Data Linkage between Markets: Does the Emergence of an Informed Insurer Cause Consumer Harm?

Claudia Herresthal[1], Tatiana Mayskaya[2], and Arina Nikandrova[3]

[1]University of Bonn

[2]Higher School of Economics

[3]City, University of London

October 17, 2022

Abstract

A merger of two companies that are active in seemingly unrelated markets creates data linkage: by selling a product in one market, the merged company acquires informational advantage in a competitive insurance market. In the insurance market, the informed insurer earns an economic rent through cream-skimming. Some of this rent is competed away in the product market. Overall, the data linkage makes the consumers better off.

Keywords: insurance market, asymmetric information, data linkage, digital market

JEL Codes: D82, G22, L22, L41, L86

Contents

1 Introduction

The permeating expansion of tech giants has put regulating digital markets high on the agenda of competition authorities in Europe. Thus, the European Commission is working on the Digital Markets Act (DMA). Similarly, the UK government is on the course to set up a Digital Markets Unit (DMU) with the power to regulate digital firms with substantial and entrenched market power. The central aim of the new legislation is to promote competition in digital markets for the benefit of consumers. This project examines whether promoting competition and consumer protection always go hand in hand in digital markets.

One aspect that makes digital markets so special is that online companies collect a vast amount of data about their customers. Providing a tech company with granular consumer data is a double-edged sword. On the one hand, concentrating consumers' information in the hands of a few tech giants may dangerously increase their market power. On the other hand, the companies which know more about their consumers may be able to provide a better service, thus increasing the overall efficiency of the market. Understanding the interaction between efficiency and consumer exploitation in digital markets is essential for designing effective and proportionate market interventions and is the central focus of this paper.

The paper is motivated by a recent acquisition of Fitbit by Google, which sparked heated debates on whether the acquisition would benefit the consumers. Fitbit is a manufacturer of wireless-enabled wearable technology for fitness monitoring. Prior to the acquisition of Fitbit, Google was not active in the market for wearables. Due to the lack of market overlap, under traditional merger analysis, the transaction should not raise serious competition concerns. Nevertheless, the European Commission undertook an in-depth investigation and cleared the transaction only subject to significant commitments from Google to restrict the use of the Fitbit data for advertising purposes.[1] Some commentators, however, argue that the Commission's decision did not go far enough because the Commission failed to investigate other serious theories of harm to consumers. In particular, the commentators argue that combining Fitbit health data with other data that Google harvests may give Google informational advantage in healthcare and health insurance markets — the markets where Google was not active in the

[1] A summary of the European Commission's decision can be found here, with the full case discussed here.

past.[2,3] Informational advantage may enable Google to identify low-risk individuals and offer them more attractive terms. The concern of the commentators is that such cream-skimming by Google would cause "higher prices or lack of cover for bad risks and, in the extreme case, market unraveling over time" (p.5 in Bourreau et al. (2020)).

In this paper, we study the welfare consequences of a merger between two companies operating in two different markets: an insurance market and a product market. The merger creates data linkage between the markets: the merged entity becomes an informed insurer by collecting information that is relevant in the insurance market as a by-product of operating in the product market. In the example, the merger between Google and Fitbit allows Google to acquire a customer's health data, thereby learning the customer's risk profile and becoming an informed insurer, through selling a Fitbit device to that customer. We show that in this setting, it is not a foregone conclusion that Google's superior data would or could cause consumer harm, and so the commentators' concern may be misplaced.

First, we show that in a competitive insurance market, the emergence of an informed insurer does not harm the insured. When insurance companies are not informed about the risk profile of their insured, they screen consumers by offering menus of insurance contracts and allowing each customer to self-select a contract that is tailored to his or her needs. The possibility of such screening implies that, contrary to the concern expressed in Bourreau et al. (2020), an informed insurer does not cause market unraveling. Moreover, competition guarantees that the informed insurer's offers do not make either the low- or high-risk consumers worse off.

Despite not harming the consumers, the informed insurer reaps additional profit from its superior information. This additional profit, however, does not come at the expense of consumers; it comes from the increased efficiency of the offered contracts. This conclusion echoes the opinion of Pierre Régibeau, who was the EC Chief Competition Economist at the time of Google/Fitbit merger decision. In his policy column Régibeau (2021), he argues that more information on individual health status can lead to "better diagnostics, better treatment and, even, fairer health insurance rates" and, thus, does not necessarily cause consumer harm.

[2]The data leveraging theory of harm is outlined in a series of Vox articles by various commentators and is summarized in detail in Bourreau et al. (2020).

[3]Shortly after securing the merger deal, Google launched a new insurance firm, Coefficient Insurance, which openly admits its intention to employ an "analytics-based underwriting engine" (see the Verge article).

Second, we show that the overall consumer welfare across the linked markets increases when an insurer becomes informed through selling a product in another market. The prospect of additional profit in the insurance market makes the merged entity a more aggressive competitor in the product market. As a result, prices in the product market decrease, which benefits the consumers.

Consumer benefit from data linkage is low when the product market is highly competitive, either because it has numerous competitors or because it features low product differentiation. Intuitively, consumers benefit from data linkage through lower prices in the product market. Hence, if prior to the merger, the product market is already highly competitive, data linkage has little scope to reduce prices further and so consumers have little to gain.

While the high number of competitors and the low degree of product differentiation both nullify the consumer gain from data linkage, they do so for different reasons. When faced with numerous competitors, the merged entity struggles to win significant market share in the product market. Without serving consumers, the company cannot learn their risk profiles and so cannot improve the insurance market efficiency. Hence, the consumers do not benefit from data linkage because there is no efficiency gain to distribute. In contrast, when product differentiation is low, the merged entity needs to undercut its competitors only marginally to capture a large share of the market. Large market share at the cost of a small price decrease means that the company pockets substantial efficiency gain in the insurance market without sharing any of it with the consumers.

Additional profits in the insurance market mean that the company which is active in both markets sometimes optimally sets its price below the marginal cost in the product market. Such *below-cost pricing* unambiguously benefits the consumers, provided the product market remains contestable even if competitors exit. However, if the product market is not contestable due to, for example, high barriers to entry or high technology development costs, then in the long run, the below-cost pricing may result in monopolization of the product market to the detriment of consumers.

The optimality of the below-cost pricing in informationally linked markets poses a significant challenge to the competition authorities. Traditionally, competition policy relies on so-called as-efficient-competitor test ("the AECT") to detect anti-competitive low pricing conduct.[4] According to the AECT, the conduct is deemed to be likely to cause consumer harm if the dominant firm's pricing structure involves below-cost pricing and thus could drive an equally efficient competitor from the market. The argument goes that the below-cost pricing could ever be optimal only if the short-run losses are recouped though higher prices when competitors leave the market. In our model, the optimality of the below-cost pricing does not hinge on successful market foreclosure. The possibility of recouping losses from one market through efficiency gains in another market could make low pricing conduct permanent to the benefit of consumers. This implies that the AECT is no longer a reliable indicator of consumer harm. To determine whether the low pricing conduct is indeed anti-competitive, the competition authorities must undertake a comprehensive in-the-round analysis of all the available evidence and, in particular, pay close attention to the nature of competition in the insurance market, the relative profitability of the linked markets, and the degree of product differentiation and contestability of the product market.

A frequently discussed remedy to the data-induced increase in market power is *data-sharing remedy*. Forcing the company that is active in both markets to share the information it collects with other companies in the insurance market is a double-edged sword. On the one hand, data sharing ensures that the consumers reap all the efficiency gain from data linkage. On the other hand, data sharing lowers the efficiency gain from data linkage because it lowers the merged entity's incentives to collect data on consumer risk profiles. Hence, data-sharing remedy may hurt the consumers.

As a robustness check, we consider a monopolistic insurance market and show that the lack of competition in the insurance market does not necessarily imply that data linkage makes consumers worse off. The monopolistic structure of the insurance market lowers the consumer welfare gain from data linkage by introducing consumer exploitation in the insurance market, which is absent in the setting with perfectly competitive insurance market. The monopolistic structure of the insurance market also weakens the mechanism through which the efficiency

[4]For example, the European Commission extensively references the application of the AECT in its 2009 enforcement priorities guidance on abusive exclusionary conduct by dominant firms. In the UK, the role of the AECT has recently been hotly debated in the Royal Mail v. Ofcom case at the Competition Appeal Tribunal, the Court of Appeal and finally the Supreme Court. In the US, Brooke Group v. Brown & Williamson Tobacco case established that proof of below-cost pricing and of the recoupment of predatory investment are both required to establish predatory pricing (see Bolton et al. (2000)).

gains from data linkage are passed through to the consumers in the product market. Nevertheless, the overall effect of data linkage on consumer welfare is often positive, for example, when the share of the low risk consumers is sufficiently low.

The rest of the paper is organized as follows. This section concludes with a literature review. Section 2 sets out the baseline model. Section 3 derives an equilibrium of the model. Section 4 presents our main result on the welfare consequences of data linkage and analyzes how these consequences change with competitiveness of the product market. Section 5 discusses banning below-cost pricing and data-sharing remedy. Section 6 undertakes various robustness checks of the key assumptions of the model. All proofs are relegated to the Appendix.

Related Literature

Our paper contributes to the active policy debate on designing the appropriate framework for assessing digital mergers and regulating big tech companies such as Google or Facebook. Recent reports published in the UK (Furman et al. (2019); Competition and Markets Authority (2022)), the EU (Crémer et al. (2019)), the US (Morton et al. (2019)), and Australia (Australian Competition & Consumer Commission (2019)) all point to the need for furthering our understanding of digital market ecosystems and the role of data within them.

In academic literature, there is also a perception that more research should be directed towards studying the role of data in competition. Thus, de Cornière and Taylor (2021) adopt the competition-in-utility space approach to identify conditions for data to be pro- or anti-competitive. According to de Cornière and Taylor (2021), data is pro-competitive when data increases mark-ups thus inducing firms to compete more fiercely to attract more consumers; data is anti-competitive when it enables firms to extract consumer surplus in a more efficient manner. In our baseline model with competitive insurance market, data has a pro-competitive effect, while in the model with monopolistic insurer, data has both pro- and anti-competitive effects. The pro-competitive effect of data is also reminiscent of various strands of well-established literature on aftermarkets and waterbed effect, comprehensively reviewed in Davis et al. (2012).

On a broader scale, we contribute to the vast literature that studies various aspects of information revelation and information externalities. The majority of this literature focuses on a single market (see, for example, Acemoglu et al. (2022); Ali et al. (2022); Bergemann et al. (2022); Choi et al. (2019); Hagiu and Wright (2022); Ichihashi (2020, 2021)). Our paper stands with recently emerging papers, such as Condorelli and Padilla (2021) and Argenziano and Bonatti (2021), which model data linkage between seemingly unrelated markets. In contrast to us, Condorelli and Padilla (2021) focus on entry deterrence, while Argenziano and Bonatti (2021) focus on privacy regulations.

The paper that is closest to ours is Chen et al. (2022) which models two horizontally differentiated Hotelling duopolies linked by data. In Chen et al. (2022), in both markets, the firms compete by setting prices. Serving a consumer in the data collection market enables the firm that is active in both markets to charge the consumer a personalized price for an improved product in the data application market. In contrast, in our model, consuming a product in one market does not enhance user experience in the other market, but consumption in one market reveals the consumer's risk profile that is relevant in the other market. Since we directly model data application market as an insurance market, our model is better suited for addressing the concerns of the commentators in relation to the Google/Fitbit merger. Furthermore, our modelling approach allows us to easily investigate the welfare consequences of an increased number of competitors in the data collection market.

Like in our paper, in Chen et al. (2022), data linkage unambiguously intensifies competition in the data collection market. However, in contrast to our paper, in Chen et al. (2022), the effect of data linkage on the data application market is ambiguous. Hence, despite the more competitive data collection market, in Chen et al. (2022) the overall consumer welfare may decrease as a result of data linkage. In our baseline model, cream-skimming by the informed insurer does not hurt consumers in the insurance market and so consumers are unambiguously better off. The difference in the results emerges because in the data application market, we have perfect competition with vertical product differentiation instead of imperfect competition with horizontal product differentiation. In addition, we find that data linkage in a model with a vertically-differentiated monopolistic data application market has an ambiguous welfare effect, just like in Chen et al. (2022). Hence, we view our paper as complementary to Chen et al. (2022).

2 Model

The model encompasses two markets. One market is a competitive insurance market (e.g., health insurance); the other market is a differentiated product Bertrand market (e.g., market for gadgets).

Insurance market

The insurance market is modelled as in Rothschild and Stiglitz (1976) (RS), with at least three companies.[5]

A risk-averse consumer faces uncertainty about her future income. She has an income endowment of y and, with positive probability, could suffer a loss of l. Hence, her income is either $x = y$ or $x = y - l$. Given income x, the utility of the consumer is $u(x)$, which is increasing and concave.

An insurance contract is characterized by premium p and cover q. The consumer can accept at most one contract. By accepting a contract (p, q), the consumer agrees to pay p irrespective of her future income in exchange for the payment q from the insurance company in case of the loss. If the consumer does not accept any contract, she does not pay anything but also does not receive a compensation in case of the loss.

There are two types of consumers — a high-risk type, for whom the probability of the loss is π_H, and a low-risk type, for whom the probability of the loss is π_L, with $0 < \pi_L < \pi_H < 1$. The consumer is privately informed about her type before choosing an insurance contract. Unless specified otherwise, companies are uninformed about the consumer's type with a prior belief $\gamma \in (0, 1)$ that the consumer is of low risk.

Each company offers a menu of insurance contracts. After all offers are made, the consumer chooses a contract. The payoff from a contract (p, q) is $p - \pi_i q$ if the consumer of type $i \in \{L, H\}$ accepts this contract and 0 otherwise.

[5] We require at least three companies to ensure that after one company becomes an informed insurer, the market remains competitive.

Product market

The product market is modelled as a differentiated product Bertrand market where consumers have random utility.

There are $N+1$ companies on the market, where $N \geq 1$. Each company $n = 0, 1, \ldots, N$ produces a single variety for which it sets the price t_n. The demand of company n is denoted by s_n. Company n's profit from the product market is $s_n t_n$; that is, it is assumed that all companies have marginal cost of zero in this market.

There is a unit mass of consumers. The random utility from buying variety n at price t_n is

$$V_n = V - t_n + \mu_n \sigma, \tag{1}$$

where V is identical across consumers and products and μ_n is a random taste parameter which is known to the consumer but unobserved by companies. Parameter $\sigma > 0$ is a known constant that reflects consumer's taste heterogeneity, or equivalently, is related to the degree of product differentiation. There is no outside option.[6] It is assumed that μ_n are i.i.d. and follow the double exponential distribution:

$$\Pr(\mu_n < x) = \exp\{-\exp(-x - \text{Euler's constant})\}. \tag{2}$$

Interaction between the markets

The insurance market and the product market serve the same population of consumers, in which the share of low-risk consumers is γ. Each consumer buys a product first and then chooses an insurance contract.

We study the effect of a merger between a company in the insurance market and a company in the product market. It is common knowledge that the merged entity, which we refer to as company 0, is the only company that operates in both markets. Company 0 collects data on its customers in the product market to exploit these data in the insurance market. If company 0 serves a consumer in the product market, it learns the consumer's risk type and can then

[6] In Appendix B.4, we show that our main results are robust to the introduction of the outside option.

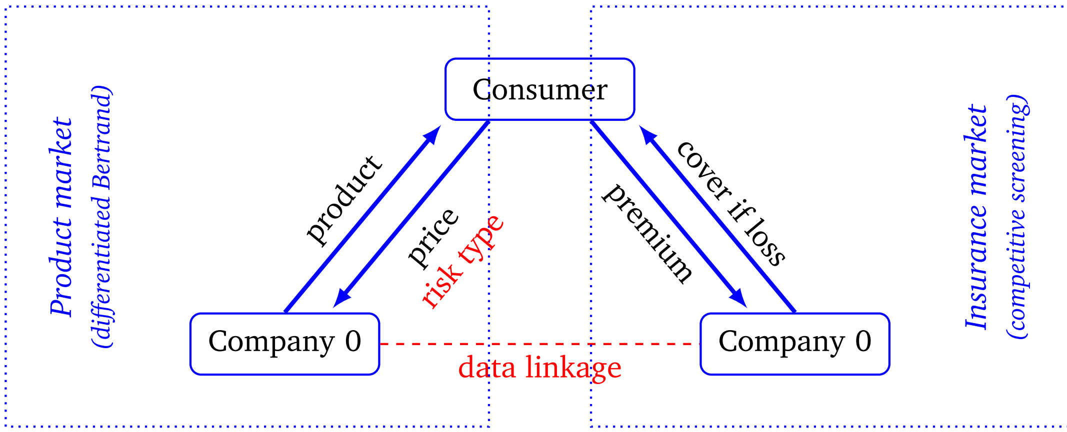


Figure 1: Sketch of the model.

offer a personalized contract in the insurance market.[7] Thus, the merger creates data linkage between the markets. Figure 1 provides a schematic depiction of the model.

Our aim is to compare market outcomes and consumer welfare with and without data linkage.

3 Equilibrium

Since company 0 is able to use information from the product market in the insurance market, we study the insurance market first. We consider two alternative assumptions: when company 0 is uninformed and when company 0 is informed about the consumer's type. Then, we move to characterizing an equilibrium in the product market. The equilibrium demand for company 0's variety in the product market defines the probability with which company 0 is informed in the insurance market.

[7]In our model, company 0 encounters the same consumer in both markets and, thus, potentially, a consumer in the product market may have an incentive to avoid company 0's variety. Alternatively, the merged entity does not encounter the same customer in both markets, but uses the information on the customers it serves in the product market to better predict the risk type of its customers in the insurance market. Under this alternative assumption, consumers would never want to hide their type by avoiding company 0's product. In our baseline model, there is no material difference between the two assumptions. However, in Section 6.2 where we consider the monopolistic insurance market, the two assumptions could lead to different results.

3.1 Insurance Market

Company 0 is uninformed

If company 0 is uninformed, then the market corresponds to a classical competitive screening model studied in RS. It is well-known that in the RS model, the separating pure strategy equilibrium — which we refer to as the RS equilibrium — does not exist when the proportion of low-risk consumers, γ, is sufficiently high. Inderst and Wambach (2001) extend the original RS model by introducing capacity constraints and search cost. In this extended model, the RS equilibrium always exists.[8] In our model, for expositional simplicity, we do not explicitly introduce capacity constraints and search cost but implicitly rely on these ideas to justify the use of the RS equilibrium for all γ.[9]

The RS equilibrium is of separating type, in which each company offers a menu of two contracts, (p_L, q_L) and (p_H, q_H), intended for the low-risk and high-risk consumer, respectively. Figure 2 illustrates the equilibrium contracts on a plane where the horizontal axis represents the consumer's income in the absence of loss and the vertical axis represents the consumer's income after suffering loss. The black point corresponds to the consumer's income endowment without any insurance.

Due to competition, each company breaks even on each contract, which means that its profit is $p_i - \pi_i q_i = 0$ for each $i \in \{L, H\}$. In Figure 2 the red straight line through the endowment corresponds to zero profit on the high-risk consumers and the green straight line is zero profit on the low-risk consumers.

The equilibrium contract for the high-risk consumer features full insurance, that is, $q_H = l$. This contract is efficient because it maximizes the expected utility of the high-risk consumer subject to the company's break-even constraint $p_H = \pi_H q_H$. In Figure 2, this contract corresponds to the red point labeled "RS".

In contrast, the contract for the low-risk consumer is inefficient. To prevent the high-risk consumer from choosing the contract intended for the low-risk consumer, the companies de-

[8] Inderst and Wambach (2001) argue that, in the presence of capacity constraints, any deviation from the RS menu of contracts is not profitable because it is more attractive to the high-risk consumers.

[9] For alternative justifications of the RS equilibrium see, for example, Dubey and Geanakoplos (2002), Bisin and Gottardi (2006), Guerrieri et al. (2010) or Azevedo and Gottlieb (2017).

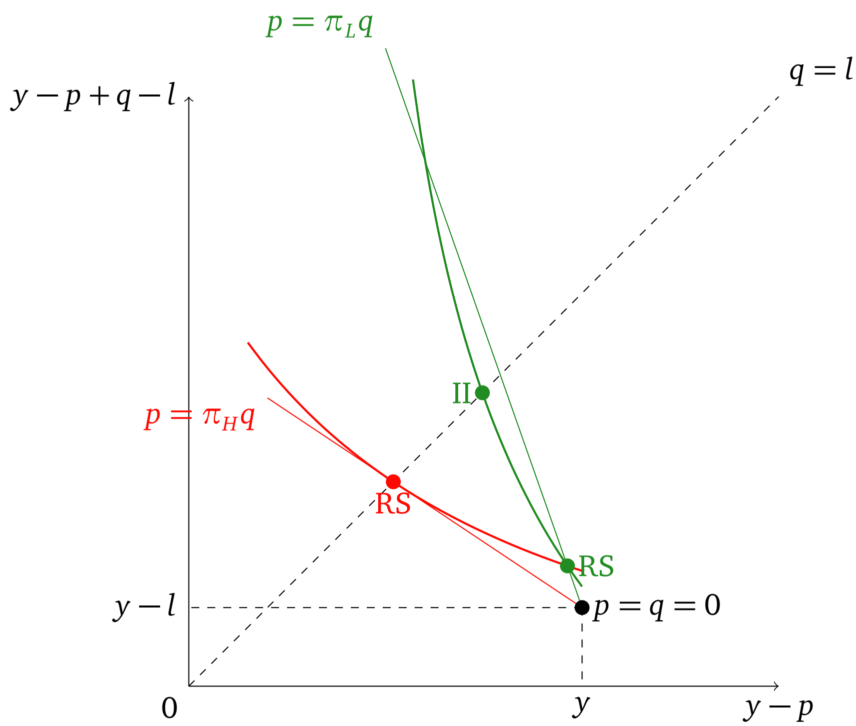


Figure 2: Equilibrium in the insurance market, drawn for the utility function $u(x) = 1 - e^{-x}$. Red and green RS points correspond to the RS equilibrium contract for the high- and low-risk consumer, respectively. Point II corresponds to the contract that the informed insurer offers to the low-risk consumer.

grade the contract intended for the low-risk type. They lower the cover and the premium just enough to satisfy the incentive compatibility constraint for the high-risk type:

$$u(y - \pi_H l) = \pi_H u(y - \pi_L q_L + q_L - l) + (1 - \pi_H) u(y - \pi_L q_L). \tag{3}$$

Equation (3) has a unique solution on $q_L \in (0, l)$, which we denote as q_L^{RS}. In Figure 2, the contract for the low-risk consumer corresponds to the green point labeled "RS", which lies on the intersection of the low-risk zero-profit line and the indifference curve for the high-risk consumers.

