The taste parameter is assumed to be uniformly distributed over some interval $[\underline{\theta}, \overline{\theta}]$, $\underline{\theta} \geq 0$.

Timing. There are three points in time, $t \in \{0, 1, 2\}$. At date 0, firms simultaneously choose qualities s_i. In $t = 1$, firms simultaneously choose prices p_i upon the publicly observed quality decisions in the previous period. Lastly, consumers decide from whom to buy in $t = 2$. Figure 11 summarizes the simple timing of events.

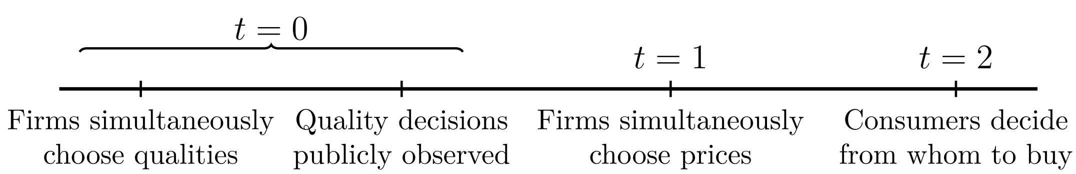


Figure 11: Timeline

If the firms choose the same level of quality, their products can potentially only differ in the price. Since consumers prefer a lower price, competition solely in prices drives the profit margins (or markups) to zero. In order to soften price competition, firms have an incentive to differentiate their products in quality. Since firms are ex-ante symmetric and do not choose the same qualities in equilibrium, if (s_1^*, s_2^*) is an equilibrium in qualities, so is (s_2^*, s_1^*). Without loss of generality we assume that firm 1 is the *low-quality* firm while firm 2 is the *high-quality* firm, that is, suppose $\Delta s := s_2 - s_1 > 0$.[13] I am interested in subgame-perfect Nash equilibria.

13 In the presence of multiple equilibria, a coordination issue emerges and one needs to break the symmetry between the two firms somehow. Here, the symmetry is broken by assigning the role of quality-leader ex-ante.

C2 Maximal Differentiation under Full Market Coverage and Constant Costs

We briefly review the driving forces at play under the standard assumptions.[14] The standard model assumes that per-unit costs c are the same for all qualities. Additionally the following restrictions on parameters are imposed:

$$\overline{\theta} = \underline{\theta} + 1 \tag{C0}$$

$$\overline{\theta} > 2\underline{\theta} \tag{A1}$$

$$c + \frac{1}{3}(\overline{s} - \underline{s})(\overline{\theta} - 2\underline{\theta}) \leq \underline{\theta}\underline{s}. \tag{A2}$$

Since (C0) and (A1) together imply $\underline{\theta} \in [0, 1)$, they can be understood as demanding that, relative to $\underline{\theta}$, there is sufficient consumer heterogeneity. As will become clear from the prices derived below, the LHS of (A2) is the highest price the low-quality firm might set in equilibrium. The RHS is the lowest possible valuation a consumer can have for the low-quality product. Hence, (A2) ensures that all consumers buy the good (*full market coverage*).

The standard result states that given quality choices $s_1 < s_2$ made in $t = 0$, the prices

$$p_1(s_1, s_2) = c + \frac{1}{3}\Delta s(\overline{\theta} - 2\underline{\theta}) \quad \text{and} \quad p_2(s_1, s_2) = c + \frac{1}{3}\Delta s(2\overline{\theta} - \underline{\theta}) \tag{C3}$$

form a Nash equilibrium in $t = 1$. In $t = 0$, there are two pure-strategy Nash equilibria in the choice of qualities and both exhibit maximal product differentiation. Specifically, for $s_1 < s_2$, firm 1 chooses the lowest possible quality $\underline{s}$ and firm 2 chooses the highest possible quality $\overline{s}$. Reversing the role of the two firms yields the other equilibrium.

The intuition of the result is as follows: In $t = 1$, when qualities $s_1 < s_2$ are already chosen, the consumer who is indifferent between the two firms is characterized by a taste parameter $\hat{\theta}$ such that $\hat{\theta}s_1 - p_1 = \hat{\theta}s_2 - p_2$, hence $\hat{\theta} = (p_2 - p_1)/\Delta s$. Firm 1 receives the consumers with θ below the threshold $\hat{\theta}$, while firm 2 receives those with $\theta > \hat{\theta}$. Firm's profits Π_1 and Π_2 take the form

$$\Pi_1(p_1, p_2) = \underbrace{(p_1 - c)}_{\text{profit margin}} \cdot \underbrace{\left[\frac{(p_2 - p_1)}{\Delta s} - \underline{\theta}\right]}_{\text{market share}}, \quad \Pi_2(p_1, p_2) = (p_2 - c) \cdot \left[\overline{\theta} - \frac{(p_2 - p_1)}{\Delta s}\right]. \tag{C4}$$

In $t = 1$, each firm chooses a price, taking the price of the other firm as given, in order to maximize profits. In $t = 0$, each firm takes into account the Nash equilibrium in prices in the next period, which gives rise to profits as a function of quality choices, specifically $\Pi_1(s_1, s_2) = \frac{1}{9}\Delta s(\overline{\theta} - 2\underline{\theta})^2$ and $\Pi_2(s_1, s_2) = \frac{1}{9}\Delta s(2\overline{\theta} - \underline{\theta})^2$. As profits are increasing in the quality differential, firm 1 chooses the lowest possible quality, while firm 2 chooses the highest possible quality. Note that as a direct consequence the quality-leader enjoys the larger profits

14 as in section 7.5.1 in Tirole (1988)

- an important observation for later.

The driving forces behind the result of maximal product differentiation are twofold. Firstly, assumption (A2) ensures that the entire market is always covered. Whatever quality choices firms make in $t = 0$ under (A2), they will always be able to optimally respond with their price choices in such a way that the indifferent consumer is left unchanged.[15] This implies that the quantity effect cancels out and only the margin effect is left. For firm 1, for example, we have

$$\frac{\partial \Pi_1(s_1)}{\partial s_1} = \underbrace{\frac{\partial (p_1(s_1) - c)}{\partial s_1}}_{\textit{margin effect}} \underbrace{[\hat{\theta}(s_1) - \underline{\theta}]}_{>0} + (p_1(s_1) - c) \underbrace{\frac{\partial [\hat{\theta}(s_1) - \underline{\theta}]}{\partial s_1}}_{=0,\ \textit{quantity effect}} . \tag{C5}$$

Since prices positively depend on the amount of product differentiation, both firms have an incentive to implement maximal product differentiation. Crucial for this result is that there is no upper limit on the price. For both firms it is optimal to increase prices in response to more product differentiation, keeping the indifferent consumer and as a result the market shares constant. Especially for the high-quality firm which charges the higher price, this means that potentially very large (also relative to costs) prices are set without the risk of loosing customers. Secondly, higher quality is not associated with higher costs.

C3 No Full Market Coverage and Costs Varying with Quality

Let's consider the following generalized setup. Suppose costs are increasing in quality, that is, suppose there is a smooth "cost" function $c : \mathbb{R}_+ \to \mathbb{R}_+$ with $c' \geq 0$ and $c'' \geq 0$ where the argument is thought of as quality. A firm incurs higher costs when choosing a higher quality, and, at a higher level of quality, increasing quality even further is even more costly.