In sum, in equilibrium, the high-risk type gets full insurance $q_H = l$ and pays premium $p_H = \pi_H l$, while the low-risk type gets cover $q_L = q_L^{RS}$ which solves (3) and pays premium $p_L = \pi_L q_L^{RS}$.

Company 0 is informed

Suppose company 0 observes the consumer's type, thus becoming an informed insurer.

Since uninformed insurance companies break even contract-by-contract, they do not change their offers in response to the emergence of an informed insurer.

To the high-risk consumer, the informed insurer offers the same contract as an uninformed insurer because it is an efficient contract and competition on the high-risk consumers pushes insurer's profit to zero.[10]

To the low-risk consumers, the informed insurer can always offer a full insurance contract that they prefer to their RS contract, thus *cream-skimming* low-risk consumers. To maximize its profit, the informed insurer undercuts the RS contracts by an arbitrarily small amount. Formally, the informed insurer offers a full insurance contract $(p, q) = (p^I, l)$ with premium p^I that leaves the low-risk consumer just indifferent between this contract and the partial insurance contract offered to him by an uninformed company:

$$u(y - p^I) = \pi_L u(y - \pi_L q_L^{RS} + q_L^{RS} - l) + (1 - \pi_L)u(y - \pi_L q_L^{RS}), \tag{4}$$

where q_L^{RS} solves (3). The informed insurer can offer the efficient full insurance contract to the low-risk consumer because she does not need to worry about high-risk consumers buying the contract that is intended for the low-risk consumers.

In Figure 2, the informed insurer's contract for the low-risk consumer corresponds to the green point labeled "II". This point lies below the low-risk zero-profit line, which indicates that in equilibrium, the informed company will earn a positive profit on low-risk consumers. Formally, the informed insurer's profit from the low-risk type consumers is

$$\Pi = p^I - \pi_L l, \tag{5}$$

where p^I solves (4).

[10] Alternatively, we could assume that company 0 does not make an offer if it faces the high-risk consumer. Then, uninformed companies would face a population of consumers with a lower share of low-risk consumers. However, it would not affect the uninformed companies' offers because the RS contracts do not depend on the share of low-risk consumers.

3.2 Product Market

Consumers

Since when cream-skimming the informed insurer makes the consumer just indifferent between own offer and the offers of the uninformed insurers, each company 0's insurance customer individually has no incentives to conceal his type by avoiding variety 0 in the product market. Thus, the consumer's choice of a variety in the product market is not affected by the data linkage; that is, each consumer chooses the variety that yields the highest utility given (1).

Utility (1) and double exponential distribution of a random taste parameter, (2), gives rise to logit demand:[11]

$$s_n = \frac{\exp\left(-\frac{t_n}{\sigma}\right)}{\sum_{i=0}^{N} \exp\left(-\frac{t_i}{\sigma}\right)}. \tag{6}$$

Companies

All companies, except for company 0, operate only in the product market. Hence, they choose prices to maximize their profit from the product market, $s_n t_n$.

Company 0, however, gets an additional profit from the insurance market. Each low-risk consumer that company 0 serves in the product market brings company 0 an additional profit Π, defined in (5), in the insurance market. All other consumers — low-risk consumers served by other product companies as well as all high-risk consumers — bring no additional profit to company 0. If company 0 knows that it faces the high-risk consumer, it offers him the RS contract $q_H = l$, $p_H = \pi_H l$ and earns zero profit. If company 0 is uninformed, it offers the pair of RS contracts and earns zero profit irrespective of the consumer's choice. Hence, in addition to $s_0 t_0$, company 0 also obtains expected profit $\gamma \Pi s_0$ due to data linkage — that is, company 0's total profit is

$$s_0(t_0 + \gamma \Pi). \tag{7}$$

In the product market, Π is effectively an exogenous parameter because it depends only on the primitives of the insurance market, as shown in (5). In the model without data linkage,

[11] For derivation, see Anderson et al. (1992).

company 0's profit (7) does not contain the term $\gamma\Pi$. Hence, for ease of notation, we refer to the model without data linkage as the model with $\Pi = 0$.

Equilibrium

We are looking for a symmetric equilibrium in the product market. Let t^* be the price set by each company $n = 1, 2, \ldots, N$; let t_0^* be the price set by company 0. Proposition 1 solves for the equilibrium prices and market shares.

Proposition 1. *In equilibrium, the prices are*

$$t_0^* = \frac{\sigma}{1 - s_0^*} - \gamma\Pi \tag{8}$$

and

$$t^* = \frac{\sigma}{1 - s^*}, \tag{9}$$

where

$$s^* = \frac{1 - s_0^*}{N} \tag{10}$$

is the demand for each variety $n = 1, 2, \ldots, N$, and s_0^* is the demand for variety 0, implicitly defined in

$$\frac{(N+1)s_0^* - 1}{(1 - s_0^*)(N - 1 + s_0^*)} - \ln \frac{1 - s_0^*}{N s_0^*} = \frac{\gamma\Pi}{\sigma}. \tag{11}$$

The solution $s_0^* \in (0, 1)$ to (11) always exists and is unique. Moreover, $s_0^* = s^* = 1/(N + 1)$ if $\Pi = 0$ and $s_0^* > 1/(N + 1) > s^*$ if $\Pi > 0$.

Proposition 1 illuminates the effect of data linkage $\Pi > 0$ on the product market equilibrium. Without data linkage ($\Pi = 0$), the prices and the market shares of all companies are the same: $t_0^* = t^*$, $s_0^* = s^* = 1/(N + 1)$. Data linkage incentivizes company 0 to set a price which is different from other companies. Profit expression (7) shows that, in the product market, $\gamma\Pi$ plays the role of a per-consumer subsidy to company 0. Intuitively, this subsidy makes company 0 compete more aggressively for customers and lower its price t_0^*, as shown in (8). Lower price results in an increase of its market share s_0^*. Formally, (11), which can be rewritten as

$$\frac{1}{1-s_0^*}+\ln s_0^*-\frac{1}{1-s^*}-\ln s^*=\frac{\gamma\Pi}{\sigma}, \tag{12}$$

shows that the presence of data linkage Π introduces a wedge between s_0^* and s^*.

In response to more aggressive pricing by company 0, other companies also lower their prices t^*, as formally stated in Lemma 1.

Lemma 1. *Data linkage lowers prices t_0^* and t^*.*

More aggressive competition in the presence of data linkage is a key force driving our results.

4 Results

In this section, we discuss our main findings. Table A.1 in Appendix A.3 comprehensively summarizes all the results of our model.

Welfare implications of data linkage

Our main result (stated in Theorem 1) is that consumers are better off in the presence of data linkage.

In the insurance market, the utility of either type of consumers is not affected by the presence of the informed insurer. The high-risk consumers get the same full insurance contract as in RS model. The low-risk consumers get full insurance instead of partial insurance contract, but the premium for the former is such that he is indifferent between the two contracts. Hence, data linkage affects the consumer welfare only through the product market.

In the product market, the consumer welfare is defined as the expected utility from the best offered product; that is, $W=\mathbb{E}\left[\max_n V_n\right]$. Proposition 2 derives this welfare.

Proposition 2. *In the product market, in equilibrium, the consumer welfare is*

$$W=V+\sigma\ln\left\{\exp\left(-\frac{t_0^*}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)\right\}. \tag{13}$$

As expected, the consumer welfare in the product market decreases with prices t^* and t_0^*. Since data linkage intensifies competition and, by Lemma 1, reduces the prices in the product market, it benefits the consumers.[12]

Theorem 1. *The consumers benefit from data linkage.*

The benefit of data linkage comes from an increase in efficiency of the insurance market. The knowledge of a consumer type allows company 0 to offer a more efficient insurance contract to the low-risk consumers. Hence, the total efficiency gain in the insurance market is $s_0^*\gamma\Pi$.[13] By Theorem 1, some of this efficiency gain passes to consumers through the product market. Consumers, however, do not reap the efficiency gain in its entirety, as follows from Proposition 3.

Proposition 3. *Company 0 gains from data linkage. Other companies in the insurance market are not affected by data linkage. Other companies in the product market lose from data linkage. However, jointly, company 0 and all other companies in the product market gain from data linkage.*

In the insurance market, other companies get none of the efficiency gain because they break even on each contract independently of whether company 0 is informed. In the product market, all companies are affected by the data linkage. Their joint profit increases, which indicates that collectively, companies which are present in the product market enjoy some of the insurance market efficiency gain. However, it turns out that company 0 is the only company that benefits from data linkage. In addition to pocketing some efficiency gain from the insurance market, company 0 also raids some of the competitors' profits in the product market.

Proposition 4 shows that all the effects of data linkage intensify when per-consumer efficiency gain Π in the insurance market increases. In particular, higher Π incentivizes company 0 to become a more aggressive competitor and win a larger market share. At the limit as Π tends to infinity, company 0's market share s_0^* tends to 1, which means that company 0 captures the entire product market.However, apparent monopolization of the product market does not

[12]The conclusion of Theorem 1 contrasts Chen et al. (2022) where lower prices in the data collection market may come at a cost of consumer surplus loss in the data application market.

[13]The difference in prices t_0^* and t^* distorts consumer choice of varieties, thus lowering the available economic surplus in the product market. Hence, the total efficiency gain from data linkage across both markets is lower than $s_0^*\gamma\Pi$.

harm the consumers because other product companies continue to exert a competitive constraint: company 0 sets its price very low to prevent competitors from winning customers.[14] This mechanism assures that the consumer gain from data linkage increases in Π.

Proposition 4. *In the product market, prices t_0^* and t^* decrease with Π; company 0's market share s_0^* increases with Π, while other companies market shares s^* decrease with Π; at the limit $\Pi \to +\infty$, $s_0^* \to 1$ and $s^* \to 0$. The gains of consumers and of company 0 as well as the loss of other companies in the product market, which data linkage induces, increase with Π.*

The role of competitiveness of the product market

The product market becomes more competitive as the number of varieties N increases. The effect of an increased competition is standard: prices, market shares and profits decrease and consumer welfare increases in N.[15] Notably, the direction of the effect of N on welfare is the same with and without data linkage, which means that the relationship between N and the welfare change due to data linkage *a priori* is not clear and depends on whether data linkage weakens the effect of higher N. Proposition 5 shows, however, that there is no ambiguity and the welfare change due to data linkage is lower in more competitive markets.

Proposition 5. *The welfare changes due to data linkage — that is, the consumer welfare gain, company 0's profit gain and the loss in the joint profit of other companies in the product market — all decrease in N and go to 0 as $N \to +\infty$.*

According to Proposition 5, an increase in competition in the product market lessens and eventually nullifies the welfare effect of data linkage. The intuition is based on the observation that, the market share of company 0 decreases in N, even in the presence of data linkage. As the market share of company 0 decreases, fewer consumers reveal their risk type to company 0 and, thus, the efficiency gain in the insurance market declines. In other words, competition in the product market limits company 0's ability to collect data, thus dissipating the insurance market efficiency gains.

[14] For further discussion of the monopolization concern, see Section 5.1.

[15] For the proof, see Appendix A.3.

Alternatively, Proposition 5 can be viewed through the lens of similarity between higher N and introducing data linkage: both lead companies in the product market to compete more aggressively. Hence, higher N reduces the scope for data linkage to lower prices.

Another parameter that affects market competitiveness is the degree of taste heterogeneity, or product differentiation, σ. As σ increases, consumers in the product market become less price sensitive, which increases the market power of each company in the product market. At the limit $\sigma = +\infty$, the consumer choice is random and does not depend on price at all.

If $\sigma = 0$, then each consumer views all varieties as equivalent and, according to utility specification in (1), chooses the cheapest variety. Hence, the product market behaves as the homogeneous-product Bertrand market. Without data linkage, prices of all companies tend to marginal cost, which is equal to zero by assumption, while market shares are equal to $1/(N+1)$. With data linkage, all prices also tend to zero, but company 0 captures the entire market, as stated in Proposition 6. Intuitively, additional profit due to data linkage makes it profitable for company 0 to undercut its competitors by an arbitrary small amount, thus reaping all the insurance market efficiency gain without passing any of it onto consumers.

Proposition 6. *Suppose $\sigma = 0$. Then, with data linkage, $t_0^* = t^* = 0$ and $s_0^* = 1$. Moreover, consumers do not gain from data linkage.*

The two extremes, $\sigma = 0$ and $N = +\infty$, both correspond to perfect competition in the product market. However, they have dramatically different consequences for the insurance market efficiency gain and company 0's gain from data linkage. If $\sigma = 0$, then company 0's market share is 1 and so the efficiency gain, $s_0^*\gamma\Pi$, is at its maximum, $\gamma\Pi$.[16] If $N = +\infty$, company 0's market share is 0 and so there is no efficiency gain. In both cases, consumers do not gain from data linkage, but for different reasons. If $\sigma = 0$, company 0 pockets all the efficiency gain, while if $N = +\infty$, there is no efficiency gain to distribute.

As σ increases, company 0 faces a non-trivial trade-off. On the one hand, like all the other companies in the product market, company 0 can exploit the reduced price sensitivity of consumers by raising its price. On the other hand, as each consumer now has stronger preference

[16] If $\sigma = 0$, all varieties provide exactly the same utility and, therefore, a change in consumer choice of varieties due to data linkage does not induce welfare loss in the product market. Thus, the total efficiency gain from data linkage across both markets is equal to the efficiency gain in the insurance market.

for a particular variety, company 0 has to lower its price more aggressively to increase its market share and, with it, the knowledge of consumer risk profiles. According to Proposition 7, company 0 resolves this trade-off by lowering price at low values of σ, when a small decrease in price expands its market share dramatically, and increasing price at high values of σ, when an increase in price does not cause a significant decrease in the market share. Despite non-monotonicity of company 0's price, its market share monotonically decreases with σ. At the limit $\sigma \to +\infty$, consumers become so price insensitive that company 0 is unable to lure additional consumers by lowering its price, thus behaving as all other companies and capturing $1/(N+1)$ of all consumers.

Proposition 7. *As σ increases, with data linkage, s_0^* decreases from 1 to $1/(N+1)$, while t_0^* first decreases and then increases.*

One takeaway message from this section is that if the product market is perfectly competitive — either because $N = +\infty$ or because $\sigma = 0$ — the consumer gain from data linkage is 0. Moreover, according to Proposition 5, the consumer gain monotonically decreases with the number of competitors in the product market; that is, when the product market is more competitive, consumers are expected to gain less from the data linkage. However, this simple message does not carry over to the degree of taste heterogeneity — the relationship between the consumer gain and σ is not monotone, as we now explain.

The increase in σ has an additional effect that is unrelated to market competitiveness. For fixed prices, higher σ makes more extreme taste realizations more likely, thus increasing consumer welfare. This additional effect causes non-monotonicity in the relationship between the consumer gain from data linkage and σ. In Appendix A.3, we show that this relationship is hump-shaped.

5 Policy Implications

5.1 The Monopolization Concern

Policy-makers and commentators are increasingly concerned that data linkage between markets may lead to the emergence of dominant companies with entrenched market power and that this may harm the consumers.[17]

Within our model, data linkage may indeed lead to an increase in the market share of company 0 in each market. In the insurance market, company 0 cream-skims all low-risk consumers whom it serves in the product market. The prospect of reaping additional profit in the insurance market by utilizing data from its consumer base in the product market incentivizes company 0 to increase its presence in the product market (see Proposition 4). Nevertheless, according to our model, despite the increased presence of company 0 in each market, data linkage benefits the consumers.

One aspect that we do not model, however, is the possibility that the company with informational advantage could induce its competitors to exit the market. In this section, we discuss this possibility in the context of our model.

In the insurance market, informational advantage of company 0 does not induce other companies to exit the market. While the informed insurer tempts away some of the low-risk consumers, the uninformed insurers keep substantial market share by serving the remaining consumers without making losses. Hence, there is no reason for the uninformed insurers to exit the market. Our conclusions rely on the assumption that companies compete in menus of contracts, choosing price-quality bundles to offer. Had the companies competed only in prices, the uninformed companies would not be able to screen their insured, and so company 0 would be able to use its informational advantage to push other companies out of the market, as in Chen et al. (2022).

In the product market, company 0 captures the entire market if per-consumer efficiency gain Π in the insurance market goes to infinity (see Propositions 4) or product differentiation in the product market σ goes to 0 (see Proposition 6). That is, when the gains in the insurance market are particularly large or when consumers view all varieties as equivalent, company 0 may be able to foreclose the sales of its competitors. In our model, such foreclosure does not cause consumer harm because other companies, however small, continue to discipline company 0's pricing behavior. The matters are different when foreclosure forces competitors to exit and thereupon the market ceases to be contestable. In this case, the price in the market

[17] For example, the UK Government expressed such concerns in the consultation on a new pro-competition regime for digital markets (see para 15).

may rise to the monopoly level, harming consumers as per the traditional foreclosure concern. Whether consumer harm arises, of course, depends on the barriers to entry into the market — when barriers are low, the exit of competitors is of no concern because the mere threat of competition suffices to keep prices low.

In the context of Google/Fitbit merger, our result imply that the merger may indeed result in Fitbit dominance in the market for wearables.[18] Commentators agree that, while the market for wearables is rapidly expanding, the available gains in the healthcare and insurance markets are so large that they dwarf the device profits.[19] Hence, the case of $\Pi \to +\infty$ is particularly relevant for Google/Fitbit merger and so, according to our model, the Fitbit market share could be expected to grow rapidly following the merger. This prediction can be tested when data becomes available.[20]

As a practical matter, our theoretical findings suggest that to determine whether and how data linkage between markets is capable of causing consumer harm, competition authorities should pay close attention to a number of factors. First, consumer harm in the insurance market depends on whether insurers compete in prices or in menus of contracts. Second, the relative sizes of the available gains in the linked markets as well as the degree of product differentiation in the product market affect to what extent data linkage increases the market presence of company 0 in the product market. Finally, whether an increased dominance of company 0 could lead to consumer harm depends on the barriers to entry into the product market.

5.2 Banning Below-Cost Pricing

Several remedies have been suggested to mitigate the monopolization concern discussed in Section 5.1. One of such remedies is prohibition of below-cost pricing.

The policy of prohibiting below-cost pricing is familiar from the traditional competition policy frameworks. For example, in its 2009 enforcement priorities guidance on abusive ex-

[18]Both Bourreau et al. (2020) and Chen et al. (2022) warn against the possibility of the product market monopolization in the context of Google/Fitbit merger.

[19]See, for example, Bourreau et al. (2020) or Forbes column.

[20]While it is too early to judge, the nascent evidence provides no support for the testable implication of our model. During the first year after the Google/Fitbit merger, from the first quarter of 2021 to the first quarter of 2022, Fitbit market share dropped from 4.1% to 2.7% (see the Counterpoint article).

clusionary conduct by dominant firms, the European Commission, while acknowledging that consumers benefit from low prices, deems pricing below own marginal cost anti-competitive.[21] The concern of competition authorities hinges on its focus on a single market, where below-cost pricing is associated with negative profit and so is necessarily short-lived. A company has an incentive to lower its price below marginal cost only if a period of below-cost pricing forces the competitors out of the market, subsequently allowing the company to raise its price above the competitive level for a prolonged period. Since the short-run losses must be compensated by the increased profit in the long run, in a single market, the below-cost pricing necessarily has anti-competitive intent and, while benefiting consumers in the short run, is detrimental to consumers in the long run.

In our model, to expand its market share, company 0 may optimally sell the product at a below-cost price. In particular, the marginal cost in the product market is zero and, in equilibrium, company 0 sets negative price when, for example, σ is positive but sufficiently close to 0 (see Propositions 6 and 7).[22]

Our model unambiguously predicts that banning below-cost pricing would not benefit the consumers. In contrast to a single-market reasoning of traditional competition policy, in our model with two linked markets, the below-cost pricing is profitable for company 0 even in the short run because the company can recoup the product market losses through efficiency gains in the insurance market. Hence, below-cost pricing strategy may be permanent and banning it would only degrade the channel through which company 0 passes to consumers the efficiency gains from the insurance market.

Despite the unambiguous prediction of our model, competition authorities may take a more cautious stance. As discussed in Section 5.1, even without an explicit intention to do so, the below-cost pricing of company 0 may force the competitors exit the product market in the long run. That is, the additional profit from the insurance market means that company 0 can squeeze rivals out of the market without profit sacrifice in the short run. Since company 0's

[21] Para 23 states: "Vigorous price competition is generally beneficial to consumers. With a view to preventing anti-competitive foreclosure, the Commission will normally only intervene where the conduct concerned has already been or is capable of hampering competition from competitors which are considered to be as efficient as the dominant undertaking."

[22] In Appendix A.3, we show that company 0 also sets negative price for sufficiently high Π (see column 3 in Table A.1) and for sufficiently high N when $\gamma\Pi > \sigma$ (see column 5 in Table A.1).

profit is decreasing in N, forcing competitors to exit may well be profitable for company 0.

Overall, data linkage has two effects. On the one hand, by eliminating the short run cost of the below-cost pricing, the data linkage with the insurance market aggravates anti-competitive foreclosure and monopolization concerns in the product market. On the other hand, the possibility of recouping the product market losses in another market may make the company's low pricing conduct permanent in a contestable market. Hence, competition authorities have to carefully weigh the consumer benefit from low prices against the risk that the price decrease may not be permanent.