We lift the assumption that the entire market is covered, i.e. we do not assume (C0), (A1) and (A2) anymore. In the absence of (A2), the symmetry between the two firms vanishes, since firm 1 needs to take into account that at too unfavorable quality and price choices, some consumers might not buy at all. Specifically, a consumer θ_0 is indifferent between not buying at all and buying from the low-quality firm if $p_1 = \theta_0 s_1$. Firm 1 faces only the market segment from θ_0 upwards, which alters its optimization problem to

$$\max_{p_1} \left\{ (p_1 - c(s_1)) \left[\frac{(p_2 - p_1)}{\Delta s} - \max \left\{ \underline{\theta}, \frac{p_1}{s_1} \right\} \right] \right\} . \tag{C6}$$

In order to avoid cumbersome case distinctions that do not seem to carry further intuition, we ensure that $p_1/s_1 \geq \underline{\theta}$ by assuming $\underline{\theta} = 0$.

Attention is restricted to pairs of qualities (s_1, s_2) that satisfy the following assumptions.
Assumption C1. $c(s_1)/s_1 < \overline{\theta}/2$

15 Formally, this can be seen when we insert equilibrium prices into the formula for the indifferent consumer and obtain $\hat{\theta}(s_1, s_2) = \frac{1}{3}(\underline{\theta} + \overline{\theta})$, independent of s_1, s_2.

Assumption C2. $c(s_2)/s_2 < 2\overline{\theta}$
Assumption C3.

$$\frac{\Delta c}{\Delta s} := \frac{c(s_2) - c(s_1)}{\Delta s} \in \left(2\frac{c(s_1)}{s_1} - \overline{\theta}, 2\overline{\theta} - \frac{c(s_2)}{s_2}\right) \tag{C7}$$

Assumption C3 ensures that the markups of both firms are positive. In particular, as will become clear from the equilibrium prices derived below, firm 1's markup will be positive if and only if $\Delta c/\Delta s > 2c(s_1)/s_1 - \overline{\theta}$, while firms 2's markup will be positive if and only if $\Delta c/\Delta s < 2\overline{\theta} - c(s_2)/s_2$. Assumption C3 is a condition on the difference in costs relative to the difference in quality chosen by the two firms. It means that some combinations of (s_1, s_2) kick one firm out of the market, which makes it plausible how a firm may exert a "pull effect" on the quality decisions of the other firm, as shown below. Assumptions C1 and C2 mandate that the upper and lower boundary of the admissible interval in assumption C3 are positive and negative, respectively. Since $\Delta c/\Delta s$ is positive, assumption C2 is a necessary condition, while assumption C1 is only a sufficient condition for positive profit margins of firm 2 and 1 respectively.[16] [17] Assumptions C1 - C3 can be ensured by a large enough $\overline{\theta}$, hence sufficient consumer heterogeneity.

Refined Principle of Product Differentiation

The Nash equilibrium in prices takes the following form.
Proposition 13. *Given quality choices (s_1, s_2) that satisfy assumptions C1 - C3, the following is a Nash equilibrium in prices in $t = 1$:*

$$p_1(s_1, s_2) = \frac{s_1}{3s_2 + \Delta s}\left[c(s_2) + 2\frac{s_2}{s_1}c(s_1) + \overline{\theta}\Delta s\right] \tag{C8}$$

$$= c(s_1) + \frac{s_1}{3s_2 + \Delta s}\left[\Delta c + \Delta s\left(-2\frac{c(s_1)}{s_1} + \overline{\theta}\right)\right] \tag{C9}$$

$$p_2(s_1, s_2) = \frac{s_2}{3s_2 + \Delta s}\left[2c(s_2) + c(s_1) + 2\overline{\theta}\Delta s\right] \tag{C10}$$

$$= c(s_2) + \frac{s_2}{3s_2 + \Delta s}\left[-\Delta c + \Delta s\left(-\frac{c(s_2)}{s_2} + 2\overline{\theta}\right)\right] \tag{C11}$$

Proof. The idea of the proof is analogous to the proof of the standard result presented above in the text. The details are presented in online appendix C5. □

As before, we are interested in whether the quality-leader has higher profits than the low-quality firm. The following corollary shows that this is the case as long as $\Delta c/\Delta s$ lies

16 If $c(0)$ is normalized to zero, the function $x \mapsto c(x)/x$ is increasing for positive x, since for $x > 0$ we have $\frac{\partial}{\partial x}\left(\frac{c(x)}{x}\right) = \frac{1}{x}\left[c'(x) - \frac{c(x)-c(0)}{(x-0)}\right] \geq 0$ from convexity. But we do not make this assumption here in general as it would rule out fixed costs.
17 Constant costs imply $\Delta c/\Delta s = 0$, hence, satisfy assumption C3 under assumptions C1 and C2.

closer to the lower than to the upper boundary of the admissible interval.

Corollary 2. *i) Firm 2 enjoys larger profit margins than firm 1, i.e.* $p_1 - c(s_1) < p_2 - c(s_2)$ *if and only if*

$$s_1 \left[\frac{\Delta c}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right] < s_2 \left[\left(2\overline{\theta} - \frac{c(s_2)}{s_2} \right) - \frac{\Delta c}{\Delta s} \right].$$

ii) Firm 2 enjoys larger market shares than firm 1, i.e. $\hat{\theta} - \theta_0 < \overline{\theta} - \hat{\theta}$ *if and only if*

$$\left[\frac{\Delta c}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right] < \left[\left(2\overline{\theta} - \frac{c(s_2)}{s_2} \right) - \frac{\Delta c}{\Delta s} \right]. \tag{B4}$$

iii) If firm 2 has the higher market share, i.e. if (B4) *is satisfied, it also has the higher profit margin and, as a result, higher profits.*

Proof. Follows directly from plugging in the respective formulas. □

When both firms anticipate the equilibrium in prices for given quality choices, one can express profits as a function of quality choices:

$$\Pi_1(s_1, s_2) = \Delta s \frac{s_2}{s_1} \left[\frac{s_1}{3s_2 + \Delta s} \left(\frac{c(s_2) - c(s_1)}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right) \right]^2 \tag{C12}$$

$$\Pi_2(s_1, s_2) = \Delta s \left[\frac{s_2}{3s_2 + \Delta s} \left(2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{c(s_2) - c(s_1)}{\Delta s} \right) \right]^2. \tag{C13}$$

In the original setup, profits were increasing in the quality differential. Here, in (C12) as well as in (C13), the first factor increases as products become more differentiated, but the effect on the expressions in brackets is unclear. Hence, an interior Nash equilibrium in qualities may be possible. Specifying conditions on the functional form of $c(\cdot)$ that ensure existence of an interior Nash equilibrium does not promise interesting economic results because of lenghty and tedious expressions, and I do not have a general existence proof. The following result, however, derives properties of a Nash equilibrium in qualities and shows a *refined principle of product differentiation.*