5.3 Data-Sharing Remedy

Another remedy that is frequently discussed in policy circles is data sharing remedy. Applying this remedy to our model, assume that company 0 is forced to share the information it obtains in the product market with other companies in the insurance market.

Competition in the insurance market ensures that, after sharing information, company 0 earns zero profit in the insurance market. Hence, it has no reason to compete more aggressively than other companies in the product market and so, in a symmetric equilibrium, the prices do not change as a result of data linkage.

With data sharing policy, the overall effect of data linkage on consumer welfare is positive and comes exclusively from the insurance market. Indeed, in the product market, data linkage does not change the consumer welfare because it does not affect prices. However, in the insurance market, data linkage increases the consumer welfare because the low-risk consumers served by company 0 in the product market get their first-best contract.

In contrast, without data sharing, the consumer welfare gain from data linkage comes exclusively from the product market. Hence, the data sharing remedy lowers the consumer welfare in the product market but increases it in the insurance market. Whether the total effect on consumer welfare is positive is ambiguous. On the one hand, data sharing ensures that the consumers reap all the efficiency gain from data linkage. On the other hand, data sharing lowers the total efficiency gain from data linkage because it lowers company 0's incentives to collect data on consumer risk profiles by competing aggressively in the product market.[23]

[23] In a different context, Condorelli and Padilla (2021) show that some forms of data sharing remedies may

In Appendix B.1, we show that whether the data-sharing remedy benefits the consumers depends on the taste heterogeneity in the product market. In particular, under an additional normalization assumption, there exists a threshold taste heterogeneity such that the data sharing remedy benefits consumers if and only if σ is below this threshold. Intuitively, when consumers view all varieties as equivalent and there is no data sharing, consumers do not gain from data linkage (see Proposition 6). In contrast, forcing company 0 to share its data allows consumers to reap the efficiency gain through the insurance market. Hence, data sharing has a positive effect on consumer welfare when taste heterogeneity is low.

6 Discussion and Extensions

6.1 Two Potentially Informed Insurers

In our baseline model, company 0 is the only company that operates in both markets. However, nowadays, there are several big tech competitors controlling extensive product ecosystems. For example, like Google, Apple can also enter the health insurance market as an informed insurer because Apple is active in the smartwatch and fitness monitoring device market. We argue that in our model, allowing another company, say company 1, to operate in both markets strengthens the positive effect of data linkage on consumer welfare.

In the insurance market, company 0 and company 1 do not directly compete with each other. Because each consumer buys only one item in the product market, the sets of consumers who revealed their risk type to company 0 and to company 1 are non-overlapping. Constrained by their uninformed competitors, both companies offer to the low-risk consumers a full insurance contract $(p,q)=(p^I,l)$ with premium p^I defined in (4). Thus, from each low-risk consumer served in the product market, each company makes the same profit Π as in our baseline model. At the same time, as in our baseline model, the consumers are indifferent to the presence of informed insurers.

In the product market, there are now two companies receiving a per-consumer subsidy in the form of insurance market profit $\gamma\Pi$. Intuitively, having two "subsidized" competitors

backfire and harm the consumers. As in our model, they show that the necessity to share the acquired data lowers a company's incentive to increase its market presence through intensifying competition.

intensifies competition further, which increases consumer gain from data linkage. For example, in our baseline model, when $\sigma = 0$, company 0 does not pass on to consumers any of the efficiency gain from the insurance market because competitors in the product market simply cannot match company 0's ability to lower prices. With two subsidized companies, however, the Bertrand competition ensures that the profits from the insurance market are completely competed away and consumers obtain the insurance market efficiency gain in its entirety.

6.2 Monopolistic Insurance Market

Our model provides a framework for exploring the efficiency vs consumer exploitation trade-off which is frequently discussed in relation to the extensive data collection by tech giants. Information about the risk type of a consumer allows the insurer to offer a more efficient contract to that consumer. At the same time, this information may also allow the insurer to exploit the consumer by extracting more rents. In our baseline model, the competitiveness of the insurance market prevents consumer exploitation, rendering the trade-off trivial. In this section, we re-introduce the trade-off by relaxing the assumption of perfect competition in the insurance market. In particular, we look at an extreme case when company 0 is the only company in the insurance market — that is, the insurance market is monopolistic. All technical derivations are deferred to Appendix B.2.

When company 0 is informed about the risk-type of a consumer, its monopoly power allows the company to extract all surplus and offer a contract that makes the consumer indifferent to buying the insurance. When company 0 does know the consumer risk type, it might have to leave the high-risk consumer some rent to prevent him from choosing the contract that is designed for low-risk consumers. Hence, data linkage deprives the high-risk consumer of this rent, which introduces the consumer exploitation element of the efficiency vs exploitation trade-off.

While information on the risk types might allow the monopolistic insurer to exploit the high-risk consumers, the low-risk consumers remain indifferent to buying the insurance irrespective of whether company 0 is informed. Instead, the low-risk consumers are the source of the efficiency gain — as in our baseline model, the informed monopolist offers more efficient contracts to the low-risk consumers and pockets all the efficiency gain.

As in our baseline model, part of the efficiency gain from data linkage may pass to consumers through lower prices in the product market. However, the monopolistic structure of the insurance market introduces two differences. On the one hand, not only low risk consumers but also high risk consumers, whom company 0 serves in the product market, may generate additional profit for company 0 in the insurance market. Hence, relative to the baseline model, now company 0 has higher incentives to lower prices to attract consumers in the product market; that is, the monopolistic insurance market strengthens the pro-competitive effect of data linkage on the product market. On the other hand, to prevent the informed monopolistic insurer from exploiting them, the high-risk consumers may have an incentive to conceal their type by avoiding company 0's variety in the product market. In effect, high-risk consumers become more loyal to other companies, which softens competition in the product market and may result in higher prices. In other words, the consumer exploitation in the insurance market, induced by data linkage, brings into play an anti-competitive effect in the product market.

Across both markets, the overall welfare consequences for consumers from the data linkage depend on the share of low-risk consumers. If the share of low-risk consumers is sufficiently low, data linkage does not lead to exploitation of high-risk consumers in the insurance market. Indeed, because there are so few low-risk consumers, the uninformed insurer does not serve them, and so there is no reason to leave any rent to the high-risk consumers. Hence, in this case, data linkage does not affect the insurance contract offered to the high-risk consumers and, thus, all the results from our baseline model remain valid. In particular, both consumer types benefit from data linkage.

If the share of low-risk consumers is high, data linkage does introduce high-risk consumer exploitation in the insurance market and, hence, the overall effect from the data linkage on consumers welfare is type-dependent and ambiguous. The high-risk consumer exploitation directly reduces the welfare of these consumers in the insurance market and indirectly reduces the welfare of all consumers in the product market through the anti-competitive effect. Despite the consumer exploitation and the ensuing anti-competitive effect on the product market, for a large space of parameter values, the overall effect of data linkage on the average consumer welfare remains positive.

In Appendix B.2, we prove that, on average, consumers benefit from data linkage when the product differentiation σ is sufficiently high. Intuitively, high product differentiation discourages the high-risk consumers from concealing their type from company 0, which weakens the anti-competitive effect in the product market.

We also prove that the low-risk consumers benefit from data linkage when the product market is sufficiently competitive, as measured by the number of companies N. Intuitively, when the high-risk consumers, who avoid buying from company 0, are thinly spread among many competitors, the anti-competitive effect is weak.

If the share of low-risk consumers γ is relatively high, the high-risk consumers and sometimes even the low-risk consumers are worse off from data linkage. When γ is relatively high, the uncertainty about the consumers' risk types is low and, thus, company 0 has little to gain from additional information that attracting consumers in the product market brings. Hence, in the product market, the pro-competitive effect of data-linkage is weak. At the same time, the high-risk consumers have a lot to lose from revealing their type to company 0 because when hiding among numerous low-risk consumers, the high-risk consumers get high information rent. Hence, in the product market, the anti-competitive effect of data linkage is strong.

6.3 Cross-Subsidy Equilibrium

In this section, we discuss the robustness of our results to an alternative outcome in the insurance market. In Section 3.1, following one stream of literature, we use the RS equilibrium for all γ. Another stream of literature justifies a different outcome in the RS model — the Miyazaki-Wilson contracts, which may involve cross-subsidization.[24] Formally, Miyazaki-Wilson contracts maximize the expected utility of the low-risk consumer subject to three constraints: the usual incentive compatibility constraint for the high-risk consumers, the constraint ensuring that the insurance companies break even on average across both contracts, for the low- and for the high-risk consumers, and the constraint that prohibits cross-subsidization from the high- to the low-risk consumers. For sufficiently low γ, Miyazaki-Wilson contracts co-

[24] Netzer and Scheuer (2014) derive the Miyazaki-Wilson contracts as a unique subgame perfect equilibrium outcome in an extensive form game where, at small cost, insurance companies can withdraw from the market after observing the initial contract offers. See also Bisin and Gottardi (2006).

incide with the RS contracts. For sufficiently high γ, Miyazaki-Wilson contracts involve cross-subsidization from the low- to the high-risk consumers. Hence, we refer to the Miyazaki-Wilson contracts as the cross-subsidy equilibrium; we derive this equilibrium in Appendix B.3.

From consumers' perspective, the cross-subsidy equilibrium is superior to the RS equilibrium. A positive subsidy from the low-risk consumers directly benefits the high-risk consumers, making them better off. A positive subsidy also benefits the low risk consumers, albeit indirectly. Intuitively, the cross-subsidy relaxes the incentive compatibility constraint of the high-risk consumers making them less willing to pretend to be of low risk. The relaxation of the constraint allows the companies to design better offers to the low-risk consumers, which benefits the consumers.

In contrast to our baseline model, data linkage reduces welfare of consumers in the insurance market. As in our baseline model, competition ensures that the informed insurer offers contracts which, in utility terms, are as good as the contract offered by the uninformed competitors. Thus, individually, each consumer is indifferent to whether an informed insurer serves him. Nevertheless, collectively, consumers are worse off because the informed insurer tempts away only the low-risk consumers and, thus, the capacity of the uniformed insurers to cross-subsidize the high-risk consumers goes down, which, in turn, reduces consumer welfare in the insurance market.[25]

A priori it is not clear whether data linkage increases the overall consumers across both markets. On the one hand, data linkage makes insurance market contracts worse for consumers. On the other hand, as in our baseline model, data linkage provides incentives for company 0 to compete aggressively in the product market, which benefits consumers. It is not immediately clear which of the two forces takes an upper hand. In Appendix B.3, we show that data linkage may benefit both types of consumers even if the equilibrium in the insurance market involves cross-subsidization.

[25]Noteworthy, the loss of consumer welfare in the insurance market is at most the welfare difference between the cross-subsidy and RS equilibrium.

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar, "Too much data: prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 2022. 8

Ali, S. Nageeb, Gregory Lewis, and Shoshana Vasserman, "Voluntary disclosure and personalized pricing," *Review of Economic Studies*, 7 2022. 8

Anderson, Simon P., Andre de Palma, and Jacques-Francois Thisse, *Discrete choice theory of product differentiation*, Cambridge, MA: MIT Press, 1992. 15

Argenziano, Rossella and Alessandro Bonatti, "Data linkages and privacy regulation," 2021. 8

Australian Competition & Consumer Commission, "Digital platforms inquiry," Technical Report July 2019. 7

Azevedo, Eduardo M. and Daniel Gottlieb, "Perfect competition in markets with adverse selection," *Econometrica*, 2017, *85* (1), 67–105. 12

Bergemann, Dirk, Alessandro Bonatti, and Tan Gan, "The economics of social data," *RAND Journal of Economics*, 6 2022, *53* (2), 263–296. 8

Bisin, Alberto and Piero Gottardi, "Efficient competitive equilibria with adverse selection," *Journal of Political Economy*, 6 2006, *114* (3), 485–516. 12, 29

Bolton, Patrick, Joseph F. Brodley, and Michael H. Riordan, "Predatory pricing: strategic theory and legal policy," *Georgetown Law Journal*, 2000, *88* (8), 2239–2330. 6

Bourreau, Marc, Cristina Caffarra, Zhijun Chen, Chongwoo Choe, Gregory S Crawford, Tomaso Duso, Christos Genakos, Paul Heidhues, Martin Peitz, Thomas Rønde, Monika Schnitzer, Nicolas Schutz, Michelle Sovinsky, Giancarlo Spagnolo, Otto Toivanen, Tommaso Valletti, and Thibaud Vergé, "Google/Fitbit will monetise health data and harm consumers," *CEPR Policy Insight*, 2020, *107* (September). 4, 23

Chen, Zhijun, Chongwoo Choe, Jiajia Cong, and Noriaki Matsushima, "Data-driven mergers and personalization," *RAND Journal of Economics*, 3 2022, *53* (1), 3–31. 8, 18, 22, 23

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim, "Privacy and personal data collection with information externalities," *Journal of Public Economics*, 5 2019, *173*, 113–124. 8

Competition and Markets Authority, "Mobile ecosystems," Technical Report June 2022. 7

Condorelli, Daniele and Jorge Padilla, "Data-driven envelopment with privacy-policy tying," 2021. 8, 25

Crémer, Jacques, Yves-Alexandre De Montjoye, and Heike Schweitzer, "Competition policy for the digital era," Technical Report 2019. 7

Davis, Peter, Lorenzo Coppi, Philip Kalmus, and Compass Lexecon, "The economics of secondary product markets," 2012. 7

de Cornière, Alexandre and Greg Taylor, "Data and competition: a simple framework, with applications to mergers and market structure," 2021. 7

Dubey, P. and J. Geanakoplos, "Competitive pooling: Rothschild-Stiglitz reconsidered," *Quarterly Journal of Economics*, 11 2002, *117* (4), 1529–1570. 12

Furman, Jason, Diane Coyle, Amelia Fletcher, Derek McAuley, and Philip Marsden, "Unlocking digital competition: report of the digital competition expert panel," Technical Report March 2019. 7

Guerrieri, Veronica, Robert Shimer, and Randall Wright, "Adverse selection in competitive search equilibrium," *Econometrica*, 11 2010, *78* (6), 1823–1862. 12

Hagiu, Andrei and Julian Wright, "Data-enabled learning, network effects and competitive advantage," 2022. 8

Ichihashi, Shota, "Online privacy and information disclosure by consumers," *American Economic Review*, 2 2020, *110* (2), 569–595. 8

_ , "The economics of data externalities," *Journal of Economic Theory*, 9 2021, *196*, 105316. 8

Inderst, Roman and Achim Wambach, "Competitive insurance markets under adverse selection and capacity constraints," *European Economic Review*, 12 2001, *45* (10), 1981–1992. 12

Morton, F. Scott, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, Douglas Melamed, and Jamie Morgenstern, "Report of the committee for the study of digital platforms, market structure and antitrust committee," 2019. 7

Netzer, Nick and Florian Scheuer, "A game theoretic foundation of competitive equilibria with adverse selection," *International Economic Review*, 5 2014, *55* (2), 399–422. 29, 59

Régibeau, Pierre, "Why I agree with the Google-Fitbit decision," *Vox EU & CERP*, 2021. 4

Rothschild, Michael and Joseph Stiglitz, "Equilibrium in competitive insurance markets: an essay on the economics of imperfect information," *Quarterly Journal of Economics*, 11 1976, *90* (4), 629–649. 9

Stiglitz, Joseph E., "Monopoly, non-linear pricing and imperfect information: the insurance market," *Review of Economic Studies*, 10 1977, *44* (3), 407. 44

Appendix A Technical Results

A.1 Proof of Proposition 1

Company 0 chooses price t_0 to maximize

$$\max_{t_0} s_0(t_0+\gamma\Pi) = \frac{\exp\left(-\frac{t_0}{\sigma}\right)}{\exp\left(-\frac{t_0}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)}(t_0+\gamma\Pi) \tag{A.1}$$

$$\text{FOC}: \ \exp\left(-\frac{t_0}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)\left(1-\frac{t_0+\gamma\Pi}{\sigma}\right)=0 \tag{A.2}$$

SOC always holds, so that any solution t_0 to (A.2) is a local maximum.

Company $n \geq 1$ maximizes

$$\max_{t_n} s_n t_n = \frac{\exp\left(-\frac{t_n}{\sigma}\right)}{\exp\left(-\frac{t_n}{\sigma}\right)+\exp\left(-\frac{t_0^*}{\sigma}\right)+(N-1)\exp\left(-\frac{t^*}{\sigma}\right)} t_n \tag{A.3}$$

$$\text{FOC}: \ \exp\left(-\frac{t_n}{\sigma}\right)+\left(\exp\left(-\frac{t_0^*}{\sigma}\right)+(N-1)\exp\left(-\frac{t^*}{\sigma}\right)\right)\left(1-\frac{t_n}{\sigma}\right)=0 \tag{A.4}$$

SOC always holds, so that any solution t_n to (A.4) is a local maximum.

Denote

$$s^* = \frac{\exp\left(-\frac{t^*}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)}, \quad s_0^* = \frac{\exp\left(-\frac{t_0^*}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)} \tag{A.5}$$

the equilibrium demand for companies $n=1,\ldots,N$ and company 0, respectively. Then (A.2) implies (8) and (A.4) implies (9). Definition (A.5) implies that $Ns^*+s_0^*=1$, which gives (10). Using the equation for s_0^* in (A.5) to solve for t_0^* yields

$$t_0^* = t^* + \sigma \ln\frac{1-s_0^*}{Ns_0^*}. \tag{A.6}$$

Combining (A.6) with (8), (9) and (10) yields equation (11) for equilibrium s_0^*. The left-hand side of (11) is increasing in s_0^*, equal to 0 at $s_0^* = 1/(N+1)$ and goes to $+\infty$ as $s_0^* \to 1$. Hence, the solution to (11) exists and is unique for any $\Pi \geq 0$; moreover, $s_0^* = 1/(N+1)$ if $\Pi = 0$ and $s_0^* > 1/(N+1)$ if $\Pi > 0$. Hence, by (10), $s^* = 1/(N+1)$ if $\Pi = 0$ and $s^* < 1/(N+1)$ if $\Pi > 0$.

A.2 Proof of Proposition 2

Consumer's welfare is

$$W = \mathbb{E}\left[\max_n V_n\right] = \int\limits_{-\infty}^{+\infty} v f(v)\,\mathrm{d}v, \tag{A.7}$$

where $f(v)$ is pdf of $\max_n V_n$:

$$\begin{aligned}\Pr\left(\max_n V_n < v\right) &= \Pr\left(\max_n \mu_n\sigma - t_n < v - V\right)\\ &\overset{(2)}{=} \prod_{n=0}^{N} \exp\left(-\exp\left(-\frac{v-V+t_n}{\sigma} - \text{Euler's constant}\right)\right)\\ &\overset{t_n=t^*,\, t_0=t_0^*}{=} \exp\left(-\exp\left(-\frac{v-V}{\sigma} - \text{Euler's constant}\right)S^*\right),\end{aligned} \tag{A.8}$$

where

$$S^* = \exp\left(-\frac{t_0^*}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right). \tag{A.9}$$

Then

$$f(v) = \frac{S^*}{\sigma}\exp\left(-\frac{v-V}{\sigma} - \text{Euler's constant}\right)\exp\left(-\exp\left(-\frac{v-V}{\sigma} - \text{Euler's constant}\right)S^*\right),$$

and so, the change of variables $x = \exp\left(-\frac{v-V}{\sigma} - \text{Euler's constant}\right)S^*$ in (A.7) yields

$$W = \int_0^{+\infty}\left(\sigma\ln\frac{S^*}{x} + V - \sigma\text{Euler's constant}\right)\exp(-x)\,\mathrm{d}x = V + \sigma\ln S^*. \tag{A.10}$$

Substituting (A.9) into (A.10) yields (13).

A.3 Comparative Statics

In this Appendix we prove Lemma 1, Theorem 1, Propositions 3-7 as well as additional results presented in Table A.1.[26]

Let R_0 be company 0's equilibrium profit across both markets and R be equilibrium profit of company n, for $n = 1,\ldots,N$, in the product market. Then, $R_0 + NR$ is the joint profit of company 0 and all other companies that are present in the product market.

[26] The comparative statics with respect to γ is the same as the comparative statics with respect to Π because Π and γ affect the equilibrium only through their product, $\gamma\Pi$.

	Π	$\Pi = 0$	$\Pi \to +\infty$	N	$N \to +\infty$	σ	$\sigma \to 0$	$\sigma \to +\infty$
s_0^*	$+$	$\frac{1}{N+1}$	1	$-$	0	$-$	1 if $\Pi > 0$	$\frac{1}{N+1}$
s^*	$-$	$\frac{1}{N+1}$	0	$-$	0	$+$	0 if $\Pi > 0$	$\frac{1}{N+1}$
t_0^*	$-$	$\frac{N+1}{N}\sigma$	$-\infty$	$-$	$\sigma - \gamma\Pi$	$\cup$ if $\Pi > 0$	0	$+\infty$
t^*	$-$	$\frac{N+1}{N}\sigma$	σ	$-$	σ	$+$	0	$+\infty$
R_0	$+$	$\frac{1}{N}\sigma$	$+\infty$	$-$	0	$\cup$ if $\Pi > 0$	$\gamma\Pi$	$+\infty$
R	$-$	$\frac{1}{N}\sigma$	0	$-$	0	$+$	0	$+\infty$
$R_0 + NR$	$+$	$\frac{N+1}{N}\sigma$	$+\infty$	$-$	σ	$\cup$ if $\Pi > 0$	$\gamma\Pi$	$+\infty$
W	$+$	$V + \sigma\left(\ln(N+1) - \frac{N+1}{N}\right)$	$+\infty$	$+$	$+\infty$	$+$ if $N \geq 3$	V	$+\infty$ if $N \geq 3$
Δ_{R0}	$+$	0	$+\infty$	$-$	0	$-$	$\gamma\Pi$	$\frac{\gamma\Pi N}{N^2+N+1}$
Δ_{RN}	$-$	0	$-\sigma$	$+$	0	$\cup$ if $N \geq 3$	0	$-\frac{\gamma\Pi N}{N^2+N+1}$
Δ_W	$+$	0	$+\infty$	$-$	0	$\cap$ if $N \geq 2$	0	$\frac{\gamma\Pi}{N+1}$

Table A.1: Comparative statics results. The rows correspond to the equilibrium quantities; the columns correspond to the parameters of interest. An entry with $+$ ($-$; $\cup$; $\cap$) indicates that the row quantity increases (decreases; decreases and then increases; increases and then decreases) with respect to the column parameter.