Proposition 14. *a) At any point* (s_1, s_2) *that satisfies assumption C1 - C3*

ii) if marginal costs for extra quality are small for firm 1, firm 1 wants to increase quality. Specifically,

$$c'(s_1) < 2\frac{c(s_1)}{s_1} - \overline{\theta} \qquad \Rightarrow \qquad \frac{\partial \Pi_1(s_1, s_2)}{\partial s_1} > 0. \tag{C14}$$

iiii) For firm 2, if marginal costs for extra quality are large, decreasing quality increases *profits. Specifically,*

$$2\overline{\theta} - \frac{c(s_2)}{s_2} < c'(s_2) \qquad \Rightarrow \qquad \frac{\partial \Pi_2(s_1, s_2)}{\partial s_2} < 0. \tag{C15}$$

b) *For a sequence of (s_1, s_2) where each pair of qualities satisfies assumptions C1 - C3 and stays distinct while converging to some s_0, i.e. Δs going to zero, we have*

$$\lim_{s_1, s_2 \to s_0} \frac{\partial \Pi_1(s_1, s_2)}{\partial s_1} = -\frac{1}{9}\left(c'(s_0) - 2\frac{c(s_0)}{s_0} + \overline{\theta}\right)^2 \leq 0, \tag{C16}$$

$$\lim_{s_1, s_2 \to s_0} \frac{\partial \Pi_2(s_1, s_2)}{\partial s_2} = \frac{1}{9}\left(2\overline{\theta} - \frac{c(s_0)}{s_0} - c'(s_0)\right)^2 \geq 0. \tag{C17}$$

Proof. See online appendix C6. □

The following observation follows. The threshold $c(s_1)/s_1 - \overline{\theta}$ in (C14) indeed also depends on s_1. It can be meaningfully interpreted, since by assumption C3, $\Delta c/\Delta s$ needs to lie above this threshold. Analogously for the threshold in (C15).

Proposition 4 part b) shows that, if qualities are very close together, i.e. when Δs is small, firms want to differentiate qualities. In other words, the same effect as in the original model prevails, but now it is only an "infinitesimal" effect as it holds for small differences in quality. At the same time, Proposition 4 part a) demonstrates that high or low marginal costs for firm 2 or 1, respectively, can be the driver behind a tendency to move qualities closer together. From Proposition 4 part aii) the quality-leader wants to provide only as much quality as "necessary", while from part ai) the low-quality firm provides "as much quality as feasible" with respect to the increasing marginal costs of quality. Together the forces from part a) and b) act like pull and push factors keeping the qualities of the two firms somewhat close together, but never equal.

C4 Upward Pressure on Qualities

Two questions arise naturally. Firstly, since the low-quality firm now experiences competition from above (the high-quality firm) and below (the option not to buy), does that exert a pull effect on the quality choice of firm 1? Secondly, when the leadership position in quality is the more attractive one, can the threat to be overtaken by the other firm induce the quality-leader to set high qualities whatsoever? The interplay of these forces would produce upward pressure on qualities.

The following proposition and subsequent discussion clarifies in how far there may be a pull effect on the quality choice of the low-quality firm.

Proposition 15. *At any point (s_1, s_2) that satisfies assumptions C1 - C3, if*

$$K := \overline{\theta}\,\underbrace{(s_2 - 2s_1)}_{=:A} + \underbrace{\left(2s_2 - \Delta s\frac{s_1}{s_2}\right)}_{>0}\underbrace{\left[\frac{c(s_1)}{s_1} - c'(s_1)\right]}_{=:B} + \underbrace{\Delta s\frac{s_1}{s_2}}_{>0}\underbrace{\left[\frac{\Delta c}{\Delta s} - c'(s_1)\right]}_{=:C} + \underbrace{2\Delta s}_{>0}\underbrace{(-c'(s_1))}_{=:D}$$

is non-negative, then $\partial\Pi_1/\partial s_1 > 0$ and subsequently the point can not be an equilibrium.

Proof. See online appendix C7. □

We discuss the consequences for the special case of constant costs $c \in \mathbb{R}_+$, quadratic costs and the general case. For constant costs, K reduces to $\overline{\theta}(s_2 - 2s_1) + (2s_2 - (\Delta s)s_1/s_2)\,c/s_1$. For $s_2 \geq 2s_1$ this expression is positive, subsequently the point can not be an equilibrium. This admits the following interpretation: In order for an equilibrium to exist, the low-quality firm needs to choose s_1 sufficiently close to the quality of firm 2, i.e. larger than $0.5\,s_2$ (*pull effect*).[18]

For quadratic costs, which play a prominent role in the literature on the subject, say $c(s) = \tau s^2$, K reduces to $K = \overline{\theta}(s_2 - 2s_1) - \tau s_1 \underbrace{(5s_2 - 3s_1)}_{>0}$. So $K \geq 0$ requires $(s_2 - 2s_1) > 0$ and is fulfilled if $0 \leq \tau < \overline{\theta}(s_2 - 2s_1)/(5s_2 - 3s_1)$. This again has an intuitive interpretation when we think of the costs $c(s) = \tau s^2$ as a quadratic "error term" to zero costs with "intensity" τ. A non-negative K requires that the condition $s_2 \geq 2s_1$, which precludes an equilibrium for zero costs, still suffices to preclude existence for quadratic costs provided the "intensity" τ of the "error term" is below some threshold.

For the general case, K consists of "drivers" A, B, C and D, as defined above, with positive weights. For a fixed s_2, each driver is monotone in s_1 and the level of s_1 determines whether the corresponding driver increases or decreases K, i.e. whether it exerts upward pressure or not. Specifically, A is positive iff $s_1 < 1/2\,s_2$, B is positive iff s_1 is smaller than s_0 with s_0 such that $c'(s_0) = c(s_0)/s_0$, C is positive for $s_1 \neq s_2$ and D is always negative.

That the quality-leader exerts a "pull effect" on the low-quality firm upwards rather than the other way around is intuitive also from a different point of view. Already in the original model the quality-leader enjoys greater profits. Albeit the fact that the low-quality firm will subsequently choose the lowest quality there, this indicates that there is room for a race for the "pole position in quality", as also noted in Tirole (1988, p. 297). Corollary 2 shows that this result persists in the generalized setup under the condition that the relation $\Delta c/\Delta s$ may not be too large. Specifically, if $\Delta c/\Delta s$ lies closer to the lower than to the upper boundary of the admissible interval, the lead position in quality is the more attractive one and the quality-leader will try to keep this "pole position". It seems plausible that the quality-leader is aware of the risk of being overtaken by the other firm at too low quality choices. Then

18 In the case of constant costs, one can easily show that firm 2 chooses the maximal quality. This is intuitive, as higher quality is not associated with higher costs in this case. The simplification of constant costs helps show the key idea of a "pull" effect exerted on the low-quality firm most clearly, but it also eliminates the force that previously counteracted the quality-leader's incentive to choose the extreme quality.

the risk of being overtaken may exert upward pressure on the quality choices when moving qualities closer together. This is shown formally in the sequel.

To capture this, suppose we break the symmetry between the two firms not, as done so far, by assigning the roles of quality-leader and quality-follower ex-ante, but instead by making the quality choice sequential. We call the new setup *sequential game without assigned roles* and assume firm 2 has a first mover advantage in the choice of quality. Specifically, we introduce an additional time period $t = (-1)$ in which firm 2 chooses its quality, while firm 1, upon observing firm 2's decision, continues to choose its quality in $t = 0$. The rest remains as before.

In $t = 0$, firm 1 can either "adapt" by actually becoming the quality-follower or overtake firm 2's leadership position by choosing a higher quality. We ensure assumptions C1 - C3 and (B4) for all quality pairs by assuming that for all s in $[\underline{s}, \overline{s}]$

$$\frac{c(s)}{s} < \frac{\overline{\theta}}{2} \tag{B1'}$$

$$c'(s) \in \left(2 \sup_t \frac{c(t)}{t} - \overline{\theta}, 2\overline{\theta} - \inf_t \frac{c(t)}{t}\right) \tag{B3'}$$

$$\overline{\theta} - 2 \inf_t \frac{c(t)}{t} + c'(s) < 2\overline{\theta} - \sup_t \frac{c(t)}{t} - c'(s). \tag{B4'}$$

(B1') - (B4') relate marginal costs of a further quality improvement to $\overline{\theta}$, the marginal willingness to pay of the most quality-sensitive consumer for a quality improvement. Note that with $c(s_2) - c(s_1) = \int_{s_1}^{s_2} c'(t)dt$, (B3') yields assumption C3 for all $s \in [\underline{s}, \overline{s}]$, while (B4') ensures condition (B4) for all qualities. Conditions (B1'), (B3') and (B4') can be ensured if $\overline{\theta}$ is large enough.[19]

Hence, the quality-leader always enjoys larger profits, which enables us to derive the following proposition.

Proposition 16. *A necessary condition for some (s_1, s_2) to be a subgame-perfect Nash equilibrium in the sequential game without assigned roles, is that*

$$s_2 > \frac{4}{5}\overline{s}. \tag{C18}$$

Proof. As before, the main idea is presented below in the text, while some calculations are relegated to online appendix C8. □

The intuition of the result is as follows: Suppose $s_1 < s_2$ is a Nash equilibrium in the sequential game without assigned roles. In that case one must not be able to find a

19 In the same spirit as in the original model, this can be interpreted as a condition on sufficient consumer heterogeneity, and thereby neatly connects to the set of assumptions made in the original model. There, (C0) and (A1) demand sufficient consumer heterogeneity while (A2) demands full market coverage; here, only sufficient consumer heterogeneity is needed.

profitable deviation for the quality-follower, that is, no s_3 with $s_2 < s_3 \leq \overline{s}$ such that the profit when taking the lead position in quality, exceeds the profit when choosing the optimal quality as quality-follower, that is no s_3 $\Pi_2(s_2, s_3) > \Pi_1(s_1, s_2)$. As shown in the appendix, $s_3 = (s_2^2 + s_1 s_2 - s_1^2)/s_2$ is such an profitable deviation, which is infeasible if $(4/5)\overline{s} < s_2$.

Proposition 16 shows that in the sequential game without assigned roles, a necessary condition for a Nash equilibrium to exist is that the quality-leader chooses a quality at least as high as 80% of the maximal quality. In other words, the threat of being overtaken and loosing the leadership position in quality induces the first mover to pick a high quality even in an environment where costs are increasing and convex in the level of quality.

The interplay between a pull effect on the quality choice of the low-quality firm and pressure on the high-quality firm not to leave too much room quality-wise above, gives rise to upward pressure on the quality choices.

C5 Proof of Proposition 13

The full maximization problem reads

$$\max_{p_1} \Pi_1(p_1, p_2) = \max_{p_1} \left\{ (p_1 - c(s_1)) \left[\frac{(p_2 - p_1)}{\Delta s} - \frac{p_1}{s_1} \right] \right\} \tag{C19}$$

$$\max_{p_2} \Pi_2(p_1, p_2) = \max_{p_2} \left\{ (p_2 - c(s_2)) \left[\overline{\theta} - \frac{(p_2 - p_1)}{\Delta s} \right] \right\}, \tag{C20}$$

with the additional conditions

$(p_1 - c(s_1)) \geq 0$	positive profit margin of firm 1	(Bi)
$(p_2 - c(s_2)) \geq 0$	positive profit margin of firm 2	(Bii)
$\frac{p_2 - p_1}{\Delta s} \geq \frac{p_1}{s_1}$	positive market share of firm 1	(Biii)
$\overline{\theta} \geq \frac{p_2 - p_1}{\Delta s}$	positive market share of firm 2	(Biv)
$\frac{p_1}{s_1} \geq \underline{\theta}$	firm 1's market share takes the form $\frac{(p_2 - p_1)}{\Delta s} - \frac{p_1}{s_1}$	(Bv)
$\frac{p_2 - p_1}{\Delta s} \geq \frac{p_2}{s_2}$	firm 2's market share takes the form $\overline{\theta} - \frac{(p_2 - p_1)}{\Delta s}$.	(Bvi)

I first solve the unconstrained maximization problem and then verify that the (unique) solution satisfies (Bi) - (Bvi). Solving the reaction functions

$$p_1 = R_1(p_2) := \frac{1}{2} \left[p_2 \frac{s_1}{s_2} + c(s_1) \right] \tag{C21}$$

$$p_2 = R_2(p_1) := \frac{1}{2} \left[p_1 + c(s_2) + \overline{\theta} \Delta s \right] \tag{C22}$$

yields the formula for the prices.

It remains to check whether conditions (Bi) - (Bvi) hold. (Bi) and (Bii) are ensured by (C3) as argued in the text. Since plugging in the respective formulas directly yields

$$\hat{\theta} - \theta_0 = \frac{s_2}{(3s_2 + \Delta s)} \left[\frac{\Delta c}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right] \tag{C23}$$

$$\overline{\theta} - \hat{\theta} = \frac{s_2}{(3s_2 + \Delta s)} \left[\left(2\overline{\theta} - \frac{c(s_2)}{s_2} \right) - \frac{\Delta c}{\Delta s} \right], \tag{C24}$$

(C3) ensures (Biii) and (Biv). (Bv) follows directly from the assumption $\underline{\theta} = 0$, since prices are positive. It remains to show (Bvi), which is a little more cumbersome. As a first step note that (Bvi) follows if we know that

$$\frac{p_2}{p_1} \geq \frac{s_2}{s_1}, \tag{C25}$$

since then

$$p_2 s_1 \geq p_1 s_2 \tag{C26}$$

$$\Leftrightarrow p_2 s_1 - p_2 s_2 + p_2 s_2 \geq p_1 s_2 \tag{C27}$$

$$\Leftrightarrow -p_2 \Delta s + s_2 \Delta p \geq 0 \tag{C28}$$

$$\Leftrightarrow \frac{\Delta p}{\Delta s} \geq \frac{p_2}{s_2}. \tag{C29}$$

It remains to show that (C25) holds. To that end we have

$$\frac{p_2}{p_1} \geq \frac{s_2}{s_1} \tag{C30}$$

$$\Leftrightarrow \frac{\frac{s_2}{3s_2+\Delta s}\left[2c(s_2) + c(s_1) + 2\overline{\theta}\Delta s\right]}{\frac{s_1}{3s_2+\Delta s}\left[c(s_2) + 2\frac{s_2}{s_1}c(s_1) + \overline{\theta}\Delta s\right]} \geq \frac{s_2}{s_1} \tag{C31}$$