Comparative statics with respect to Π

Let $s_0\left(\frac{\gamma\Pi}{\sigma}\right)$ be the solution to (11) (for notational simplicity, we sometimes omit the star in s_0^*). Applying the implicit function theorem to (11), we get

$$s_0'\left(\frac{\gamma\Pi}{\sigma}\right) = \frac{(1-s_0)^2 s_0 (N-1+s_0)^2}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} > 0, \tag{A.11}$$

so that s_0^* is increasing in Π. Hence, by (10), s^* is decreasing in Π. The fact that $s_0^* = s^* = 1/(N+1)$ if $\Pi = 0$ immediately follows from Proposition 1. Since the left-hand side of (11) goes to $+\infty$ as $s_0^* \to 1$, $s_0^* \to 1$ if $\Pi \to +\infty$, and so, by (10), $s^* \to 0$ if $\Pi \to +\infty$.

Since s^* is decreasing in Π from $1/(N+1)$ to 0, by (9), t^* is decreasing in Π from $\sigma(N+1)/N$ to σ.

Differentiating (8) with respect to Π and using (A.11) yield

$$t_0'(\Pi) = -\gamma \frac{(1-s_0)(N-1+s_0)^2 + N(1-s_0)^2 s_0}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} < 0. \tag{A.12}$$

Since s_0^* converges to $1/(N+1)$ as Π goes to 0, the limit of (8) is $\sigma(N+1)/N$. To derive the limit of (8) at $\Pi \to +\infty$, we substitute $\gamma\Pi$ from (11) and take the limit $s_0^* \to 1$; as a result, we get $-\infty$.

In equilibrium, company 0's total profit is $R_0 = s_0^*(t_0^* + \gamma\Pi)$, which after the substitution of t_0^* from (8) becomes

$$R_0 = \frac{\sigma s_0^*}{1 - s_0^*}. \tag{A.13}$$

Expression (A.13) increases in s_0^*. Thus, since s_0^* increases in Π from $1/(N+1)$ to 1, company 0's profit increases in Π from σ/N to $+\infty$.

Company n's profit is $R = s^* t^*$, which after the substitution of t^* from (9) becomes

$$R = \frac{\sigma s^*}{1 - s^*}. \tag{A.14}$$

Expression (A.14) increases in s^*. Thus, since s^* decreases in Π from $1/(N+1)$ to 0, company n's profit decreases in Π from σ/N to 0.

Substituting s^* from (10) to (A.14), we get the expression for the joint profit as a function of s_0^*:

$$R_0 + NR = \frac{\sigma s_0^*}{1 - s_0^*} + N\frac{\sigma(1 - s_0^*)}{N - 1 + s_0^*}. \tag{A.15}$$

The right-hand side of (A.15) is increasing in $s_0^* > 1/(N+1)$. Since s_0^* increases in Π from $1/(N+1)$ to 1, the joint profit $R_0 + NR$ increases in Π from $\sigma(N+1)/N$ to $+\infty$.

Substituting t_0^* from (8) and t^* from (9) into (13), then s^* from (10) and then Π from (11), we get the expression for the consumer welfare as a function of s_0^*:

$$W = V + \sigma\left(\ln\frac{N}{1 - s_0^*} - \frac{N}{N - 1 + s_0^*}\right). \tag{A.16}$$

Expression (A.16) increases in s_0^*. Thus, since s_0^* increases in Π from $1/(N+1)$ to 1, the consumer welfare increases in Π from $V + \sigma\left(\ln(N+1) - \frac{N+1}{N}\right)$ to $+\infty$.

Define the welfare gain of consumers from data linkage as

$$\Delta_W = W(\Pi) - W(0) > 0, \tag{A.17}$$

company 0's change in profit as

$$\Delta_{R0} = R_0(\Pi) - R_0(0) > 0, \tag{A.18}$$

and the change in the joint profit of all other companies as

$$\Delta_{RN} = NR(\Pi) - NR(0) < 0. \tag{A.19}$$

In (A.17), (A.18) and (A.19), the argument of W, R_0 and R is Π, which is equal to 0 when there is no data linkage. From comparative statics results with respect to Π, it follows that Δ_W and Δ_{R0} are positive, while Δ_{RN} is negative.

The comparative statics of Δ_{R0}, Δ_{RN} and Δ_W with respect to Π follows from the comparative statics of R_0, R and W.

Comparative statics with respect to N

Let $s_0(N)$ be the solution to (11). Applying the implicit function theorem to (11), we get

$$s_0'(N) = -\frac{(1-s_0)^2 s_0 \left(N(1-s_0)+(N-1+s_0)^2\right)}{N\left(N(1-s_0)^2 s_0+(N-1+s_0)^2\right)} < 0, \tag{A.20}$$

so that s_0^* is decreasing in N. Since for any fixed $s_0^* \in (0,1)$, the left-hand side of (11) goes to $+\infty$ as $N \to +\infty$, at the limit $N \to +\infty$, s_0^* cannot be strictly inside $(0,1)$. Since s_0^* is decreasing in N, it cannot be 1 at the limit. Hence, $s_0^* \to 0$.

To find how s^* changes with N, we differentiate (10):

$$s'(N) = -\frac{1-s_0+Ns_0'(N)}{N^2} \overset{(A.20)}{=} -\frac{(1-s_0)(N-1+s_0)^2(1-s_0+s_0^2)}{N^2\left(N(1-s_0)^2 s_0+(N-1+s_0)^2\right)} < 0, \tag{A.21}$$

so that s^* is decreasing in N. As $N \to +\infty$, since $s_0^* \to 0$, by (10), s^* converges to 0.

Since s_0^* and s^* are decreasing in N and go to 0 at the limit, t_0^*, t^*, R_0 and R are decreasing in N by (8), (9), (A.13) and (A.14), respectively, and their limits are $\sigma - \gamma\Pi$, σ, 0 and 0.

Differentiating the joint profit (A.15) with respect to N, we get

$$\frac{\mathrm{d}}{\mathrm{d}N}\left(R_0(N)+NR(N)\right) = -\frac{\sigma\left((1-s_0)^4-((N-1)(1-s_0)+N)((N+1)s_0-1)s_0'(N)\right)}{(1-s_0)^2(N-1+s_0)^2}, \tag{A.22}$$

which is negative because s_0^* is decreasing in N and $s_0^* > 1/(N+1)$. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.15) goes to σ.

Differentiating the consumer welfare (A.16) with respect to N, we get

$$W'(N) = \frac{\sigma\left(N(1-s_0)+(N-1+s_0)^2\right)\left(1-s_0+Ns_0'(N)\right)}{N(1-s_0)(N-1+s_0)^2}, \tag{A.23}$$

which is positive by (A.21). At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.16) goes to $+\infty$.

Substituting R_0 from (A.13) into (A.18) and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_{R0} = \frac{\sigma s_0^*}{1-s_0^*} - \frac{\sigma}{N}. \tag{A.24}$$

Differentiating (A.24) with respect to N and using (A.20), we get

$$\Delta'_{R0}(N) = -\frac{(N-1)(N-1+s_0)((N+1)s_0-1)\sigma}{N^2\left(N(1-s_0)^2 s_0 + (N-1+s_0)^2\right)} < 0, \tag{A.25}$$

so that Δ_{R0} is decreasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.24) goes to 0.

Substituting R from (A.14) into (A.19), then s^* from (10) and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_{RN} = \frac{\sigma(1-s_0^*)}{1-(1-s_0^*)/N} - \sigma. \tag{A.26}$$

Differentiating (A.26) with respect to N and using (A.20), we get

$$\Delta'_{RN}(N) = \frac{(N-1)(1-s_0)^2((N+1)s_0-1)\sigma}{(N-1+s_0)\left(N(1-s_0)^2 s_0 + (N-1+s_0)^2\right)} > 0, \tag{A.27}$$

so that Δ_{RN} is increasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.26) goes to 0.

Substituting W from (A.16) into (A.17) and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_W = \sigma\left(\ln\frac{1-1/(N+1)}{1-s_0^*} + \frac{N}{N-1+1/(N+1)} - \frac{N}{N-1+s_0^*}\right). \tag{A.28}$$

Differentiating (A.28) with respect to N and using (A.20), we get

$$\Delta'_W(N) = -\frac{\left(\left(N^3-1+N(1-s_0)^2\right)(1-s_0)+N\right)((N+1)s_0-1)\sigma}{N^2(N+1)\left(N(1-s_0)^2 s_0 + (N-1+s_0)^2\right)} < 0, \tag{A.29}$$

so that Δ_W is decreasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.28) goes to 0.

Comparative statics with respect to σ

By (A.11), s_0^* is decreasing in σ. Hence, by (10), s^* is increasing in σ. Since by (11) s_0^* depends on σ and Π only through Π/σ, at the limits $\sigma \to 0$ and $\sigma \to +\infty$, s_0^* behaves in the same way as at the limits $\Pi \to +\infty$ and $\Pi \to 0$. Hence, $s_0^* \to 1$ if $\sigma \to 0$ provided that $\Pi > 0$, and $s_0^* \to 1/(N+1)$ if $\sigma \to +\infty$. Then, by (10), $s^* \to 0$ if $\sigma \to 0$ provided that $\Pi > 0$, and $s^* \to 1/(N+1)$ if $\sigma \to +\infty$.

Since s^* is increasing in σ, by (9) and (A.14), t^* and R are also increasing in σ. As $\sigma \to 0$,

$s^* = 1/(N+1)$ if $\Pi = 0$ and $s^* \to 0$ if $\Pi > 0$; in either case, (9) and (A.14) converge to 0. As $\sigma \to +\infty$, $s^* \to 1/(N+1)$, and so, (9) and (A.14) go to $+\infty$.

Differentiating (8) with respect to σ, using (A.11) and substituting Π from (11) yield

$$t_0'(\sigma) = \frac{\frac{N^2}{(N-1+s_0)^2} + \frac{1}{s_0} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \tag{A.30}$$

The numerator in (A.30) is decreasing in $s_0^* \in (1/(N+1), 1)$ from a positive value to $-\infty$. If $\Pi > 0$, then s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, and so, (A.30) is negative and then positive. If $\sigma \to +\infty$, then $s_0^* \to 1/(N+1)$, and so, (8) goes to $+\infty$. If $\Pi > 0$ and $\sigma \to 0$, $s_0^* \to 1$ and thus from (11) it follows that

$$\lim_{\sigma \to 0} \frac{\sigma}{1 - s_0^*(\sigma)} = \gamma\Pi. \tag{A.31}$$

If $\Pi = 0$, then $s_0^* = 1/(N+1)$ and so (A.31) also holds. Then, (A.31) implies that (8) converges to 0 as $\sigma \to 0$.

Differentiating (A.13) with respect to σ, using (A.11) and substituting Π from (11) yield

$$R_0'(\sigma) = \frac{\frac{N\left(N-(1-s_0)^2\right)}{(N-1+s_0)^2} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \tag{A.32}$$

The numerator in (A.32) is decreasing in $s_0^* \in (1/(N+1), 1)$ from a positive value to $-\infty$. If $\Pi > 0$, then s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, and so, (A.32) is negative and then positive. If $\sigma \to +\infty$, then $s_0^* \to 1/(N+1)$, and so, (A.13) goes to $+\infty$. If $\Pi > 0$ and $\sigma \to 0$, then (A.13) converges to $\gamma\Pi$ because $s_0^* \to 1$ and (A.31) holds. If $\Pi = 0$ and $\sigma \to 0$, then $s_0^* = 1/(N+1)$ and so (A.13) goes to 0.

Differentiating the joint profit (A.15) with respect to σ, using (A.11) and substituting Π from (11) yield

$$R_0'(\sigma) + NR'(\sigma) = \frac{s_0\left((N-1)(1-s_0)+N\right)\left((N+1)s_0-1\right)}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} \times \left(\frac{N\left((N+1)\left(1+(1-s_0)s_0^2\right)-2+s_0\right)}{s_0\left((N-1)(1-s_0)+N\right)\left((N+1)s_0-1\right)} + \ln\left(\frac{1-s_0}{Ns_0}\right)\right). \tag{A.33}$$

The numerator in (A.33) is decreasing in $s_0^* \in (1/(N+1), 1)$ from $+\infty$ to $-\infty$. If $\Pi > 0$, then s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, and so, (A.33) is negative and then positive.

Differentiating the consumer welfare (A.16) with respect to σ, using (A.11) and substituting Π

from (11) yield $W'(\sigma) = w(N, s_0)$, where

$$w(N,s_0) = \frac{(N-1+s_0)^2(1-(1-s_0)s_0)\ln\left(\frac{Ns_0}{1-s_0}\right) - (1-s_0)^2 s_0 - N(N-1+s_0)\left(1+s_0^2\right)}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} - \ln s_0. \quad \text{(A.34)}$$

Function $w(N, s_0)$ is increasing in N for $N \geq \max\left\{3, \frac{1-s_0}{s_0}\right\}$:

$$\begin{aligned}\frac{\partial w(N,s_0)}{\partial N} = \frac{1-(1-s_0)s_0}{(N(1-s_0)^2 s_0 + (N-1+s_0)^2)^2}\Big\{ & N^2(N-(3-s_0)(1-s_0)) + \frac{(1-s_0)^4}{N} \\ & + 2N(1-s_0)^2(1-(1-s_0)s_0) + \left(2N-(1-s_0)(3+s_0^2)\right)(1-s_0)^2 \\ & + (1-s_0+N)(1-s_0)^2 s_0 (N-1+s_0)\ln\left(\frac{Ns_0}{1-s_0}\right)\Big\} > 0, \end{aligned} \quad \text{(A.35)}$$

and positive at $N = \max\left\{3, \frac{1-s_0}{s_0}\right\}$. Hence, $W'(\sigma) > 0$ if $N \geq 3$ and $N \geq \frac{1-s_0^*}{s_0^*}$. The latter condition is equivalent to $s_0^* \geq 1/(N+1)$, which always holds. Hence, W is increasing in σ if $N \geq 3$. At the limit $\sigma \to 0$, (A.16) converges to V because (A.31) holds and $s_0^* \to 1$ if $\Pi > 0$ and $s_0^* = 1/(N+1)$ if $\Pi = 0$. At the limit $\sigma \to +\infty$, $s_0^* \to 1/(N+1)$, and so, (A.16) goes to $+\infty$ if $N \geq 3$.

Differentiating (A.18) with respect to σ, using (A.32) and substituting $s_0^* = 1/(N+1)$ for $\Pi = 0$, we get

$$\Delta'_{R0}(\sigma) = \frac{1 + \frac{1}{N} - \frac{((N+1)s_0-1)^2}{N(N-1+s_0)^2} - \frac{1}{Ns_0} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \quad \text{(A.36)}$$

The numerator in (A.36) is decreasing in $s_0 \in (1/(N+1), 1)$ and equal to 0 at $s_0 = 1/(N+1)$. Hence, (A.36) is negative for all $s_0 \in (1/(N+1), 1)$. Since $s_0^* > 1/(N+1)$, Δ_{R0} is decreasing in σ. If $\sigma \to +\infty$, $s_0^* \to 1/(N+1)$ and thus from (11) it follows that

$$\lim_{\sigma \to +\infty} \Delta_{R0}(\sigma) \overset{\text{(A.24)}}{=} \lim_{\sigma \to +\infty} \frac{\sigma\left((N+1)s_0^*(\sigma) - 1\right)}{N\left(1 - s_0^*(\sigma)\right)} = \frac{\gamma \Pi N}{N^2 + N + 1}. \quad \text{(A.37)}$$

From results on R_0, at the limit $\sigma \to 0$, $\Delta_{R0} = \gamma\Pi$.

Differentiating (A.26) with respect to σ, using (A.11) and substituting Π from (11) yield

$$\Delta'_{RN}(\sigma) = \frac{\frac{1}{s_0}\left(\frac{s_0^2}{1-s_0} - \frac{1}{N}\right)\left(\frac{1}{N} - \frac{1}{1-s_0}\right) - 1 - \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{1}{N} + \frac{(N-1+s_0)^2}{N^2(1-s_0)^2 s_0}}. \quad \text{(A.38)}$$

The numerator in (A.38) is equal to 0 at $s_0 = 1/(N+1)$, $-\infty$ at $s_0 \to 1$, and its derivative with respect to s_0 is equal to

$$\frac{\left(s_0-\frac{1}{N+1}\right)\left(s_0+\frac{3N-2+\sqrt{(9N-8)N}}{2}\right)}{N^2(1-s_0)^3 s_0^2}\left(1+\underbrace{\frac{\frac{2N}{N+1}\left((N-1)^2-3\right)}{\sqrt{(9N-8)N}+\frac{N(1+3N)}{N+1}}}_{>0\text{ if }N\geq 3}-(N+1)s_0\right). \tag{A.39}$$

Hence, if $N \geq 3$, then, as a function of $s_0 \in (1/(N+1), 1)$, the numerator in (A.38) is positive, then negative. Since s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, (A.38) is negative and then positive as a function of $\sigma \in (0, +\infty)$. To find the limit $\sigma \to +\infty$, observe that from (A.24) and (A.26), we get

$$\frac{\Delta_{RN}}{\Delta_{R0}} = -\frac{N(1-s_0^*)}{N-1+s_0^*}. \tag{A.40}$$

As $\sigma \to +\infty$, $s_0^* \to 1/(N+1)$, and so, (A.40) converges to -1. Then $\lim_{\sigma\to+\infty} \Delta_{RN}(\sigma) = -\lim_{\sigma\to+\infty} \Delta_{R0}(\sigma)$. From results on R, at the limit $\sigma \to 0$, $\Delta_{RN} = 0$.

Differentiating (A.28) with respect to σ twice and using (A.11) yield

$$\Delta_W''(\sigma) = \frac{\gamma^2\Pi^2(N-1+s_0)^3(1-s_0)^2 s_0 \tilde{w}(N,s_0)}{\sigma^3\left(N(1-s_0)^2 s_0+(N-1+s_0)^2\right)^3}, \tag{A.41}$$

where

$$\tilde{w}(N,s_0) = N(1-s_0)\left((1-s_0)\left(N(1+s_0^2)-(1-s_0)^2(1+s_0)\right)-2s_0N\right)+(1-2s_0)(N-1+s_0)^3. \tag{A.42}$$

Function $\tilde{w}(N,s_0)$ is decreasing in $s_0 \in (0,1)$:

$$\begin{aligned}\frac{\partial \tilde{w}(N,s_0)}{\partial s_0} = &-\frac{5+6(1-2s_0)^2+5(1-2s_0)^4}{16}-(N-1)\left(2s_0+5(1-s_0)^4\right)\\ &-(N-1)^2\left(2s_0(1-s_0)(1+2s_0)+(1-2s_0)^2\right)-2\left(N-1+s_0\right)^3<0,\end{aligned} \tag{A.43}$$

negative at $s_0 = 1$ and positive at $s_0 = 1/(N+1)$ if $N \geq 2$:

$$\tilde{w}(N,1) = -N^3, \quad \tilde{w}\left(N, \frac{1}{N+1}\right) = \frac{N^3(1+N+N^2)\left((N-2)(N^2+2)+2(N-1)^2\right)}{(N+1)^5}. \tag{A.44}$$

Since s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, $\Delta_W''(\sigma)$ is negative and then positive. Hence, $\Delta_W'(\sigma)$ is decreasing and then increasing. Differentiating (A.28) with respect to σ, using (A.11) and substituting Π from (11) yield

$$\Delta_W'(\sigma) = 1 + \frac{1}{N} - \frac{N}{N-1+s_0} + \ln\left(\frac{N}{(N+1)(1-s_0)}\right) - \frac{N(1-s_0)+(N-1+s_0)^2}{N(1-s_0)+\frac{(N-1+s_0)^2}{(1-s_0)s_0}}\left(\frac{(N+1)s_0-1}{(1-s_0)(N-1+s_0)} - \ln\left(\frac{1-s_0}{Ns_0}\right)\right). \quad \text{(A.45)}$$

At $s_0 = 1/(N+1)$, (A.45) is equal to 0. At the limit $s_0 \to 1$, (A.45) goes to $+\infty$. Hence, $\Delta_W(\sigma)$ is increasing and then decreasing. To find the limit $\sigma \to +\infty$, observe that from (A.24) and (A.28), the ratio Δ_W/Δ_{R0} depends on σ only through s_0^*. As $\sigma \to +\infty$, $s_0^* \to 1/(N+1)$ and thus the ratio Δ_W/Δ_{R0} converges to $(N^2+N+1)/(N^2+N)$. Then $\lim_{\sigma\to+\infty} \Delta_W(\sigma) = \frac{N^2+N+1}{N^2+N} \lim_{\sigma\to+\infty} \Delta_{R0}(\sigma) = \frac{\gamma\Pi}{N+1}$. From results on W, at the limit $\sigma \to 0$, $\Delta_W = 0$.

Appendix B Additional Results

B.1 Data-Sharing Remedy

As discussed in Section 5.3, when company 0 is forced to share the information it obtains in the product market with other companies in the insurance market, data linkage does not affect the product market and, in particular, company 0's market share remains $s_0 = 1/(N+1)$. In the insurance market, the consumers with the known risk type — that is, $1/(N+1)$ share of consumers served by company 0 in the product market — get their first-best contract which involves full insurance, $q_i = l$, at the fair premium, $p_i = \pi_i l$. Without data linkage, low risk consumers get (4), which is lower than their utility from the first-best contract, $u(y - \pi_L l)$. Hence, overall, with data sharing, the consumer welfare gain from data linkage comes exclusively from the insurance market and equals

$$\Delta_W^S = \frac{\gamma}{N+1}\big(u(y-\pi_L l) - u(y-p^I)\big). \quad \text{(B.1)}$$

To determine whether the data-sharing remedy benefits the consumers, we need to compare (B.1) with Δ_W defined in (A.17), the consumer welfare gain from data linkage without data sharing. Theorem B.1 shows that, under an additional restriction $u'(0) \le 1$,[27] the data sharing remedy benefits consumers, $\Delta_W^S > \Delta_W$, if and only if the taste heterogeneity in the product market is sufficiently low.