$$\Leftrightarrow \frac{2c(s_2) + c(s_1) + 2\overline{\theta}\Delta s}{c(s_2) + 2\frac{s_2}{s_1}c(s_1) + \overline{\theta}\Delta s} \geq 1 \tag{C32}$$

$$\Leftrightarrow c(s_2) + c(s_1) \underbrace{\left(1 - 2\frac{s_2}{s_1}\right)}_{= -\frac{(s_2+\Delta s)}{s_1}} + \overline{\theta}\Delta s \geq 0 \tag{C33}$$

$$\Leftrightarrow c(s_2) - \frac{s_2}{s_1}c(s_1) + \frac{c(s_1)}{s_1}(-\Delta s + 2\Delta s) + \Delta s \left[\overline{\theta} - 2\frac{c(s_1)}{s_1}\right] \geq 0 \tag{C34}$$

$$\Leftrightarrow \frac{\Delta c}{\Delta s} \geq 2\frac{c(s_1)}{s_1} - \overline{\theta}, \tag{C35}$$

which is ensured by (C3).

□

C6 Proof of Proposition 14

Part a) follows immediately, if we know the following expressions for the derivatives of the profits. With α and β the expressions inside the squared brackets in (C12) and (C13), namely

$$\alpha(s_1, s_2) := \frac{s_1}{3s_2 + \Delta s} \left(\frac{\Delta c}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right) \tag{C36}$$

$$\beta(s_1, s_2) := \frac{s_2}{3s_2 + \Delta s} \left(2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{\Delta c}{\Delta s} \right). \tag{C37}$$

we claim that

$$\begin{aligned}
\frac{\partial \Pi_1(s_1, s_2)}{\partial s_1} &= -\frac{s_2^2}{s_1^2}\alpha^2 + 2\alpha \frac{s_2}{s_1} \Delta s \frac{\partial \alpha(s_1, s_2)}{\partial s_1} && \text{(C38)} \\
&= \underbrace{\frac{s_2}{s_1} \frac{\alpha}{(3s_2 + \Delta s)^2}}_{>0} \Bigg[(2\Delta s(3s_2 + \Delta s) + 3s_1 s_2) \underbrace{\left(\frac{\Delta c}{\Delta s} - c'(s_1) \right)}_{\geq 0 \text{ from convexity}} && \text{(C39)} \\
&\quad + s_2(3s_2 + \Delta s) \left(2\frac{c(s_1)}{s_1} - \overline{\theta} - c'(s_1) \right) + 4s_2 \Delta s \underbrace{\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - c'(s_1) \right)}_{>0 \text{ from (C3)}} \Bigg],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Pi_2(s_1, s_2)}{\partial s_2} &= \beta^2 + 2\beta \Delta s \frac{\partial \beta(s_1, s_2)}{\partial s_2} && \text{(C40)} \\
&= \underbrace{\frac{\beta}{(3s_2 + \Delta s)^2}}_{>0} \Bigg[(3s_2 + \Delta s)(s_2 + 2\Delta s) \underbrace{\left(\frac{\Delta c}{\Delta s} - c'(s_2) \right)}_{\leq 0 \text{ from convexity}} && \text{(C41)} \\
&\quad + 4s_1 \Delta s \underbrace{\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s} \right)}_{<0 \text{ from (C3)}} + (3s_2 + \Delta s)s_2 \left(2\overline{\theta} - \frac{c(s_2)}{s_2} - c'(s_2) \right) \Bigg].
\end{aligned}$$

To show this, note that for firm 1 the derivative can be written as follows

$$\begin{aligned}
\frac{\partial \Pi_1(s_1, s_2)}{\partial s_1} &= -\frac{s_2^2}{s_1^2}\alpha^2 + 2\alpha \frac{s_2}{s_1} \Delta s \frac{\partial \alpha(s_1, s_2)}{\partial s_1} && \text{(C42)} \\
&\overset{\text{Def } \alpha}{=} \frac{s_2}{s_1} \frac{\alpha}{(3s_2 + \Delta s)^2} \Bigg[-\frac{s_2}{s_1}(3s_2 + \Delta s)s_1 \left(\frac{\Delta c}{\Delta s} - \left(2\frac{c(s_1)}{s_1} - \overline{\theta} \right) \right) \\
&\quad + 2\Delta s \frac{\partial \alpha(s_1, s_2)}{\partial s_1} (3s_2 + \Delta s)^2 \Bigg]. && \text{(C43)}
\end{aligned}$$

For the derivative of $\alpha(s_1, s_2)$ w.r.t. s_1 it proves helpful to use two versions of the formula for α when applying the product rule, namely

$$\alpha(s_1, s_2) := \frac{1}{3s_2 + \Delta s} \left(s_1 \frac{\Delta c}{\Delta s} - \left(2c(s_1) - s_1 \overline{\theta} \right) \right) \tag{C44}$$

$$= \frac{1}{3s_2 + \Delta s}\left(\frac{s_1}{\Delta s}c(s_2) - \frac{s_2 + \Delta s}{\Delta s}c(s_1) + s_1\overline{\theta}\right). \tag{C45}$$

Then

$$\begin{aligned}
\frac{\partial \alpha(s_1, s_2)}{\partial s_1}(3s_2 + \Delta s)^2 &= \left[\overline{\theta} + \frac{s_2}{(\Delta s)^2}c(s_2) - \frac{s_2}{(\Delta s)^2}c(s_1) - \frac{s_2 + \Delta s}{\Delta s}c'(s_1)\right](3s_2 + \Delta s) \\
&\quad + s_1\overline{\theta} - 2c(s_1) + s_1\frac{\Delta c}{\Delta s} \\
&= 4\overline{\theta}s_2 - 2c(s_1) + \frac{(s_2 + \Delta s)(3s_2 + \Delta s)}{\Delta s}\left[\frac{\Delta c}{\Delta s} - c'(s_1)\right] \\
&\quad - (3s_2 + \Delta s)\frac{\Delta c}{\Delta s} + s_1\frac{\Delta c}{\Delta s} \\
&= \frac{(s_2 + \Delta s)(3s_2 + \Delta s)}{\Delta s}\left[\frac{\Delta c}{\Delta s} - c'(s_1)\right] \\
&\quad + \underbrace{[-(3s_2 + \Delta s) + s_1 + 2\Delta s]}_{=-2s_2}\frac{\Delta c}{\Delta s} + 2s_2\left(2\overline{\theta} - \frac{c(s_2)}{s_2}\right) \\
&= \frac{(s_2 + \Delta s)(3s_2 + \Delta s)}{\Delta s}\left[\frac{\Delta c}{\Delta s} - c'(s_1)\right] + 2s_2\left[2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{\Delta c}{\Delta s}\right].
\end{aligned}$$

Hence together with (C43)

$$\begin{aligned}
\frac{\partial \Pi_1(s_1, s_2)}{\partial s_1} &= \frac{s_2}{s_1}\frac{\alpha}{(3s_2 + \Delta s)^2}\Bigg[s_2(3s_2 + \Delta s)\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right) && \text{(C46)} \\
&\quad + 2(s_2 + \Delta s)(3s_2 + \Delta s)\left(\frac{\Delta c}{\Delta s} - c'(s_1)\right) \\
&\quad + (2\Delta s)2s_2\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{\Delta c}{\Delta s}\right)\Bigg] \\
&= \frac{s_2}{s_1}\frac{\alpha}{(3s_2 + \Delta s)^2}\Bigg[s_2\Delta s\left(2\frac{c(s_1)}{s_1} - \overline{\theta}\right) + s_2\Delta s\left(2\overline{\theta} - \frac{c(s_2)}{s_2}\right) && \text{(C47)} \\
&\quad - s_2(3s_2 + \Delta s)\frac{\Delta c}{\Delta s} + 2s_2\left[3s_2\left(2\frac{c(s_1)}{s_1} - \overline{\theta}\right) + 3\Delta s\left(2\overline{\theta} - \frac{c(s_2)}{s_2}\right)\right] \\
&\quad - 4s_2\Delta s\frac{\Delta c}{\Delta s} - 2(s_2 + \Delta s)(3s_2 + \Delta s)c'(s_1) \\
&\quad + 2s_2(3s_2 + \Delta s)\frac{\Delta c}{\Delta s} + 2\Delta s(3s_2 + \Delta s)\frac{\Delta c}{\Delta s}\Bigg] \\
&= \frac{s_2}{s_1}\frac{\alpha}{(3s_2 + \Delta s)^2}\Bigg[2\Delta s(3s_2 + \Delta s)\left(\frac{\Delta c}{\Delta s} - c'(s_1)\right) && \text{(C48)} \\
&\quad - 2s_2(3s_2 + \Delta s)c'(s_1) + s_2(3s_2 + \Delta s)\left(2\frac{c(s_1)}{s_1} - \overline{\theta}\right)
\end{aligned}$$

$$+s_2\left(3\Delta s+\Delta s\right)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}\right)-4s_2\Delta s\frac{\Delta c}{\Delta s}+s_2(3s_2+\Delta s)\frac{\Delta c}{\Delta s}\Bigg]$$

$$= \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\Bigg[2\Delta s(3s_2+\Delta s)\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right) \tag{C49}$$

$$+s_2(3s_2+\Delta s)\left(2\frac{c(s_1)}{s_1}-\overline{\theta}-c'(s_1)\right)$$

$$+s_2(3s_2+\Delta s)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)$$

$$+\underbrace{(4s_2\Delta s-s_2(3s_2+\Delta s))}_{=-3s_1s_2}\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)\Bigg]$$

$$= \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\Bigg[2\Delta s(3s_2+\Delta s)\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right) \tag{C50}$$

$$+s_2(3s_2+\Delta s)\left(2\frac{c(s_1)}{s_1}-\overline{\theta}-c'(s_1)\right)$$

$$+s_2(3s_2+\Delta s)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)$$

$$-3s_1s_2\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)+3s_1s_2\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right)\Bigg]$$

$$= \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\Bigg[\left(2\Delta s(3s_2+\Delta s)+3s_1s_2\right)\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right) \tag{C51}$$

$$+s_2(3s_2+\Delta s)\left(2\frac{c(s_1)}{s_1}-\overline{\theta}-c'(s_1)\right)$$

$$+\underbrace{\left(3s_2^2+s_2\Delta s-3s_1s_2\right)}_{=4s_2\Delta s}\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)\Bigg].$$

For firm 2 the proof follows analogous steps but now

$$\frac{\partial\Pi_2(s_1,s_2)}{\partial s_2} = \beta^2+2\beta\Delta s\frac{\partial\beta(s_1,s_2)}{\partial s_2} \tag{C52}$$

$$\stackrel{\text{Def }\beta}{=} \frac{\beta}{(3s_2+\Delta s)^2}\Bigg[(3s_2+\Delta s)s_2\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)$$

$$+2\Delta s\frac{\partial\beta(s_1,s_2)}{\partial s_2}(3s_2+\Delta s)^2\Bigg],$$

the two versions of β read

$$\beta(s_1,s_2) := \frac{1}{3s_2+\Delta s}\left(2s_2\overline{\theta}-c(s_2)-s_2\frac{c(s_2)-c(s_1)}{\Delta s}\right) \tag{C53}$$

$$= \frac{1}{3s_2+\Delta s}\left(2s_2\overline{\theta}+\frac{s_2}{\Delta s}c(s_1)-\frac{s_2+\Delta s}{\Delta s}c(s_2)\right), \tag{C54}$$

for the derivative of β w.r.t. s_2 we have

$$\begin{aligned}
\frac{\partial \beta(s_1,s_2)}{\partial s_2}(3s_2+\Delta s)^2 &= \left[2\overline{\theta} - \frac{s_1}{(\Delta s)^2}c(s_1) + \frac{s_1}{(\Delta s)^2}c(s_2) - \frac{s_2+\Delta s}{\Delta s}c'(s_2)\right](3s_2+\Delta s) \\
&\quad -4\left(2\overline{\theta}s_2 - c(s_2) - s_2\frac{\Delta c}{\Delta s}\right) \\
&= -\frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}c'(s_2) + \underbrace{\frac{s_1(3s_2+\Delta s)+4s_2\Delta s}{\Delta s}}_{=\frac{(s_2+\Delta s)(3s_2-\Delta s)}{\Delta s}}\left(\frac{\Delta c}{\Delta s}\right) \\
&\quad -2\overline{\theta}s_1 + 4c(s_2) \\
&= \frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right) - 2\frac{(s_2+\Delta s)\Delta s}{\Delta s}\left(\frac{\Delta c}{\Delta s}\right) \\
&\quad -2\overline{\theta}s_1 + 4c(s_2) \\
&= \frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right) \\
&\quad -2s_1\left(\overline{\theta} + \frac{\Delta c}{\Delta s}\right)\underbrace{+2(2s_1)\frac{\Delta c}{\Delta s} - 4s_2\frac{\Delta c}{\Delta s} + 4c(s_2)}_{=-4\Delta c+4c(s_2)=4c(s_1)} \\
&= \frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right) + 2s_1\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right)
\end{aligned}$$

and together with (C53) this yields

$$\begin{aligned}
\frac{\partial \Pi_2(s_1,s_2)}{\partial s_2} &= \frac{\beta}{(3s_2+\Delta s)^2}\Bigg[(3s_2+\Delta s)s_2\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{\Delta c}{\Delta s}\right) && \text{(C55)} \\
&\quad -2(s_2+\Delta s)(3s_2+\Delta s)\left(c'(s_2) - \frac{\Delta c}{\Delta s}\right) + (2\Delta s)2s_1\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right)\Bigg] \\
&= \frac{\beta}{(3s_2+\Delta s)^2}\Bigg[(3s_2+\Delta s)s_2\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - c'(s_2)\right) && \text{(C56)} \\
&\quad -s_2(3s_2+\Delta s)\frac{\Delta c}{\Delta s} + s_2(3s_2+\Delta s)c'(s_2) \\
&\quad +(2s_2+2\Delta s)(3s_2+\Delta s)\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right) + 4s_1\Delta s\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right)\Bigg] \\
&= \frac{\beta}{(3s_2+\Delta s)^2}\Bigg[(3s_2+\Delta s)s_2\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - c'(s_2)\right) && \text{(C57)} \\
&\quad +(3s_2+\Delta s)(s_2+2\Delta s)\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right) + 4s_1\Delta s\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right)\Bigg].
\end{aligned}$$