Theorem B.1. *If $u'(0) \le 1$, then there exists a finite $\hat{\sigma} > 0$ such that $\Delta_W^S > \Delta_W$ for all $\sigma < \hat{\sigma}$ and $\Delta_W^S < \Delta_W$ for all $\sigma > \hat{\sigma}$.*

[27] For example, CARA utility $u(x) = \frac{1-e^{-\lambda x}}{\lambda}$ satisfies $u'(0) \le 1$. Restriction $u'(0) \le 1$ makes consumer utility comparable with company's profit.

Proof. From Table A.1, Δ_W has a hump-shaped form in $\sigma \in (0, +\infty)$, increasing from 0 at $\sigma \to 0$ and then decreasing to $\gamma\Pi/(N+1)$ at $\sigma \to +\infty$. Lemma B.1 shows that $\Delta_W^S < \gamma\Pi/(N+1)$. Thus, there exists a threshold such that $\Delta_W^S > \Delta_W$ if and only if σ is below this threshold.

Lemma B.1. *If $u'(0) \leq 1$, then (B.1) is lower than $\gamma\Pi/(N+1)$, where Π is defined by (5).*

Proof. Since $u'(0) \leq 1$ and $u(x)$ is concave, $x - u(x)$ is increasing for all $x > 0$. Since $x - u(x)$ is increasing in $x > 0$ and $p^I > \pi_L l$, $(y - p^I) - u(y - p^I) < (y - \pi_L l) - u(y - \pi_L l)$. Thus, (B.1) is lower than $\gamma(p^I - \pi_L l)/(N+1)$. □

□

B.2 Monopolistic Insurance Market

Suppose that only one company, company 0, operates in the insurance market. When markets are informationally linked, this company learns the risk type of the insurance consumer by serving this consumer in the product market. All other aspects of the model remain as described in Section 2.

B.2.1 Insurance Market

The equilibrium in monopolistic insurance market has been derived in Stiglitz (1977). Figure B.1 illustrates the equilibrium in the same space as Figure 2.

If the monopolist knows the risk-type of its customer, then he offers a full-insurance contract to each type, with low-risk consumers paying a lower premium — see points II in Figure B.1. The contract makes each consumer, independent of his risk type, indifferent to buying the insurance; that is, the monopolist fully extracts consumer surplus.

If the monopolist does not know the risk type of the consumer, the equilibrium contract for the high-risk consumer still features full insurance, that is, $q_H = l$. However, to prevent the high-risk consumer from choosing the contract designed for the low-risk consumer, the equilibrium contract for the low-risk consumer features partial or no insurance.

If γ, the share of the low-risk consumers in the population, is below certain threshold γ^*, the monopolist does not serve the low-risk consumer at all and to the high-risk consumer offers a full-insurance at a premium which makes him indifferent to buying insurance. Formally, $p_L(\gamma) = q_L(\gamma) = 0$ and $p_H(\gamma)$ is defined from

$$u(y - p_H) = \pi_H u(y - l) + (1 - \pi_H) u(y). \tag{B.2}$$

In Figure B.1 the contract to the low-risk and high risk consumer correspond to the black point and red point II, respectively.

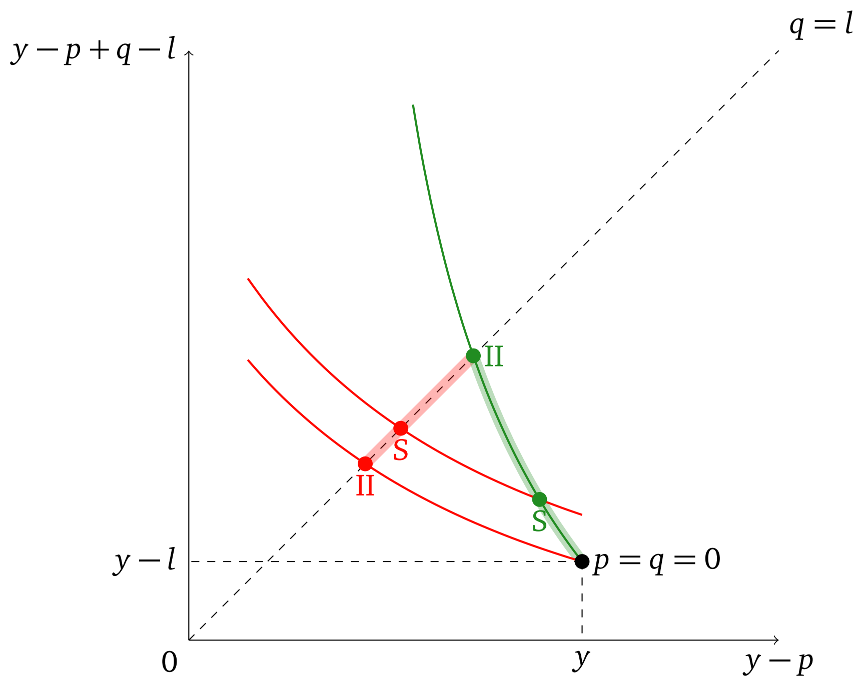


Figure B.1: Equilibrium in the monopolistic insurance market, drawn for the utility function $u(x) = 1-e^{-x}$. Red and green S points correspond to the Stiglitz equilibrium contract for the high- and low-risk consumer, respectively. Red and green points II correspond to the contract that the informed monopolistic insurer offers to the high- and low-risk consumer, respectively.

For γ above γ^*, the low-risk consumer gets partial insurance at a premium that makes him indifferent to buying no insurance:

$$\pi_L u(y-p_L+q_L-l)+(1-\pi_L)u(y-p_L)=\pi_L u(y-l)+(1-\pi_L)u(y). \tag{B.3}$$

This partial insurance contract should not be attractive to the high-risk consumers, that is, the incentive compatibility constraint of the high-risk consumers is satisfied:

$$u(y-p_H)=\pi_H u(y-p_L+q_L-l)+(1-\pi_H)u(y-p_L). \tag{B.4}$$

Finally, the pair of the offered contracts maximizes the monopolist's expected profit and so satisfies the additional optimality condition:[28]

$$\frac{(1-\pi_L)\pi_L}{\pi_H-\pi_L}\left(\frac{u'(y-p_H)}{u'(y-p_L)}-\frac{u'(y-p_H)}{u'(y-p_L+q_L-l)}\right)=\frac{1-\gamma}{\gamma}. \tag{B.5}$$

[28]Equality (B.5) follows from first order conditions for the principal's optimization problem: maximize $\gamma(p_L-\pi_L q_L)+(1-\gamma)(p_H-\pi_H l)$ subject to (B.3) and (B.4).

Equations (B.3), (B.4) and (B.5) define optimal $q_L(\gamma)$, $p_L(\gamma)$ and $p_H(\gamma)$. In Figure B.1, the low-risk contract corresponds to the green point S and the high-risk contract corresponds to red point S.

Threshold γ^* is defined as the lowest γ, for which $q_L(\gamma)$, $p_L(\gamma)$ and $p_H(\gamma)$, the solution to (B.3)-(B.5), satisfy the participation constraint for the high-risk type; that is, at $\gamma = \gamma^*$, equality (B.2) holds. Thus, there is no discrete change in the contracts as we move from the region where $\gamma < \gamma^*$ to the region where $\gamma > \gamma^*$. When $\gamma \leq \gamma^*$, the contracts do not change with γ. As γ increases above γ^*, the cover in the low-risk contract, $q_L(\gamma)$, increases, while the premium in the high-risk contract, $p_H(\gamma)$, decreases. In Figure B.1, as γ increases, the green point S, which corresponds to the low-risk contract, moves up along the highlighted segment of the indifference curve of the low-risk consumer from the black point to green point II. At the same time, red point S, which corresponds to the high-risk contract, moves up along the highlighted segment of the 45-degree line from red point II to green point II.

B.2.2 Product Market

Irrespective of whether markets are informationally linked, the monopolistic insurer keeps the low-risk consumer indifferent to buying the insurance. Hence, the low-risk consumers are indifferent to the monopolistic insurer knowing their type and so they have no incentives to conceal their type by avoiding variety 0 in the product market. Thus, the demand of low-risk consumers for each variety is given by (6).

In contrast, depending on γ, the high-risk consumers may have incentives to hide their risk type from the monopolistic insurer.

Low γ

When $\gamma < \gamma^*$, the high-risk consumer has no incentives to avoid variety 0 in the product market because the insurer's knowledge of his type does not affect the offered contract — he receives full insurance at a premium which makes him indifferent to buying insurance (see red point II in Figure B.1). Thus, the demand of high-risk consumers is the same as the demand of low-risk consumers and is given by (6).

Since all consumers are indifferent to revealing their risk type to company 0 and data linkage does not affect the contract for the high-risk consumers, the analysis from Section 3.2 applies. However, Π is now defined as

$$\Pi = (p_L(1) - \pi_L l) - (p_L(\gamma) - \pi_L q_L(\gamma)) = p_L(1) - \pi_L l, \tag{B.6}$$

which is the difference in the insurer's profit from contracts corresponding to green point II and to black point in Figure B.1. Premium $p_L(1)$ makes the low-risk type indifferent to buying the insurance:

$$u(y - p_L(1)) = \pi_L u(y - l) + (1 - \pi_L)u(y). \tag{B.7}$$

From Section 3.2, it follows that all consumers are strictly better off in the presence of data linkage. The mechanism behind the welfare improvement is exactly the same as in the case of competitive insurance market. Data linkage promotes contract efficiency for the low-risk consumers without harming either low- or high-risk consumers in the insurance market; then, in the product market, some of this efficiency gain is passed on to consumers through lower prices. Moreover, with the monopolistic insurer, the efficiency improvement is more stark than in a competitive market, because without data linkage, the monopolistic insurer does not serve the low-risk consumers at all, and so data linkage opens the insurance market to new consumers.

High γ

When $\gamma > \gamma^*$, the high-risk consumers get a disutility from revealing their type — instead of getting contract marked by red point S, they get contract marked by red point II in Figure B.1. In equilibrium, the high-risk consumers take into account this disutility when choosing a variety in the product market.

Formally, let s_0^L (s_0^H) be the low-risk (high-risk) consumer equilibrium demand for variety 0. As a result of data linkage between the markets, company 0 knows the risk type of the consumer with probability s_0^L if the consumer is of low risk and with probability s_0^H if the consumer is of high risk. Then, conditional on company 0 not knowing the consumer's risk type, the probability that the consumer is of low risk is

$$\gamma' = \frac{\gamma(1 - s_0^L)}{\gamma(1 - s_0^L) + (1 - \gamma)(1 - s_0^H)}. \tag{B.8}$$

Thus, in equilibrium, when hiding his risk type, the high-risk consumer pays premium $p_H(\gamma')$ and, relative to revealing his type, gains

$$\delta V = u(y - p_H(\gamma')) - \pi_H u(y - l) - (1 - \pi_H)u(y). \tag{B.9}$$

In contrast to the case of low γ, serving a consumer in the product market allows company 0 to earn additional profit in the insurance market on both high- and low-risk consumers. Moreover, the profit that company 0 gets in the insurance market from consumers it does not serve in the product market is also affected by data linkage through γ'. Overall, company 0's additional profit from data linkage is

$$\delta\Pi = \gamma s_0^L \Pi_L(1) + \gamma(1 - s_0^L)\Pi_L(\gamma') + (1 - \gamma)s_0^H \Pi_H(0) + (1 - \gamma)(1 - s_0^H)\Pi_H(\gamma'), \tag{B.10}$$

where

$$\Pi_L(x) = (p_L(x) - \pi_L q_L(x)) - (p_L(\gamma) - \pi_L q_L(\gamma)), \quad \Pi_H(x) = p_H(x) - p_H(\gamma) \tag{B.11}$$

with $q_L(1) = l$.

We are looking for a symmetric equilibrium in the product market. Let t^* be the price set by each company $n = 1, 2, \ldots, N$; let t_0^* be the price set by company 0. We view δV and $\delta\Pi$, defined in (B.9) and (B.10), as functions of s_0^L and s_0^H, and let δV_L (δV_H) and $\delta\Pi_L$ ($\delta\Pi_H$) denote partial derivatives of these functions with respect to s_0^L (s_0^H).

Proposition B.1. *In equilibrium, the prices t_0^* and t^* and the demands s_0^L and s_0^H solve the system of four equations:*

$$t_0^* = t^* + \sigma \ln \frac{1 - s_0^L}{N s_0^L}, \tag{B.12}$$

$$\sigma \ln \frac{s_0^L(1 - s_0^H)}{(1 - s_0^L) s_0^H} = \delta V, \tag{B.13}$$

$$t^* \left\{ \gamma(1 - s_0^L) \left(1 - \frac{1 - s_0^L}{N} \right) + (1 - \gamma)(1 - s_0^H) \left(1 - \frac{1}{N} + \frac{s_0^H}{N} \frac{\sigma - s_0^L(1 - s_0^L)\delta V_L}{\sigma + s_0^H(1 - s_0^H)\delta V_H} \right) \right\} \\ = \sigma \left(\gamma(1 - s_0^L) + (1 - \gamma)(1 - s_0^H) \right), \tag{B.14}$$

$$s_0^L(1 - s_0^L)\left(\gamma t_0^* + \delta\Pi_L\right) + s_0^H(1 - s_0^H) \frac{\sigma - s_0^L(1 - s_0^L)\delta V_L}{\sigma + s_0^H(1 - s_0^H)\delta V_H} \left((1 - \gamma) t_0^* + \delta\Pi_H\right) \\ = \sigma \left(\gamma s_L^0 + (1 - \gamma) s_H^0 \right). \tag{B.15}$$

Proof. Company 0 chooses price t_0 to maximize

$$\max_{t_0} \left(\gamma s_0^L + (1 - \gamma) s_0^H \right) t_0 + \delta\Pi, \tag{B.16}$$

where the demand functions are

$$s_0^L = \frac{\exp\left(-\frac{t_0}{\sigma}\right)}{\exp\left(-\frac{t_0}{\sigma}\right) + N \exp\left(-\frac{t^*}{\sigma}\right)}, \quad s_0^H = \frac{\exp\left(-\frac{\delta V + t_0}{\sigma}\right)}{\exp\left(-\frac{\delta V + t_0}{\sigma}\right) + N \exp\left(-\frac{t^*}{\sigma}\right)}. \tag{B.17}$$

Equations (B.17) imply that

$$\frac{\mathrm{d}s_0^L(t_0)}{\mathrm{d}t_0} = -\frac{s_0^L(1-s_0^L)}{\sigma}, \quad \frac{\mathrm{d}s_0^H(t_0)}{\mathrm{d}t_0} = -\frac{s_0^H(1-s_0^H)}{\sigma}\frac{\sigma - s_0^L(1-s_0^L)\delta V_L}{\sigma + s_0^H(1-s_0^H)\delta V_H}. \tag{B.18}$$

Using (B.18), we can show that the first order condition for (B.16) is (B.15).

Company $n \geq 1$ chooses price t to maximize

$$\max_t \left(\gamma s^L + (1-\gamma)s^H\right) t, \tag{B.19}$$

where the demand functions are

$$s^L = \frac{\exp\left(-\frac{t}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right) + \exp\left(-\frac{t}{\sigma}\right) + (N-1)\exp\left(-\frac{t^*}{\sigma}\right)}, \quad s^H = \frac{\exp\left(-\frac{t}{\sigma}\right)}{\exp\left(-\frac{\delta V + t_0^*}{\sigma}\right) + \exp\left(-\frac{t}{\sigma}\right) + (N-1)\exp\left(-\frac{t^*}{\sigma}\right)}, \tag{B.20}$$

where δV depends on t through the demands for variety 0, s_0^L and s_0^H, defined as

$$s_0^L = \frac{\exp\left(-\frac{t_0^*}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right) + \exp\left(-\frac{t}{\sigma}\right) + (N-1)\exp\left(-\frac{t^*}{\sigma}\right)}, \quad s_0^H = \frac{\exp\left(-\frac{\delta V + t_0^*}{\sigma}\right)}{\exp\left(-\frac{\delta V + t_0^*}{\sigma}\right) + \exp\left(-\frac{t}{\sigma}\right) + (N-1)\exp\left(-\frac{t^*}{\sigma}\right)}. \tag{B.21}$$

Equations (B.21) imply that

$$\frac{\mathrm{d}s_0^L(t)}{\mathrm{d}t} = \frac{s_0^L(1-s_0^L)}{\sigma\left(1+(N-1)\exp\left(\frac{t-t^*}{\sigma}\right)\right)}, \quad \frac{\mathrm{d}s_0^H(t)}{\mathrm{d}t} = \frac{s_0^H(1-s_0^H)}{\sigma\left(1+(N-1)\exp\left(\frac{t-t^*}{\sigma}\right)\right)}\frac{\sigma - s_0^L(1-s_0^L)\delta V_L}{\sigma + s_0^H(1-s_0^H)\delta V_H}. \tag{B.22}$$

Using (B.22), we differentiate the demand functions (B.20):

$$\frac{\mathrm{d}s^L(t)}{\mathrm{d}t} = -\frac{s^L(1-s^L)}{\sigma}, \tag{B.23}$$

$$\frac{\mathrm{d}s^H(t)}{\mathrm{d}t} = -\frac{s^H}{\sigma}\left(\frac{N-1}{N-1+\exp\left(\frac{t^*-t}{\sigma}\right)} + \left(\frac{1}{1+(N-1)\exp\left(\frac{t-t^*}{\sigma}\right)} - s^H\right)\frac{\sigma - s_0^L(1-s_0^L)\delta V_L}{\sigma + s_0^H(1-s_0^H)\delta V_H}\right). \tag{B.24}$$

In equilibrium, $t = t^*$, and so, from (B.20) and (B.21), $s^L = (1-s_0^L)/N$ and $s^H = (1-s_0^H)/N$. Using that and expressions (B.23) and (B.24), we can show that the first order condition for (B.19) is (B.14).

Equations (B.12) and (B.13) are re-arrangements of equations in (B.17) when $t_0 = t_0^*$. □

Equilibrium described in Proposition B.1 features the difference in the high-risk and low-risk consumer demand for variety 0. Since high-risk consumers suffer a disutility from revealing their type to the monopolistic insurer, they tend to avoid variety 0 in the product market, which implies that $s_0^H < s_0^L$. As a result, $\gamma' < \gamma$, which is illustrated in Figure B.2. The fact that in equilibrium, $\gamma' < \gamma$ indicates that

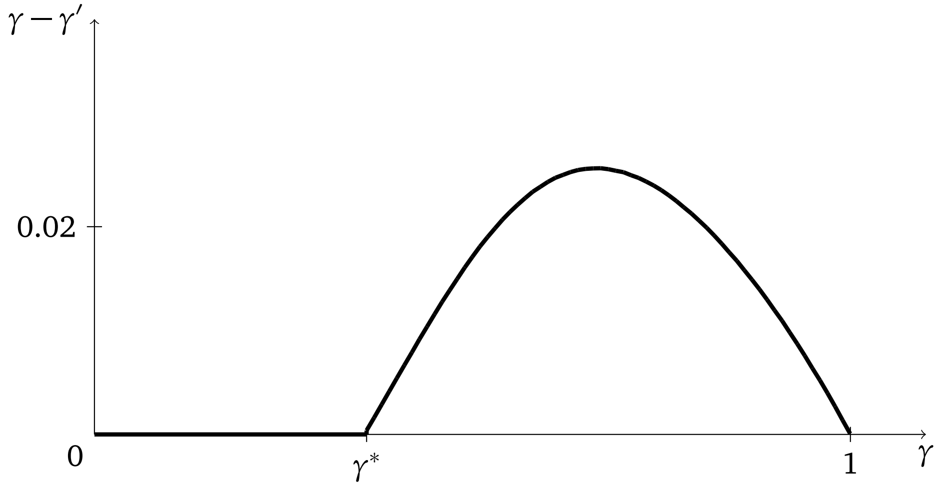


Figure B.2: The difference between γ, the share of low-risk consumers, and γ', the share of low risk consumers among those consumers whom company 0 does not serve in the product market. Parameters: $u(x) = 1 - e^{-x}$, $l = y = 1$, $\sigma = 0.1$, $N = 5$, $\pi_H = 0.8$, $\pi_L = 0.6$.

the low-risk consumers, who have no incentives to hide their type from company 0, pose negative externality to the high-risk consumers, who suffer a disutility from revealing their type. Indeed, when the high-risk consumer does not reveal his risk type to company 0, data linkage changes his utility in the insurance market from $u(y - p_H(\gamma))$ to $u(y - p_H(\gamma'))$; this change is negative because $\gamma' < \gamma$.

Proposition B.2 derives the change in the consumer welfare as a result of data linkage.