For the limits in part b) note that for firm 2, if the limit exists, (C38) implies

$$\lim_{s_1,s_2\to s_0} \frac{\partial \Pi_1(s_1,s_2)}{\partial s_1} = -\lim_{s_1,s_2\to s_0} \alpha(s_1,s_2)^2 + \lim_{s_1,s_2\to s_0} 2\alpha\Delta s \frac{\partial \alpha(s_1,s_2)}{\partial s_1}, \tag{C58}$$

with

$$\begin{aligned}
\lim_{s_1,s_2\to s_0} \alpha(s_1,s_2) &= \frac{1}{3}\left(c'(s_0) - 2\frac{c(s_0)}{s_0} + \overline{\theta}\right) =: K_1 \\
\lim_{s_1,s_2\to s_0} \frac{\partial \alpha(s_1,s_2)}{\partial s_1} &= \lim_{s_1,s_2\to s_0}\left[\underbrace{\frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}\left(\frac{\Delta c}{\Delta s} - c'(s_1)\right)}_{\to 3s_2^2\cdot 0} + 2s_2\underbrace{\left(2\overline{\theta} - \frac{c(s_2)}{s_2} - \frac{\Delta c}{\Delta s}\right)}_{\to 2\overline{\theta} - \frac{c(s_0)}{s_0} - c'(s_0)}\right] \\
&= 2s_0\left(2\overline{\theta} - \frac{c(s_0)}{s_0} - c'(s_0)\right) =: K_2.
\end{aligned}$$

Plugged into (C58) this yields

$$\lim_{s_1,s_2\to s_0} \frac{\partial \Pi_1(s_1,s_2)}{\partial s_1} = -K_1^2 + 2K_1K_2 \lim_{s_1,s_2\to s_0} \Delta s = -K_1^2. \tag{C59}$$

Analogously for firm 2 we know from (C40) that, if the limit exists,

$$\lim_{s_1,s_2\to s_0} \frac{\partial \Pi_2(s_1,s_2)}{\partial s_2} = \lim_{s_1,s_2\to s_0} \beta(s_1,s_2)^2 + \lim_{s_1,s_2\to s_0} 2\beta\Delta s \frac{\partial \beta(s_1,s_2)}{\partial s_2}, \tag{C60}$$

with

$$\begin{aligned}
\lim_{s_1,s_2\to s_0} \beta(s_1,s_2) &= \frac{1}{3}\left(2\overline{\theta} - \frac{c(s_0)}{s_0} - c'(s_0)\right) =: K_3, \\
\lim_{s_1,s_2\to s_0} \frac{\partial \beta(s_1,s_2)}{\partial s_2} &= \lim_{s_1,s_2\to s_0}\left[\underbrace{\frac{(s_2+\Delta s)(3s_2+\Delta s)}{\Delta s}\left(\frac{\Delta c}{\Delta s} - c'(s_2)\right)}_{\to 3s_2^2\cdot 0} + 2s_1\underbrace{\left(2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta c}{\Delta s}\right)}_{\to 2\frac{c(s_0)}{s_0} - \overline{\theta} - c'(s_0)}\right] \\
&= 2s_0\left(2\frac{c(s_0)}{s_0} - \overline{\theta} - c'(s_0)\right) =: K_4.
\end{aligned}$$

Plugged into (C60) this yields

$$\lim_{s_1,s_2\to s_0} \frac{\partial \Pi_2(s_1,s_2)}{\partial s_2} = K_3^2 + 2K_3K_4 \lim_{s_1,s_2\to s_0} \Delta s = K_3^2, \tag{C61}$$

which concludes the proof.

□

C7 Proof of Proposition 15

The proposition is a direct consequence of the following claim.

Claim. $\partial\Pi_1/\partial s_1$ can be bounded from below as follows

$$\frac{\partial\Pi_1(s_1,s_2)}{\partial s_1} \geq \underbrace{\frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}}_{>0}\left[\underbrace{(3s_2+\Delta s)}_{>0}K+\underbrace{s_1\Delta s\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)}_{>0}\right] \quad \text{(C62)}$$

with K as defined in the proposition.

Proof of claim. For the lower bound of $\partial\Pi_1/\partial s_1$, we start with (C49) to obtain

$$\begin{aligned}
\frac{\partial\Pi_1(s_1,s_2)}{\partial s_1} &\overset{\text{(C49)}}{=} \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\Bigg[2\Delta s(3s_2+\Delta s)\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right)\\
&\quad +s_2(3s_2+\Delta s)\left(2\frac{c(s_1)}{s_1}-\overline{\theta}-c'(s_1)\right)\\
&\quad +s_2(3s_2+\Delta s)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)-3s_1s_2\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)\Bigg]\\
&= \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\Bigg[2\Delta s(3s_2+\Delta s)\left(\frac{\Delta c}{\Delta s}-c'(s_1)\right) && \text{(C63)}\\
&\quad +s_2(3s_2+\Delta s)\left(2\frac{c(s_1)}{s_1}-\overline{\theta}-c'(s_1)\right)\\
&\quad +s_2(3s_2+\Delta s)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-c'(s_1)\right)\\
&\quad -(3s_2+\Delta s)\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)+s_1\Delta s\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)\Bigg]\\
&= \frac{s_2}{s_1}\frac{\alpha}{(3s_2+\Delta s)^2}\left[(3s_2+\Delta s)K+s_1\Delta s\left(2\overline{\theta}-\frac{c(s_2)}{s_2}-\frac{\Delta c}{\Delta s}\right)\right] && \text{(C64)}
\end{aligned}$$