Proposition B.2. *As a result of data linkage, the low-risk consumers welfare increases by*

$$\Delta_W^L = \sigma \left(\ln \frac{N}{(N+1)(1-s_0^L)} + \frac{N+1}{N} \right) - t^*, \tag{B.25}$$

while the high-risk consumers welfare increases by

$$\Delta_W^H = \sigma \left(\ln \frac{N}{(N+1)(1-s_0^H)} + \frac{N+1}{N} \right) - t^* + u(y - p_H(\gamma')) - u(y - p_H(\gamma)). \tag{B.26}$$

Proof. Reasoning similarly as in the proof for Proposition 2, we derive the consumer welfare in the product market with data linkage:

$$W_{\text{linked}}^i = V + \sigma \ln S^i, \quad i = L, H, \tag{B.27}$$

where

$$S^L = \exp\left(-\frac{t_0^*}{\sigma}\right) + N \exp\left(-\frac{t^*}{\sigma}\right) \tag{B.28}$$

for the low-risk consumers and

$$S^H = \exp\left(-\frac{\delta V + t_0^*}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right) \tag{B.29}$$

for the high-risk consumers. Substituting δV from (B.13) into S^H, and t_0^* from (B.12) into S^L and S^H yield

$$W^i_{\text{linked}} = V - t^* + \sigma \ln \frac{N}{1-s_0^i}. \tag{B.30}$$

Without data linkage, the consumer welfare in the product market is given in Table A.1:

$$W^i_{\text{independent}} = V + \sigma\left(\ln(N+1) - \frac{N+1}{N}\right). \tag{B.31}$$

Hence, as result of data linkage, consumer welfare gain in the product market is

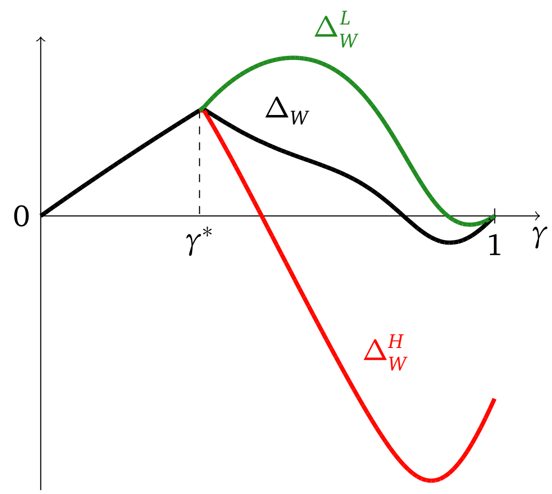
$$W^i_{\text{linked}} - W^i_{\text{independent}} = \sigma\left(\ln\frac{N}{(N+1)(1-s_0^i)} + \frac{N+1}{N}\right) - t^* \tag{B.32}$$

In the insurance market, as a result of data linkage, the low-risk consumer welfare does not change, while the high-risk consumer welfare changes from $u(y - p_H(\gamma))$ to $u(y - p_H(\gamma'))$.[29] Together with (B.32), it gives us (B.25) and (B.26). □

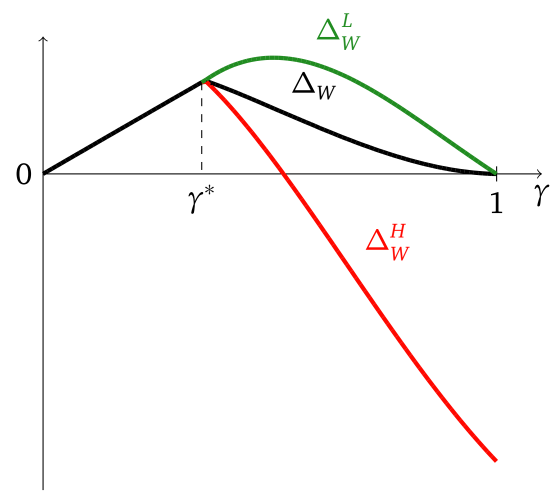
Proposition B.2 implies that the low-risk consumers gain from data linkage more than the high-risk consumers; that is, $\Delta^L_W > \Delta^H_W$. Indeed, as we discussed above, the term $u(y - p_H(\gamma')) - u(y - p_H(\gamma))$ is negative. Moreover, by (B.13), since $\delta V > 0$, $s_0^L > s_0^H$, that is, the low-risk consumers demand more variety 0 than the high-risk consumers. Hence, the direct comparison of (B.25) and (B.26) gives $\Delta^L_W > \Delta^H_W$.

A priori it is not clear whether consumers benefit from data linkage, that is, the signs of Δ^L_W and Δ^H_W are ambiguous. Figure B.3 shows that both low- and high-risk consumers may benefit from data linkage, that is, both Δ^L_W and Δ^H_W may be positive. According to the figure, both Δ^L_W and Δ^H_W are positive if γ is sufficiently close to γ^*. If γ is high, the high-risk consumers are worse off from data linkage, while whether the welfare gain of the low-risk consumers is positive depends on other parameters. Figure B.3a demonstrates that the low-risk consumer welfare gain could be negative for high γ. This happens when σ is sufficiently low, $\sigma = 0.05$. According to Figure B.3b, increasing σ from 0.05 to 0.5

[29] Data linkage changes the utility of a high-risk consumer in the insurance market from $u(y - p_H(\gamma))$ to $u(y - p_H(\gamma'))$ only if this consumer does not reveal his risk type to company 0. If a high-risk consumer reveals his risk type to company 0, his utility in the insurance market changes from $u(y - p_H(\gamma))$ to $\pi_H u(y - l) + (1 - \pi_H)u(y)$. However, part of this change is already incorporated into $W^H_{\text{linked}} - W^H_{\text{independent}}$ through δV (see (B.9)). As a result, the welfare change is the same for those who reveal their type to company 0 and those who do not.



(a) Parameters: $l = y = 1$, $\sigma = 0.05$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$.



(b) Parameters: $l = y = 1$, $\sigma = 0.5$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$.

Figure B.3: *Consumer welfare gain from data linkage.* Numerical results for the model with monopolistic insurance market with utility function $u(x) = 1 - e^{-x}$. The high- (low-) risk consumer gain is in red (green); the average gain $\Delta_W = \gamma\Delta_W^L + (1-\gamma)\Delta_W^H$ is in black.

is sufficient for making the low-risk consumer and even the average consumer welfare gain positive for all γ. All these observations are confirmed in Theorem B.2. In addition, Theorem B.2 states that for sufficiently high N, Δ_W^L is positive.

Theorem B.2. *If either N or σ is sufficiently high, then the low-risk consumers are better off with data linkage. If σ is sufficiently high and $u'(0) \leq 1$, then the average consumer welfare increases with data linkage. If γ is sufficiently close to 1, then high-risk consumers are worse off with data linkage.*

Proof.

Preliminaries We start by deriving the expressions for δV_L, δV_H, $\delta\Pi_L$ and $\delta\Pi_H$.

As we can see from (B.9), δV depends on s_0^L and s_0^H only through γ'. Denote

$$\delta V' = -p_H'(\gamma')u'(y - p_H(\gamma')) \tag{B.33}$$

the derivative of δV with respect to γ'. Differentiating (B.9) with respect to s_0^L and s_0^H and using the definition of γ', (B.8), we get

$$\delta V_L = -\frac{(1-\gamma')\gamma'\delta V'}{1-s_0^L}, \quad \delta V_H = \frac{(1-\gamma')\gamma'\delta V'}{1-s_0^H}. \tag{B.34}$$

Hence, δV_L and δV_H enter the equilibrium conditions (B.14) and (B.15) through

$$\frac{\sigma - s_0^L(1-s_0^L)\delta V_L}{\sigma + s_0^H(1-s_0^H)\delta V_H} = \frac{\sigma + s_0^L(1-\gamma')\gamma'\delta V'}{\sigma + s_0^H(1-\gamma')\gamma'\delta V'}. \tag{B.35}$$

Note that $s_0^L > s_0^H$ by (B.13), and $\delta V' > 0$ because, as we discuss in Section B.2.1, $p_H(\gamma)$ decreases in γ. Hence, the ratio (B.35) is greater than 1.

Differentiating (B.10) with respect to s_0^L and s_0^H and using the definition of γ', (B.8), we get

$$\delta\Pi_L = \gamma\left(\Pi_L(1) - \Pi_L(\gamma')\right) - \gamma(1-\gamma')\left(\gamma'\Pi_L'(\gamma') + (1-\gamma')\Pi_H'(\gamma')\right), \tag{B.36}$$

$$\delta\Pi_H = (1-\gamma)\left(\Pi_H(0) - \Pi_H(\gamma')\right) + \gamma'(1-\gamma)\left(\gamma'\Pi_L'(\gamma') + (1-\gamma')\Pi_H'(\gamma')\right). \tag{B.37}$$

Using the definition (B.11), we rewrite

$$\gamma'\Pi_L'(\gamma') + (1-\gamma')\Pi_H'(\gamma') = \gamma'\left(p_L'(\gamma') - \pi_L q_L'(\gamma')\right) + (1-\gamma')p_H'(\gamma'). \tag{B.38}$$

From the analysis of the insurance market, we know that for any $\gamma > \gamma^*$, functions $q_L(\gamma)$, $p_L(\gamma)$ and $p_H(\gamma)$ solve (B.3)-(B.5). Applying the implicit function theorem to get $q_L'(\gamma)$, $p_L'(\gamma)$ and $p_H'(\gamma)$ from (B.3)-(B.5), we derive that

$$\gamma\left(p_L'(\gamma) - \pi_L q_L'(\gamma)\right) + (1-\gamma)p_H'(\gamma) = 0, \quad \forall\, \gamma > \gamma^*. \tag{B.39}$$

In particular, equality (B.39) holds for γ' because in equilibrium, γ' must be greater than γ^*. Hence, (B.36) and (B.37) can be simplified to

$$\delta\Pi_L = \gamma\left(\Pi_L(1) - \Pi_L(\gamma')\right), \quad \delta\Pi_H = (1-\gamma)\left(\Pi_H(0) - \Pi_H(\gamma')\right). \tag{B.40}$$

Limit $\sigma \to +\infty$

From (B.13), we can see that in the limit, both types have equal demand for variety 0:

$$\lim_{\sigma\to+\infty} s_0^L(\sigma) = \lim_{\sigma\to+\infty} s_0^H(\sigma) \equiv s_0, \tag{B.41}$$

and furthermore,

$$\text{if } s_0 = 0, \text{ then } \lim_{\sigma\to+\infty} \frac{s_0^L(\sigma)}{s_0^H(\sigma)} = 1; \quad \text{if } s_0 = 1, \text{ then } \lim_{\sigma\to+\infty} \frac{1-s_0^H(\sigma)}{1-s_0^L(\sigma)} = 1. \tag{B.42}$$

Using (B.41) and (B.42), we get

$$\lim_{\sigma\to+\infty}\frac{t^*(\sigma)}{\sigma}\overset{(B.14)}{=}\frac{1}{1-\frac{1-s_0}{N}},\qquad \lim_{\sigma\to+\infty}\frac{t_0^*(\sigma)}{\sigma}\big(1-s_0^L(\sigma)\big)\overset{(B.15)}{=}1. \tag{B.43}$$

Substituting (B.43) into

$$\lim_{\sigma\to+\infty}\frac{t_0^*(\sigma)}{\sigma}\big(1-s_0^L(\sigma)\big)\overset{(B.12),(B.41)}{=}(1-s_0)\lim_{\sigma\to+\infty}\frac{t^*(\sigma)}{\sigma}+\lim_{\sigma\to+\infty}\big(1-s_0^L(\sigma)\big)\ln\frac{1-s_0^L(\sigma)}{Ns_0^L(\sigma)}, \tag{B.44}$$

we get

$$1=\frac{1-s_0}{1-\frac{1-s_0}{N}}+\lim_{\sigma\to+\infty}\big(1-s_0^L(\sigma)\big)\ln\frac{1-s_0^L(\sigma)}{Ns_0^L(\sigma)}. \tag{B.45}$$

Cases $s_0=0$ and $s_0=1$ both contradict (B.45). Hence,

$$\lim_{\sigma\to+\infty}\big(1-s_0^L(\sigma)\big)\ln\frac{1-s_0^L(\sigma)}{Ns_0^L(\sigma)}=(1-s_0)\ln\frac{1-s_0}{Ns_0}, \tag{B.46}$$

and (B.45) gives an equation for $s_0\in(0,1)$. This equation has a unique solution $s_0=\frac{1}{N+1}$. Hence,

$$\lim_{\sigma\to+\infty}s_0^L(\sigma)=\lim_{\sigma\to+\infty}s_0^H(\sigma)=\frac{1}{N+1},\qquad \lim_{\sigma\to+\infty}\frac{t^*(\sigma)}{\sigma}=\frac{N+1}{N}. \tag{B.47}$$

Therefore,

$$\lim_{\sigma\to+\infty}\gamma'(\sigma)\overset{(B.8)}{=}\gamma. \tag{B.48}$$

The asymptotics (B.47) and (B.48) implies that Δ_W^L/σ and Δ_W^H/σ converge to 0 as $\sigma\to+\infty$ (see (B.25) and (B.26)) and, thus, a more refined asymptotics is needed to get the signs of Δ_W^L and Δ_W^H.

Equality (B.13) implies that

$$\lim_{\sigma\to+\infty}\left(\frac{s_0^L(\sigma)(1-s_0^H(\sigma))}{(1-s_0^L(\sigma))s_0^H(\sigma)}\sigma-\sigma\right)=\delta V(\gamma). \tag{B.49}$$

where $\delta V(\gamma)$ is δV evaluated at $\gamma'=\gamma$:

$$\delta V(\gamma)=u(y-p_H(\gamma))-\pi_H u(y-l)-(1-\pi_H)u(y). \tag{B.50}$$

Using (B.35), we rewrite equation (B.14) as

$$t^*\left(1-\frac{1-s_0^L}{N}\right)-\sigma=\frac{\sigma(1-\gamma)(1-s_0^H)(1-s_0^L)s_0^H}{N\left(\gamma(1-s_0^L)+(1-\gamma)(1-s_0^H)\right)\left(\sigma+s_0^H(1-\gamma')\gamma'\delta V'\right)}\frac{t^*}{\sigma}\left(\frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H}\sigma-\sigma\right). \tag{B.51}$$

Hence, by (B.47) and (B.49),

$$\lim_{\sigma\to+\infty}\left(t^*(\sigma)\left(1-\frac{1-s_0^L(\sigma)}{N}\right)-\sigma\right)=\frac{(1-\gamma)\delta V(\gamma)}{N(N+1)}. \tag{B.52}$$

Using (B.35) and (B.40), we rewrite equation (B.15) as

$$\sigma-(1-s_0^L)t_0^*=\frac{\gamma(1-s_0^L)^2 s_0^L\left(\Pi_L(1)-\Pi_L(\gamma')\right)+(1-\gamma)(1-s_0^L)(1-s_0^H)s_0^H\left(\Pi_H(0)-\Pi_H(\gamma')\right)}{\gamma(1-s_0^L)s_0^L+(1-\gamma)(1-s_0^H)s_0^H}+$$
$$\frac{(1-\gamma)(1-s_0^L)\left(s_0^H\right)^2}{\gamma(1-s_0^L)s_0^L+(1-\gamma)(1-s_0^H)s_0^H}\left(1+\frac{(1-s_0^H)(1-s_0^L)(1-\gamma')\gamma'\delta V'}{\sigma+s_0^H(1-\gamma')\gamma'\delta V'}\left(\frac{t_0^*}{\sigma}+\frac{\Pi_H(0)-\Pi_H(\gamma')}{\sigma}\right)\right)$$
$$\times\left(\frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H}\sigma-\sigma\right). \tag{B.53}$$

Hence, by (B.47), (B.48) and (B.49),

$$\lim_{\sigma\to+\infty}\left(\sigma-\left(1-s_0^L(\sigma)\right)t_0^*(\sigma)\right)=\frac{(1-\gamma)\delta V(\gamma)}{N+1}+\frac{N}{N+1}\left(\gamma\Pi_L(1)+(1-\gamma)\Pi_H(0)\right). \tag{B.54}$$

Rewriting equation (B.12) as

$$\frac{\sigma-(1-s_0^L)t_0^*}{1-s_0^L}+\frac{t^*\left(1-\frac{1-s_0^L}{N}\right)-\sigma}{1-\frac{1-s_0^L}{N}}=\sigma\left((N+1)s_0^L-1\right)\left(\frac{1}{(1-s_0^L)(N-1+s_0^L)}-\frac{\ln\frac{1-s_0^L}{Ns_0^L}}{(N+1)s_0^L-1}\right), \tag{B.55}$$

and using (B.47), (B.52), (B.54) and

$$\lim_{s_0^L\to\frac{1}{N+1}}\left(\frac{1}{(1-s_0^L)(N-1+s_0^L)}-\frac{\ln\frac{1-s_0^L}{Ns_0^L}}{(N+1)s_0^L-1}\right)=\frac{(N+1)(N^2+N+1)}{N^3}, \tag{B.56}$$

we get

$$\lim_{\sigma\to+\infty}\sigma\left((N+1)s_0^L(\sigma)-1\right)=\frac{N(1-\gamma)\delta V(\gamma)}{N^2+N+1}+\frac{N^3\left(\gamma\Pi_L(1)+(1-\gamma)\Pi_H(0)\right)}{(N+1)(N^2+N+1)}. \tag{B.57}$$

Finally, rewriting (B.25) and (B.26) as

$$\Delta_W^L = \sigma\left((N+1)s_0^L - 1\right)\left(\frac{1}{N(N-1+s_0^L)} - \frac{\ln\frac{(N+1)(1-s_0^L)}{N}}{(N+1)s_0^L - 1}\right) - \frac{t^*\left(1-\frac{1-s_0^L}{N}\right)-\sigma}{1-\frac{1-s_0^L}{N}}, \tag{B.58}$$

$$\Delta_W^H = \Delta_W^L + \sigma\ln\left(1 - \frac{1}{\sigma}\left(\frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H}\sigma - \sigma\right)\frac{s_0^H(1-s_0^L)}{1-s_0^H}\right) + u(y-p_H(\gamma')) - u(y-p_H(\gamma)), \tag{B.59}$$

and using (B.47), (B.48), (B.49), (B.52), (B.57) and

$$\lim_{s_0^L\to\frac{1}{N+1}}\left(\frac{1}{N(N-1+s_0^L)} - \frac{\ln\frac{(N+1)(1-s_0^L)}{N}}{(N+1)s_0^L-1}\right) = \frac{N^2+N+1}{N^3}, \quad \lim_{\sigma\to+\infty}\sigma\ln\left(1+\frac{\text{const}}{\sigma}\right) = \text{const}, \tag{B.60}$$

we get

$$\lim_{\sigma\to+\infty}\Delta_W^L(\sigma) = \frac{\gamma\Pi_L(1)+(1-\gamma)\Pi_H(0)}{N+1} > 0, \tag{B.61}$$

$$\begin{aligned}
\lim_{\sigma\to+\infty}\gamma\Delta_W^L(\sigma)+(1-\gamma)\Delta_W^H(\sigma) &= \frac{\gamma\Pi_L(1)+(1-\gamma)(\Pi_H(0)-\delta V(\gamma))}{N+1} \\
&\overset{(B.11),(B.50)}{=} \frac{\gamma\Pi_L(1)+(1-\gamma)(p_H(0)-p_H(\gamma)-u(y-p_H(\gamma))+\pi_H u(y-l)+(1-\pi_H)u(y))}{N+1} \\
&\overset{(B.2)}{=} \frac{\gamma\Pi_L(1)+(1-\gamma)(p_H(0)-p_H(\gamma)-u(y-p_H(\gamma))+u(y-p_H(0)))}{N+1} > 0.
\end{aligned} \tag{B.62}$$

The last inequality holds by assumption $u'(0)\le 1$. Indeed, since $u'(0)\le 1$ and $u(x)$ is concave, $x-u(x)$ is increasing for all $x>0$. Since $x-u(x)$ is increasing in $x>0$ and $p_H(0)>p_H(\gamma)$, $(y-p_H(0))-u(y-p_H(0)) < (y-p_H(\gamma))-u(y-p_H(\gamma))$.

Limit $N\to+\infty$

Equality (B.13) implies that the limit of the ratio $\frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H}$ is positive and finite. Hence, there are three cases: (1) both s_0^L and s_0^H are strictly between 0 and 1 at the limit, (2) both s_0^L and s_0^H converge to 1 but the limit of the ratio $\frac{1-s_0^H}{1-s_0^L}$ is positive and finite, and (3) both s_0^L and s_0^H converge to 0 but the limit of the ratio $\frac{s_0^L}{s_0^H}$ is positive and finite.

If at the limit $N\to+\infty$, both s_0^L and s_0^H are strictly between 0 and 1, then (B.15) implies that at the limit, t_0^* is finite. Equality (B.14) immediately gives that $t^*\to\sigma$. Since $t^*\to\sigma$ and the limit of s_0^L belongs to $(0,1)$, equality (B.12) implies that $t_0^*\to-\infty$. Contradiction.

If at the limit $N\to+\infty$, both s_0^L and s_0^H converge to 1 but $\frac{1-s_0^H}{1-s_0^L}$ is positive and finite, then (B.14) gives that $t^*\to\sigma$ and (B.15) implies that $t_0^*\to+\infty$. At the same time, equality (B.12) implies that $t_0^*-t^*\to-\infty$. Contradiction.

Hence,

$$\lim_{N\to+\infty} s_0^L(N) = \lim_{N\to+\infty} s_0^H(N) = 0, \quad \lim_{N\to+\infty} \gamma'(N) \overset{(B.8)}{=} \gamma, \quad \lim_{N\to+\infty} \frac{s_0^L(N)}{s_0^H(N)} \overset{(B.13)}{=} \exp\left(\frac{\delta V(\gamma)}{\sigma}\right), \tag{B.63}$$

where $\delta V(\gamma)$ is defined as in (B.50). Then, equalities (B.14) and (B.15) give

$$\lim_{N\to+\infty} t^*(N) = \sigma, \quad \lim_{N\to+\infty} t_0^*(N) = \sigma - \frac{\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right)\Pi_L(1) + (1-\gamma)\Pi_H(0)}{\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right) + (1-\gamma)}. \tag{B.64}$$

Note that the asymptotics (B.63) and (B.64) imply that Δ_W^L converges to 0 (see (B.25)) and, thus, a more refined asymptotics is needed to get the sign of Δ_W^L.