with

$$\begin{aligned}
K &= \overline{\theta}(s_2-2s_1)+c(s_2)\left(\frac{s_1}{s_2}-1\right)+2\frac{s_2}{s_1}c(s_1)-2(s_2+\Delta s)c'(s_1)+\Delta c\left(2+\frac{s_1}{\Delta s}\right) && \text{(C65)}\\
&= \overline{\theta}(s_2-2s_1)-2(s_2+\Delta s)c'(s_1) && \text{(C66)}\\
&\quad +\frac{1}{s_2\Delta s}\left[c(s_2)\left(s_2^2-s_1^2+s_1s_2\right)+c(s_1)\left(2\frac{s_2^3}{s_1}-4s_2^2+s_1s_2\right)\right]\\
&= \overline{\theta}(s_2-2s_1)-2(s_2+\Delta s)c'(s_1) && \text{(C67)}\\
&\quad +\frac{1}{s_2\Delta s}\Bigg[s_2^2\underbrace{\left(c(s_2)+2c(s_1)\left(\frac{s_2}{s_1}-2\right)\right)}_{\geq c(s_1)\left[1+2\frac{s_2}{s_1}-4\right]=c(s_1)\left[\frac{2\Delta s-s_1}{s_1}\right]}+c(s_2)s_1\Delta s+c(s_1)s_1s_2\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\geq \bar{\theta}(s_2 - 2s_1) - 2(s_2 + \Delta s)c'(s_1) && \text{(C68)}\\
&\quad + \frac{1}{s_2 \Delta s}\left[c(s_1)\frac{s_2^2}{s_1}(2\Delta s) + c(s_2)s_1\Delta s + c(s_1)(-s_2\Delta s)\right] \\
&= \bar{\theta}(s_2 - 2s_1) + 2s_2\left(\frac{c(s_1)}{s_1} - c'(s_1)\right) - 2c'(s_1)\Delta s + \frac{1}{s_2}\underbrace{(c(s_2)s_1 - c(s_1)s_2)}_{=c(s_2)s_1 - c(s_1)s_1 + c(s_1)s_1 - c(s_1)s_2} && \text{(C69)}\\
&= \bar{\theta}(s_2 - 2s_1) + 2s_2\left(\frac{c(s_1)}{s_1} - c'(s_1)\right) - 2c'(s_1)\Delta s + \frac{s_1\Delta s}{s_2}\left(\frac{\Delta c}{\Delta s} - \frac{c(s_1)}{s_1}\right) && \text{(C70)}\\
&= \bar{\theta}(s_2 - 2s_1) + \left(2s_2 - \Delta s\frac{s_1}{s_2}\right)\left[\frac{c(s_1)}{s_1} - c'(s_1)\right] + \Delta s\frac{s_1}{s_2}\left[\frac{\Delta c}{\Delta s} - c'(s_1)\right] + 2\Delta s\left(-c'(s_1)\right).
\end{aligned}
$$

□

C8 Proof of Proposition 16

It remains to derive the profitable deviation $s_3 = (s_2^2 + s_1 s_2 - s_1^2)/s_2$. To that end, let s_3 be some quality choice with $s_2 < s_3 \leq \bar{s}$. Then with $\Delta_{ij}s := (s_j - s_i)$ and $\Delta_{ij}c := c(s_j) - c(s_i)$ the following inequalities are equivalent

$$
\begin{aligned}
\Pi_2(s_2, s_3) &> \Pi_1(s_1, s_2) && \text{(C71)}\\
\Leftrightarrow \Delta_{23}s\,\beta(s_2, s_3)^2 &> \Delta_{12}s\frac{s_2}{s_1}\alpha(s_1, s_2)^2 \\
\Leftrightarrow \Delta_{23}s\left[\frac{s_3}{3s_3 + \Delta_{23}s}\left(2\bar{\theta} - \frac{c(s_3)}{s_3} - \frac{\Delta_{23}c}{\Delta_{23}s}\right)\right]^2 &> \Delta_{12}s\frac{s_2}{s_1}\left[\frac{s_1}{3s_2 + \Delta_{12}s}\left(\frac{\Delta_{12}c}{\Delta_{12}s} - 2\frac{c(s_1)}{s_1} + \bar{\theta}\right)\right]^2 \\
\Leftrightarrow \left(\frac{\Delta_{23}s}{\Delta_{12}s}\right)\left(\frac{s_1}{s_2}\right)\left(\frac{s_3^2}{s_1^2}\right)\frac{(3s_2 + \Delta_{12}s)^2}{(3s_3 + \Delta_{23}s)^2} &> \left(\frac{\frac{c(s_2) - c(s_1)}{\Delta_{12}s} - 2\frac{c(s_1)}{s_1} + \bar{\theta}}{2\bar{\theta} - \frac{c(s_3)}{s_3} - \frac{c(s_3) - c(s_2)}{\Delta_{23}s}}\right)^2 && \text{(C72)}
\end{aligned}
$$

with α and β for $s_i < s_j$ as defined in (C36) and (C37) at the beginning of Appendix C6. Suppose we can choose s_3 in the admissible interval such that

$$
\begin{aligned}
\frac{(s_3 - s_2)}{(s_2 - s_1)} &= \frac{s_1}{s_2} && \text{(C73)}\\
\Leftrightarrow s_3 &= \frac{s_2^2 + s_1 s_2 - s_1^2}{s_2}. && \text{(C74)}
\end{aligned}
$$

With $s_1/s_2 \leq s_3/s_2$ and this particular choice of s_3 we have

$$
(s_3 - s_2) \leq \frac{s_3}{s_2}(s_2 - s_1), \tag{C75}
$$

which implies

$$
\frac{3s_2 + (s_2 - s_1)}{3s_3 + (s_3 - s_2)} \geq \frac{s_2}{s_3}. \tag{C76}
$$

Hence, for this specific choice of s_3, the LHS of (C72) reads

$$\left(\frac{s_1}{s_2}\right)\left(\frac{s_1}{s_2}\right)\left(\frac{s_3^2}{s_1^2}\right)\frac{(3s_2+\Delta_{12}s)^2}{(3s_3+\Delta_{23}s)^2} \geq \left(\frac{s_3^2}{s_2^2}\right)\left(\frac{s_2^2}{s_3^2}\right) = 1, \tag{C77}$$

while we know that the RHS of (C72) is smaller than 1 if and only if

$$2\overline{\theta} - \frac{c(s_3)}{s_3} - \frac{c(s_3)-c(s_2)}{(s_3-s_2)} > \frac{c(s_2)-c(s_1)}{(s_2-s_1)} - 2\frac{c(s_1)}{s_1} + \overline{\theta}. \tag{C78}$$

But (C78) holds, since from (B4') we know

$$\begin{aligned} 2\overline{\theta} - \frac{c(s_3)}{s_3} - c'(s_3) &> c'(s_3) - 2\frac{c(s_1)}{s_1} + \overline{\theta} \\ \Leftrightarrow 2\overline{\theta} - \frac{c(s_3)}{s_3} - \frac{\Delta_{23}c}{(s_3-s_2)} + 2\frac{c(s_1)}{s_1} - \overline{\theta} - \frac{\Delta_{12}c}{(s_2-s_1)} &> \underbrace{c'(s_3) - \frac{\Delta_{23}c}{(s_3-s_2)}}_{>0} + \underbrace{c'(s_3) - \frac{\Delta_{12}c}{(s_2-s_1)}}_{>0}. \end{aligned}$$

Hence, (C72) holds and this particular choice of s_3 is in fact a profitable deviation. When is this choice of s_3 infeasible? Suppose $s_2 < \overline{s}$. For $s_3 = s_2$ the LSH of (C73) is zero. As s_3 increases, the expression on the LHS increases. Hence, either (C73) holds for some s_3 - in which case we have found a profitable deviation - or $(\overline{s} - s_2) < s_1/s_2(s_2 - s_1)$. This deviation is infeasible if

$$(\overline{s} - s_2) < \frac{s_1}{s_2}(s_2 - s_1) = \frac{s_1}{s_2}\left(1 - \frac{s_1}{s_2}\right)s_2, \tag{C79}$$

which, since the RHS is smaller equal than $s_2/4$, holds if

$$\frac{4}{5}\overline{s} < s_2. \tag{C80}$$

□