The limits (B.63), (B.64) and equality (B.12) give

$$\lim_{N\to+\infty} \ln N s_0^L(N) = \frac{\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right)\Pi_L(1) + (1-\gamma)\Pi_H(0)}{\sigma\left(\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right) + (1-\gamma)\right)}. \tag{B.65}$$

The limits (B.63) and equality (B.14) give

$$\lim_{N\to+\infty} \frac{N}{s_0^L(N)}\left(\frac{\sigma}{t^*(N)} - 1 + \frac{1}{N}\right) = \gamma + (1-\gamma)\exp\left(-\frac{\delta V(\gamma)}{\sigma}\right). \tag{B.66}$$

Rewriting (B.25) as

$$N\Delta_W^L = t^*\left(\frac{N}{s_0^L}\left(\frac{\sigma}{t^*} - 1 + \frac{1}{N}\right)s_0^L\left(\ln\frac{N}{(N+1)(1-s_0^L)} + \frac{N+1}{N}\right) - \frac{1}{N} - (N-1)\ln\left(\left(1+\frac{1}{N}\right)\left(1-\frac{N s_0^L}{N}\right)\right)\right), \tag{B.67}$$

and using (B.63), (B.64), (B.65), (B.66) and

$$\lim_{N\to+\infty} (N-1)\ln\left(\left(1+\frac{1}{N}\right)\left(1-\frac{\text{const}}{N}\right)\right) = 1 - \text{const}, \tag{B.68}$$

we get

$$\lim_{N\to+\infty} N\Delta_W^L(N) = \sigma\left(\exp\left(\frac{\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right)\Pi_L(1) + (1-\gamma)\Pi_H(0)}{\sigma\left(\gamma \exp\left(\frac{\delta V(\gamma)}{\sigma}\right) + (1-\gamma)\right)}\right) - 1\right) > 0. \tag{B.69}$$

Limit $\gamma \to 1$

By (B.8), at the limit $\gamma \to 1$, we have two possibilities: either $\gamma' \to 1$ or $s_0^L \to 1$. Suppose $s_0^L \to 1$.

Then, by (B.13), the limit of $\frac{1-s_0^H}{1-s_0^L}$ is finite. Hence, by (B.8), $\gamma' \to 1$. Thus, in any case, we have

$$\lim_{\gamma\to 1} \gamma'(\gamma) = 1. \tag{B.70}$$

Hence, the ratio (B.35) converges to 1 and, by (B.13),

$$\lim_{\gamma\to 1} \frac{s_0^L(\gamma)(1-s_0^H(\gamma))}{(1-s_0^L(\gamma))s_0^H(\gamma)} = \exp\left(\frac{\delta V(1)}{\sigma}\right), \tag{B.71}$$

where $\delta V(1)$ is δV evaluated at $\gamma' = 1$:

$$\delta V(1) = u(y - p_H(1)) - \pi_H u(y-l) - (1-\pi_H)u(y). \tag{B.72}$$

Since the ratio (B.35) converges to 1 and since, by (B.71), the limits of $\frac{1-s_0^H}{1-s_0^L}$ and $\frac{s_0^H}{s_0^L}$ are finite, equality (B.14) implies that

$$\lim_{\gamma\to 1} t^*(\gamma)\left(1 - \frac{1-s_0^L(\gamma)}{N}\right) = \sigma, \tag{B.73}$$

while equality (B.15) implies that

$$\lim_{\gamma\to 1} \left(1-s_0^L(\gamma)\right) t_0^*(\gamma) = \sigma \tag{B.74}$$

because $\delta\Pi_L \to 0$ and $\delta\Pi_H \to 0$ as $\gamma \to 1$ (see (B.40) and (B.70)).

The limits (B.73), (B.74) and equality (B.12) give

$$\lim_{\gamma\to 1} \frac{1-s_0^L(\gamma)}{1-\frac{1-s_0^L(\gamma)}{N}} + (1-s_0^L(\gamma))\ln\frac{1-s_0^L(\gamma)}{Ns_0^L(\gamma)} = 1 \quad\Rightarrow\quad \lim_{\gamma\to 1} s_0^L(\gamma) = \frac{1}{N+1}. \tag{B.75}$$

Thus, by (B.71) and (B.73), we have

$$\lim_{\gamma\to 1} s_0^H(\gamma) = \frac{1}{N\exp\left(\frac{\delta V(1)}{\sigma}\right)+1}, \quad \lim_{\gamma\to 1} t^*(\gamma) = \frac{N+1}{N}\sigma. \tag{B.76}$$

Finally, asymptotics (B.70) and (B.76) imply that the limit of (B.26) is

$$\lim_{\gamma\to 1} \Delta_W^H(\gamma) = \sigma \ln \frac{N + \exp\left(-\frac{\delta V(1)}{\sigma}\right)}{N+1} < 0. \tag{B.77}$$

□

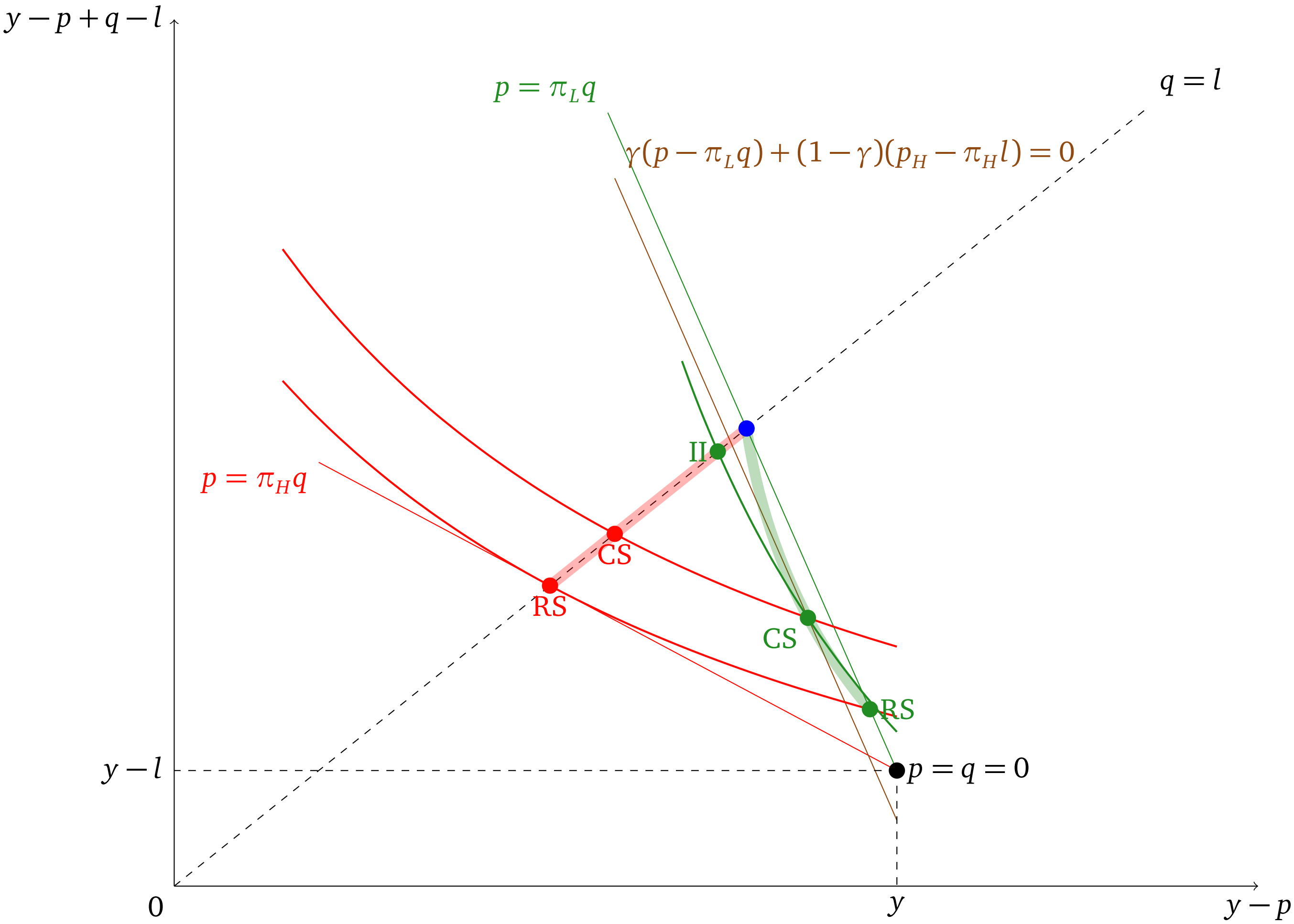


Figure B.4: Cross-subsidy equilibrium in the insurance market, drawn for the utility function $u(x) = 1-e^{-x}$. Red and green RS points correspond to the RS equilibrium contract for the high- and low-risk consumer, respectively. For sufficiently high γ, red and green CS points correspond to the cross-subsidy equilibrium and point II corresponds to the contract that the informed insurer offers to the low-risk consumer. Blue point corresponds to the competitive equilibrium when $\gamma = 1$.

B.3 Cross-Subsidy Equilibrium

Insurance market

Following Netzer and Scheuer (2014), we look for the cross-subsidy equilibrium. In this equilibrium, the high-risk consumer gets full insurance, $q_H = l$. The cover for the low-risk consumer, q_L, and the premiums, p_H and p_L, solve the following optimization problem:

$$\max_{p_L, q_L, p_H} \quad \pi_L u(y - p_L + q_L - l) + (1 - \pi_L) u(y - p_L) \tag{B.78}$$

$$u(y - p_H) = \pi_H u(y - p_L + q_L - l) + (1 - \pi_H) u(y - p_L) \tag{B.79}$$

$$\gamma(p_L - \pi_L q_L) + (1 - \gamma)(p_H - \pi_H l) = 0 \tag{B.80}$$

$$p_L - \pi_L q_L \geq 0 \tag{B.81}$$

At the optimum, the incentive compatibility constraint of the high-risk consumers (B.79) and the average break-even condition of the insurance company (B.80) bind. Constraint (B.81) ensures that low risk consumers cross-subsidize the high risk consumers and not the other way round. Whether this constraint binds depends on the proportion of the low-risk consumers γ.

If γ is below certain threshold $\hat{\gamma}$, the constraint (B.81) binds with equality, which means that there is no cross-subsidization between contracts. The cross-subsidy outcome then coincides with the RS outcome in which insurance companies break even contract-by-contract.

For γ above $\hat{\gamma}$, the constraint (B.81) is slack and so the low-risk consumers subsidize the high-risk consumers. The optimum satisfies optimality condition (B.5), which is familiar from the optimization problem of the monopolistic insurer. Equations (B.79), (B.80) and (B.5) define optimal $q_L(\gamma)$, $p_L(\gamma)$ and $p_H(\gamma)$. In Figure B.4, as γ increases from $\hat{\gamma}$ to 1, the red point CS, which corresponds to the high-risk cross-subsidy contract, moves up along the highlighted segment of the 45-degree line from red point RS, which corresponds to the RS contract for high-risk consumers, to the blue point. At the same time, the green point CS, which corresponds to the low-risk cross-subsidy contract, moves along the light green highlighted curve from the green point RS, which corresponds to the RS contract for low-risk consumers, to the blue point.

Suppose company 0 observes the consumer's type, thus becoming an informed insurer. If $\gamma \leq \hat{\gamma}$, all the analysis from the baseline model remains valid. Suppose $\gamma > \hat{\gamma}$. To avoid making losses on a contract, the informed insurer does not serve the high-risk consumers. To the low-risk consumers, the informed insurer offers a full insurance contract that they prefer to their cross-subsidy contract, thus cream-skimming low-risk consumers. In Figure B.4, the informed insurer's contract for the low-risk consumer corresponds to the green point labeled "II". The uniformed insurance companies continue to offer cross-subsidy contracts. However, because of the informed insurer's cream-skimming, the uninformed insurance companies face a population with lower proportion of low-risk consumers γ.

Product market

As in our baseline model and in contrast to the monopolistic insurance market (see Section B.2), the competitive insurance market guarantees that both types of consumers have no incentives to conceal their type by avoiding variety 0 in the product market. Thus, the demand of all consumers for each variety is given by (6).

Since all the analysis from the baseline model remains valid for $\gamma \leq \hat{\gamma}$, for the remainder of the section, we assume $\gamma > \hat{\gamma}$.

Let s_0 be the demand for variety 0. Then, the uninformed insurance companies face a population with the proportion of low-risk consumers equal to[30]

$$\gamma'(s_0) = \frac{\gamma(1-s_0)}{\gamma(1-s_0)+(1-\gamma)} < \gamma. \tag{B.82}$$

Company 0's additional profit from data linkage is

$$\Pi(\gamma') = p^I(\gamma') - \pi_L l, \tag{B.83}$$

per each low-risk consumer served in the product market. In (B.83), p^I is the premium that the informed insurer sets for the low-risk consumers, defined by the indifference condition:

$$u(y-p^I(\gamma)) = \pi_L u(y-p_L(\gamma)+q_L(\gamma)-l)+(1-\pi_L)u(y-p_L(\gamma)), \tag{B.84}$$

where $q_L(\gamma)$ and $p_L(\gamma)$ are the cross-subsidy contract for the low-risk consumer, given the share γ of the low-risk consumers in the population. If $\gamma \leq \hat{\gamma}$, then $q_L(\gamma) = q_L^{RS}$ and $p_L(\gamma) = \pi_L q_L^{RS}$.

Proposition B.3 characterizes the symmetric equilibrium in the product market.

Proposition B.3. *In equilibrium, the prices are*

$$t_0^* = \frac{\sigma}{1-s_0^*} - \gamma\left(\Pi(\gamma') - \frac{(\gamma-\gamma')(1-\gamma')\Pi'(\gamma')}{1-\gamma}\right) \tag{B.85}$$

and

$$t^* = \frac{\sigma}{1-s^*}, \tag{B.86}$$

where

$$s^* = \frac{1-s_0^*}{N} \tag{B.87}$$

[30]By defining γ' as (B.82) we implicitly assume that company 0 does not serve high-risk consumers even in the RS equilibrium, that is, when γ' is lower than $\hat{\gamma}$. This assumption is for notation simplicity and immaterial for our analysis because the informed insurer can never make positive profit on the high-risk consumers.

is the demand for each variety $n = 1, 2, \ldots, N$*, and* s_0^* *is the demand for variety 0, implicitly defined in*

$$\frac{(N+1)s_0^* - 1}{(1-s_0^*)(N-1+s_0^*)} - \ln \frac{1-s_0^*}{Ns_0^*} = \frac{\gamma}{\sigma}\left(\Pi(\gamma') - \frac{(\gamma-\gamma')(1-\gamma')\Pi'(\gamma')}{1-\gamma}\right), \tag{B.88}$$

$$\frac{\gamma^2(1-\gamma')^3}{(1-\gamma)^2\sigma}\left((\gamma-\gamma')\Pi''(\gamma') - 2\Pi'(\gamma')\right) < \frac{1}{s_0^*(1-s_0^*)^2}, \tag{B.89}$$

where $\gamma' = \gamma'(s_0^*)$ *is defined in* (B.82).

Proof. Company 0 chooses price t_0 to maximize

$$\max_{t_0} s_0(t_0 + \gamma\Pi(\gamma')) = \frac{\exp\left(-\frac{t_0}{\sigma}\right)}{\exp\left(-\frac{t_0}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right)}\left(t_0 + \gamma\Pi\left(\frac{\gamma N\exp\left(-\frac{t^*}{\sigma}\right)}{N\exp\left(-\frac{t^*}{\sigma}\right) + (1-\gamma)\exp\left(-\frac{t_0}{\sigma}\right)}\right)\right) \tag{B.90}$$

$$\text{FOC}: \frac{1}{1-s_0} - \frac{t_0 + \gamma\Pi(\gamma')}{\sigma} + \frac{(\gamma-\gamma')(1-\gamma')}{1-\gamma}\frac{\gamma\Pi'(\gamma')}{\sigma} = 0, \tag{B.91}$$

$$\text{SOC}: \frac{\gamma^2(1-\gamma')^3}{(1-\gamma)^2\sigma}\left((\gamma-\gamma')\Pi''(\gamma') - 2\Pi'(\gamma')\right) < \frac{1}{s_0(1-s_0)^2}, \tag{B.92}$$

where $s_0 = \frac{\exp\left(-\frac{t_0}{\sigma}\right)}{\exp\left(-\frac{t_0}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right)}$ and $\gamma' = \gamma'(s_0)$ is defined in (B.82).

Equation (B.91) gives (B.85), and inequality (B.92) gives (B.89).

Equations (B.86) and (B.87) follow from the same argument as in Section A.1. In particular, company n's optimization problem is the same as in Section A.1. Combining (A.6) with (B.85), (B.86) and (B.87) yields equation (B.88) for equilibrium s_0^*. □

Welfare implication of data linkage

The derivations of the welfare implications in both markets rely on Lemma B.2.

Lemma B.2. *If* $\gamma > \hat{\gamma}$*, then both* $p_H(\gamma)$ *and* $p^I(\gamma)$ *are decreasing in* γ*.*

Proof. Substituting p_H from (B.80):

$$p_H = \pi_H l - \frac{\gamma}{1-\gamma}(p_L - \pi_L q_L) \tag{B.93}$$

into equations (B.79) and (B.5) and applying the implicit function theorem to these equations, we get

$$p_L'(\gamma)=\frac{1}{H(\gamma)}\times\left\{\frac{\gamma\pi_L}{u'(y-p_L+q_L-l)}+\frac{(1-\gamma)\pi_H}{u'(y-p_H)}\right.$$
$$\left.+\gamma(p_L-\pi_Lq_L)\left(\frac{\gamma}{1-\gamma}\frac{(1-\pi_L)\pi_L}{\pi_H-\pi_L}\frac{u'(y-p_H)u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^3}+\frac{\pi_Hu''(y-p_H)}{u'(y-p_H)^2}\right)\right\}, \quad \text{(B.94)}$$

$$q_L'(\gamma)=\frac{1}{H(\gamma)}\times\left\{\left(\frac{(1-\gamma)(1-\pi_H)u'(y-p_L)}{\gamma u'(y-p_H)}+1\right)\frac{\gamma}{u'(y-p_L+q_L-l)}+\frac{(1-\gamma)\pi_H}{u'(y-p_H)}\right.$$
$$+\gamma(p_L-\pi_Lq_L)\left(\frac{\gamma}{1-\gamma}\frac{(1-\pi_L)\pi_L}{\pi_H-\pi_L}\left(\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^2}-\frac{u''(y-p_L)}{u'(y-p_L)^2}\right)\frac{u'(y-p_H)}{u'(y-p_L+q_L-l)}\right.$$
$$\left.\left.+\left(\frac{(1-\pi_H)u'(y-p_L)}{u'(y-p_L+q_L-l)}+\pi_H\right)\frac{u''(y-p_H)}{u'(y-p_H)^2}\right)\right\}, \quad \text{(B.95)}$$

where

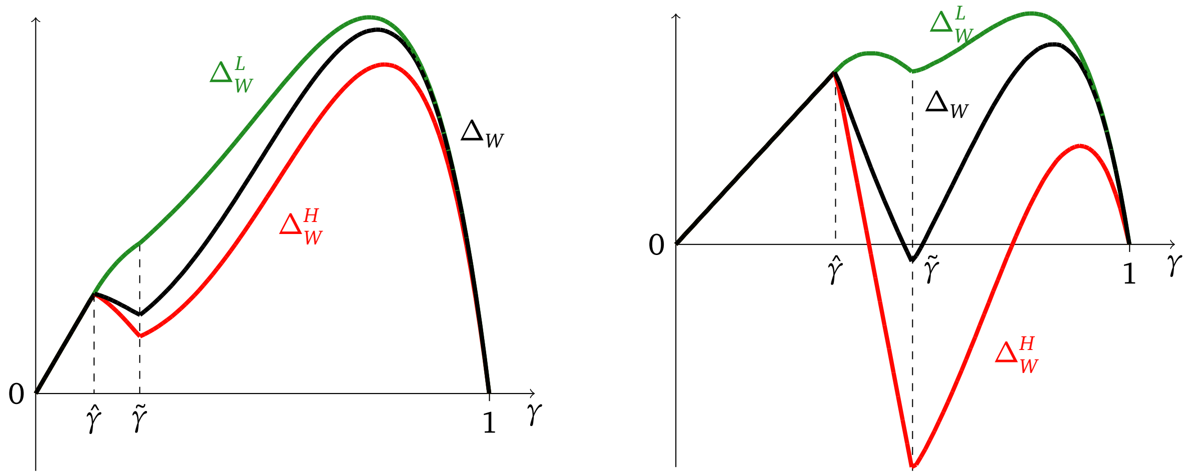
$$H(\gamma)=-\frac{\gamma^2(1-\gamma)(1-\pi_L)\pi_Lu'(y-p_L)}{\pi_H-\pi_L}\left(\pi_H\frac{u''(y-p_L)}{u'(y-p_L)^3}+(1-\pi_H)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^3}\right)$$
$$-\frac{\gamma^3(1-\pi_L)\pi_Lu'(y-p_H)}{(\pi_H-\pi_L)u'(y-p_L+q_L-l)}\left(\pi_L\frac{u''(y-p_L)}{u'(y-p_L)^2}+(1-\pi_L)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^2}\right)$$
$$-\frac{(1-\gamma)\gamma(\pi_H-\pi_L)u'(y-p_L)u''(y-p_H)}{\pi_Lu'(y-p_H)^2}\left(\frac{\gamma\pi_L}{u'(y-p_L+q_L-l)}+\frac{(1-\gamma)\pi_H}{u'(y-p_H)}\right). \quad \text{(B.96)}$$

Note that $H(\gamma)>0$ because u is increasing and concave.

Then, differentiating (B.93) and (B.84) with respect to γ and using (B.94), (B.95) and equality (B.5), we get

$$p_H'(\gamma)=-\frac{(\pi_H-\pi_L)u'(y-p_L)}{H(\gamma)u'(y-p_H)}\left\{\frac{\gamma}{u'(y-p_L)}+\frac{(1-\gamma)(1-\pi_H)}{(1-\pi_L)u'(y-p_H)}\right.$$
$$\left.-(p_L-\pi_Lq_L)\frac{\pi_L(1-\pi_L)}{(\pi_H-\pi_L)^2}\frac{\gamma^2u'(y-p_H)}{1-\gamma}\left(\pi_H\frac{u''(y-p_L)}{u'(y-p_L)^3}+(1-\pi_H)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^3}\right)\right\}, \quad \text{(B.97)}$$

$$p^{I\prime}(\gamma)=\frac{\gamma(\pi_H-\pi_L)(p_L-\pi_Lq_L)u'(y-p_L)u'(y-p_H)}{H(\gamma)u'(y-p^I)}\left\{\frac{u''(y-p_H)}{u'(y-p_H)^3}\right.$$
$$\left.+\frac{\pi_L(1-\pi_L)}{(\pi_H-\pi_L)^2}\frac{\gamma}{1-\gamma}\left(\pi_L\frac{u''(y-p_L)}{u'(y-p_L)^3}+(1-\pi_L)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^3}\right)\right\}. \quad \text{(B.98)}$$



(a) Parameters: $l = y = 3$, $\sigma = 1$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.7$.

(b) Parameters: $l = y = 1$, $\sigma = 1$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.7$.

Figure B.5: *Consumer welfare gain from data linkage.* Numerical results for the model with the cross-subsidy equilibrium in the insurance market with utility function $u(x) = 1 - e^{-x}$. The high- (low-) risk consumer gain is in red (green); the average gain $\Delta_W = \gamma\Delta_W^L + (1-\gamma)\Delta_W^H$ is in black.

Both derivatives are negative because u is increasing and concave and because $p_L - \pi_L q_L > 0$ for all $\gamma > \hat{\gamma}$. □

Armed with Lemma B.2, Proposition B.4 shows that data linkage has different implications for the consumer welfare in different markets.

In the insurance market, the welfare of the consumers is determined by the offers of the uninformed companies. Hence, as a result of data linkage, the high-risk consumers welfare in the insurance market increases by

$$\Delta_W^{I,H} = u(y - p_H(\gamma')) - u(y - p_H(\gamma)), \tag{B.99}$$

while the low-risk consumers welfare increases by

$$\Delta_W^{I,L} = u(y - p_I(\gamma')) - u(y - p_I(\gamma)). \tag{B.100}$$

Expression (B.100) follows because the expected utility of low-risk consumers from the uninformed companies' offer, $\pi_L u(y - p_L(\gamma) + q_L(\gamma) - l) + (1 - \pi_L)u(y - p_L(\gamma))$, is equal to $u(y - p^I(\gamma))$ by (B.84)).

In the product market, the consumer welfare is given in (A.16); that is, the consumer welfare expression from the baseline model remains valid. Hence, data linkage increases the consumer welfare of both types equally and by amount (A.28), which we now denote by Δ_W^P:

$$\Delta_W^P = \sigma\left(\ln\frac{1 - 1/(N+1)}{1 - s_0^*} + \frac{N}{N - 1 + 1/(N+1)} - \frac{N}{N - 1 + s_0^*}\right). \tag{B.101}$$

Proposition B.4. *In the insurance market, data linkage decreases the welfare of consumers of each risk type; that is, $\Delta_W^{I,H} < 0$ and $\Delta_W^{I,L} < 0$. In the product market, data linkage increases the consumer welfare; that is, $\Delta_W^P > 0$.*

Proof. In the insurance market, by Lemma B.2, as γ decreases, offers of the uninformed companies become worse for consumers in utility terms. Indeed, for the high-risk (low-risk) consumers, $u(y - p_H(\gamma))$ ($u(y - p_I(\gamma))$) is increasing in γ because $p_H'(\gamma) < 0$ ($p^{I'}(\gamma) < 0$). Due to cream-skimming by the informed insurer, data linkage reduces γ in the population of consumers faced by the uninformed insurer, that is, $\gamma' < \gamma$. Hence, data linkage makes both type of consumers worse off in the insurance market.

In the product market, Δ_W^P is positive if equilibrium s_0^* is higher than $1/(N+1)$, the equilibrium market share of company 0 in the absence of data linkage. By Proposition B.3, s_0^* is implicitly defined in (B.88). The right-hand side of (B.88) is positive because $\Pi'(\gamma') < 0$. The last assertion follows because by (B.83), $\Pi'(\gamma') = p^{I'}(\gamma)$, which is less than zero by Lemma B.2. The left-hand side of (B.88) is increasing in s_0^* and equal to 0 at $s_0^* = 1/(N+1)$. Hence, $s_0^* > 1/(N+1)$ as required. □

The overall consumer welfare gain from data linkage across both markets is

$$\Delta_W^L = \Delta_W^P + \Delta_W^{I,L}, \quad \Delta_W^H = \Delta_W^P + \Delta_W^{I,H} \tag{B.102}$$

for low- and high-risk consumers, respectively.

Figure B.5 depicts an example of Δ_W^L and Δ_W^H as a function of γ. If $\gamma < \hat{\gamma}$, the cross-subsidy equilibrium coincides with the RS equilibrium and, thus, consumers experience no welfare loss in the insurance market. Hence, the consumer welfare gain is the same for both types and positive. In Figure B.5, threshold $\tilde{\gamma}$ separates the region $(\hat{\gamma}, \tilde{\gamma})$ where $\gamma' < \hat{\gamma}$ from the region $(\tilde{\gamma}, 1)$ where $\gamma' > \hat{\gamma}$. Figure B.5a shows that, across the two markets, data linkage may increase the overall welfare of both low- and high-risk consumers. In contrast, Figure B.5b shows that the welfare of high-risk consumers and the average welfare may decrease in the presence of data linkage. The figures suggest that the overall welfare gain is positive for both consumer types when the stakes in the insurance market are high.

B.4 Outside Option

In this section, we assume that consumers have an outside option in the product market, that is, they might choose not to buy any product. The utility from the outside option is $\mu_{N+1}\sigma$, where μ_{N+1} is independent of other μ_n's and follows the double exponential distribution (2); thus, on average, the outside value is 0.

The presence of the outside option changes company n's demand s_n from (6) to

$$s_n = \frac{\exp\left(\frac{V-t_n}{\sigma}\right)}{1+\sum_{i=0}^{N}\exp\left(\frac{V-t_i}{\sigma}\right)}. \tag{B.103}$$

As a result of this change in the demand, the symmetric equilibrium in the product market takes a slightly different form.

Proposition B.5. *In equilibrium, the prices are*

$$t_0^* = \frac{\sigma}{1-s_0^*} - \gamma\Pi \tag{B.104}$$

and

$$t^* = \frac{\sigma}{1-s^*}, \tag{B.105}$$

where s^* is the demand for each variety $n = 1, 2, \ldots, N$, and s_0^* is the demand for variety 0, implicitly defined by the system of two equations:

$$\frac{1}{1-s^*} - \ln\frac{1-s_0^*-Ns^*}{s^*} = \frac{V}{\sigma}, \tag{B.106}$$

$$\frac{1}{1-s_0^*} + \ln s_0^* - \frac{1}{1-s^*} - \ln s^* = \frac{\gamma\Pi}{\sigma}. \tag{B.107}$$

Proof. Company 0 chooses price t_0 to maximize

$$\max_{t_0} s_0(t_0+\gamma\Pi) = \frac{\exp\left(\frac{V-t_0}{\sigma}\right)}{\exp\left(\frac{V-t_0}{\sigma}\right)+N\exp\left(\frac{V-t^*}{\sigma}\right)+1}(t_0+\gamma\Pi) \tag{B.108}$$

$$\text{FOC}: \ \exp\left(\frac{V-t_0}{\sigma}\right)+\left(N\exp\left(\frac{V-t^*}{\sigma}\right)+1\right)\left(1-\frac{t_0+\gamma\Pi}{\sigma}\right)=0 \tag{B.109}$$

SOC always holds, so that any solution t_0 to (B.109) is a local maximum.

Company $n \geq 1$ maximizes

$$\max_{t_n} s_n t_n = \frac{\exp\left(\frac{V-t_n}{\sigma}\right)}{\exp\left(\frac{V-t_n}{\sigma}\right)+\exp\left(\frac{V-t_0^*}{\sigma}\right)+(N-1)\exp\left(\frac{V-t^*}{\sigma}\right)+1}t_n \tag{B.110}$$

$$\text{FOC}: \ \exp\left(\frac{V-t_n}{\sigma}\right)+\left(\exp\left(\frac{V-t_0^*}{\sigma}\right)+(N-1)\exp\left(\frac{V-t^*}{\sigma}\right)+1\right)\left(1-\frac{t_n}{\sigma}\right)=0 \tag{B.111}$$

SOC always holds, so that any solution t_n to (B.111) is a local maximum.

Denote

$$s^* = \frac{\exp\left(\frac{V-t^*}{\sigma}\right)}{\exp\left(\frac{V-t_0^*}{\sigma}\right) + N\exp\left(\frac{V-t^*}{\sigma}\right) + 1}, \quad s_0^* = \frac{\exp\left(\frac{V-t_0^*}{\sigma}\right)}{\exp\left(\frac{V-t_0^*}{\sigma}\right) + N\exp\left(\frac{V-t^*}{\sigma}\right) + 1} \tag{B.112}$$

the equilibrium demand for companies $n = 1, \ldots, N$ and company 0, respectively. Then (B.109) implies (B.104) and (B.111) implies (B.105). Definition (B.112) implies that

$$t^* = V + \sigma \ln \frac{1 - s_0^* - Ns^*}{s^*}, \tag{B.113}$$

$$t_0^* = t^* + \sigma \ln \frac{s^*}{s_0^*}. \tag{B.114}$$

Combining (B.113) with (B.105) yields (B.106). Combining (B.114) with (B.104) and (B.105) yields (B.107).

□

Equality (B.107) is identical to equality (12) in the baseline model, and it shows that the presence of data linkage Π introduces a wedge between s_0^* and s^*.

Equality (B.106) reflects the presence of the outside option. If $V = +\infty$ (so that, effectively, there is no outside option), then we get (10). As V reduces, s^* also reduces, succumbing to the increasing attractiveness of the outside option.

Lemma B.3. *The system (B.106) and (B.107) is equivalent to*

$$s_0^* = s^*\left(\frac{1}{s^*} - \exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) - N\right), \tag{B.115}$$

$$\frac{1}{s^*\left(\exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) + N\right)} - \frac{1}{1-s^*} + \ln\left(\frac{1}{s^*} - \exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) - N\right) = \frac{\gamma\Pi}{\sigma}. \tag{B.116}$$

The solution s^* to (B.116) exists and unique on the region $(0, \bar{s}]$, where $\bar{s} \in \left(0, \frac{1}{N+1}\right)$ uniquely solves

$$\frac{1}{\bar{s}} - \exp\left(\frac{1}{1-\bar{s}} - \frac{V}{\sigma}\right) - N = 1. \tag{B.117}$$

Given the solution s^* to (B.116), s_0^* defined in (B.115) belongs to $[s^*, 1 - Ns^*)$. If $\Pi = 0$, then $s_0^* = s^* = \bar{s}$. If $\Pi > 0$, then $s^* < \bar{s}$ and $s_0^* > s^*$.

Proof. Equation (B.115) is equivalent to equation (B.106). Substituting s_0^* from (B.115) into (B.107), we get (B.116).

The left-hand side of (B.117) is decreasing in $\bar{s}$ and less than 1 at $\bar{s} = 1/(N+1)$, and it approaches $+\infty$ as $\bar{s}$ goes to 0. Hence, the solution to (B.117) exists and unique on the region $\left(0, \frac{1}{N+1}\right)$. Moreover, for all $s^* \in (0, \bar{s})$,

$$\frac{1}{s^*} - \exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) - N > 1. \tag{B.118}$$

The left-hand side of (B.116) is decreasing in s^* and equal to 0 at $s^* = \bar{s}$, and it approaches $+\infty$ as s^* goes to 0. Hence, the solution s^* to (B.116) exists and unique on the region $(0, \bar{s}]$. Moreover, $s^* = \bar{s}$ if $\Pi = 0$ and $s^* < \bar{s}$ if $\Pi > 0$.

Given $s^* \in (0, \bar{s})$, s_0^* defined in (B.115) is less than $1 - Ns^*$, and, by (B.118), is greater than s^*. If $s^* = \bar{s}$, then, by (B.115), $s_0^* = \bar{s}$. □

Proposition B.6 gives the expression for the consumer's welfare.

Proposition B.6. *In the product market, in equilibrium, the consumer welfare is*

$$W = V + \sigma \ln\left\{\exp\left(-\frac{t_0^*}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right) + \exp\left(-\frac{V}{\sigma}\right)\right\}. \tag{B.119}$$

Proof. The proof is analogous to the proof of Proposition 2. □

According to Table B.1, the results of Theorem 1, Lemma 1 and Propositions 3, 4 and 6 are fully robust, while Proposition 5 is partially robust to the introduction of the outside option.

	Π	$\Pi \to +\infty$	N	$N \to +\infty$	$\sigma \to 0$ $(V > 0)$
s_0^*	$+$	1	$-$	0	1 if $\Pi > 0$
s^*	$-$	0	$-$	0	0 if $\Pi > 0$
t_0^*	$-$	$-\infty$	$-$	$\sigma - \gamma\Pi$	0
t^*	$-$	σ	$-$	σ	0
R_0	$+$	$+\infty$	$-$	0	$\gamma\Pi$
R	$-$	0	$-$	0	0
$R_0 + NR$	$+$	$+\infty$		σ	$\gamma\Pi$
W	$+$	$+\infty$	$+$	$+\infty$	V
Δ_{R0}	$+$	$+\infty$		0	$\gamma\Pi$
Δ_{RN}	$-$			0	0
Δ_W	$+$	$+\infty$		0	0

Table B.1: Comparative statics results for the model with outside option. The rows correspond to the equilibrium quantities; the columns correspond to the parameters of interest. An entry with + (−) indicates that the row quantity increases (decreases) with respect to the column parameter.

Comparative statics with respect to Π

Since the left-hand side of (B.116) is decreasing in s^* and independent of Π, and the right-hand side of (B.116) is increasing in Π and independent of s^*, the solution s^* to (B.116) is decreasing in Π. Moreover, since the left-hand side of (B.116) approaches $+\infty$ as s^* goes to 0, the solution s^* to (B.116) is 0 if $\Pi = +\infty$.

Since the right-hand side of (B.115) is decreasing in s^* and independent of Π, and since s^* is decreasing in Π, s_0^* is increasing in Π. As $\Pi \to +\infty$, since s^* goes to 0, s_0^* goes to 1.

Since s^* is decreasing in Π, by (B.105), t^* is decreasing in Π. As $\Pi \to +\infty$, since s^* goes to 0, t^* goes to σ.

Substituting s_0^* from (B.115) and $\gamma\Pi$ from (B.116) into (B.104) yields

$$t_0^* = \frac{\sigma}{1-s^*} - \sigma \ln\left(\frac{1}{s^*} - \exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) - N\right). \tag{B.120}$$

The right-hand side of (B.120) is increasing in s^*. Then, since s^* is decreasing in Π, t_0^* is decreasing in Π. As $\Pi \to +\infty$, since s^* goes to 0, t_0^* goes to $-\infty$.

As in the baseline model, the expressions for company 0's profit and company n's profit are given in (A.13) and (A.14). Hence, similar to the baseline model, R_0 increases in Π and R decreases in Π, R_0 goes to $+\infty$ and R goes to 0 as $\Pi \to +\infty$.

Substituting s_0^* from (B.115) to (A.13), we get the expression for the joint profit as a function of s^*:

$$R_0 + NR = \frac{\sigma}{s^*\left(\exp\left(\frac{1}{1-s^*} - \frac{V}{\sigma}\right) + N\right)} - \sigma + N\frac{\sigma s^*}{1-s^*}. \tag{B.121}$$

As $\Pi \to +\infty$, since s^* goes to 0, $R_0 + NR$ goes to $+\infty$. The right-hand side of (B.121) is decreasing in s^*: its derivative with respect to s^* is

$$-\frac{\sigma\left((s_0^* - s^*)(1-s_0^*)(2-s^*-s_0^*) + (1-s_0^* - Ns^*)(s^* + (1-s_0^*)^2)\right)}{(1-s_0^*)^2(1-s^*)^2 s^*} < 0, \tag{B.122}$$

where s_0^* is defined in (B.115). Then, since s^* is decreasing in Π, $R_0 + NR$ is increasing in Π.

The consumer welfare W defined in (B.119) is increasing in Π because both prices, t_0^* and t^*, are decreasing in Π. Moreover, as $\Pi \to +\infty$, since t_0^* goes to $-\infty$ while t^* stays finite, W goes to $+\infty$.

The comparative statics of Δ_{R0}, Δ_{RN} and Δ_W with respect to Π follows from the comparative statics of R_0, R and W.

Comparative statics with respect to N

Since the left-hand side of (B.116) is decreasing in N and in s^*, the solution s^* to (B.116) is decreasing in N. Moreover, since the upper bound on s^*, $1/(N+1)$, approaches 0 as $N \to +\infty$, s^* is 0 if $N = +\infty$.

By (B.107), since s^* is decreasing in N and equal to 0 at the limit $N \to +\infty$, s_0^* is also decreasing in N equal to 0 at the limit $N \to +\infty$.

Since s_0^* and s^* are decreasing in N and go to 0 at the limit, t_0^*, t^*, R_0 and R are decreasing in N by (B.104), (B.105), (A.13) and (A.14), respectively, and their limits are $\sigma - \gamma\Pi$, σ, 0 and 0.

The consumer welfare W defined in (B.119) is increasing in N because both prices, t_0^* and t^*, are decreasing in N. Moreover, as $N \to +\infty$, since t_0^* and t^* stay finite, W goes to $+\infty$.

By (A.13), company 0's change in profit, defined in (A.18), is

$$\Delta_{R0} = \frac{\sigma s_0^*}{1-s_0^*} - \frac{\sigma \bar{s}}{1-\bar{s}}. \tag{B.123}$$

because $s_0^* = \bar{s}$ if $\Pi = 0$. Then, as $N \to +\infty$, since both s_0^* and $\bar{s}$ go to 0, Δ_{R0} goes to 0.

By (B.119), the consumer welfare gain from data linkage, defined in (A.17), is

$$\Delta_W = \sigma \ln \frac{\exp\left(-\frac{t_0^*}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right) + \exp\left(-\frac{V}{\sigma}\right)}{(N+1)\exp\left(-\frac{1}{1-\bar{s}}\right) + \exp\left(-\frac{V}{\sigma}\right)} \tag{B.124}$$

because, by (B.104) and (B.105), $t_0^* = t^* = \frac{\sigma}{1-\bar{s}}$ if $\Pi = 0$. Then, as $N \to +\infty$, since $t_0^* \to \sigma - \gamma\Pi$, $t^* \to \sigma$ and $\bar{s} \to 0$, Δ_W goes to 0.

Finally, we find the limit of $R_0 + NR$ and Δ_{NR}, defined in (A.19). Fix any $\Pi \geq 0$. As $N \to +\infty$, by (B.107), since both s_0^* and s^* go to 0, s_0^*/s^* goes to $\exp\left(\frac{\gamma\Pi}{\sigma}\right)$. Thus, by (B.106), $1/s^* - N$ goes to $\exp\left(\frac{\gamma\Pi}{\sigma}\right) + \exp\left(1 - \frac{V}{\sigma}\right)$. By (A.14), the joint profit of companies $n = 1, \ldots, N$ is

$$NR = \frac{\sigma N s^*}{1-s^*} = \frac{\sigma N}{(1/s^* - N) + N - 1}. \tag{B.125}$$

Since $1/s^* - N$ is finite at the limit, NR converges to σ. Therefore, $R_0 + NR$ converges to σ and the change in NR due to data linkage, Δ_{NR}, converges to 0.

The limit case of $\sigma \to 0$

Suppose $V > 0$.

Then, if $\Pi > 0$, by (B.116),

$$\lim_{\sigma \to 0} \sigma \left(\frac{1}{s^*(\sigma)N} - \frac{1}{1-s^*(\sigma)} + \ln\left(\frac{1}{s^*(\sigma)} - N\right)\right) = \gamma\Pi. \tag{B.126}$$

Hence, $s^* \to 0$ and

$$\lim_{\sigma \to 0} \frac{\sigma}{s^*(\sigma)N} \times \lim_{s^* \to 0}\left(1 - \frac{s^*N}{1-s^*} + s^*N\ln\left(\frac{1}{s^*} - N\right)\right) = \lim_{\sigma \to 0} \frac{\sigma}{s^*(\sigma)N} = \gamma\Pi. \tag{B.127}$$

Since $s^* \to 0$, $s_0^* \to 1$ by (B.115), and $t^* \to 0$ by (B.105). Moreover, since $\sigma/s^* \to N\gamma\Pi$ by (B.127), (B.115) implies that $\sigma/(1-s_0^*) \to \gamma\Pi$. Therefore, $t_0^* \to 0$ by (B.104).

Note that if $\Pi = 0$, $s_0^* = s^* = \bar{s}$ and both prices t_0^* and t^* converge to 0 by (B.104) and (B.105).

Since $\Pi s_0^* \to \Pi$ and $t_0^* \to 0$, $R_0 = s_0^*(t_0^* + \gamma\Pi)$ converges to $\gamma\Pi$. Since $t^* \to 0$, $R = s^* t^*$ converges to 0. Therefore, $R_0 + NR$ converges to $\gamma\Pi$.

Substituting (B.104) and (B.105) into (B.119), then using (B.107), we get

$$
\begin{aligned}
W &= V - t_0^* + \sigma \ln\left\{1 + N\exp\left(\frac{t_0^* - t^*}{\sigma}\right) + \exp\left(\frac{t_0^* - V}{\sigma}\right)\right\} \\
&= V - t_0^* + \sigma \ln\left\{1 + N\exp\left(\frac{1}{1-s_0^*} - \frac{1}{1-s^*} - \frac{\gamma\Pi}{\sigma}\right) + \exp\left(\frac{t_0^* - V}{\sigma}\right)\right\} \\
&= V - t_0^* + \sigma \ln\left\{1 + \frac{Ns^*}{s_0^*} + \exp\left(\frac{t_0^* - V}{\sigma}\right)\right\}. \qquad (B.128)
\end{aligned}
$$

Hence, since $t_0^* \to 0$ and s^*/s_0^* is finite (0 if $\Pi > 0$ and 1 if $\Pi = 0$), W converges to V.

The limiting behavior of Δ_{R0}, Δ_{RN} and Δ_W follows from the limiting behavior of R_0, R and W.