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

Claudia Herresthal† Tatiana Mayskaya‡ Arina Nikandrova§

May 5, 2024

Abstract

A merger of two companies active in seemingly unrelated markets creates data linkage: by operating in a product market, the merged company acquires an informational advantage in an insurance market where companies compete in menus of contracts. In the insurance market, the informed insurer earns rent through cream-skimming. Some of this rent is passed on to consumers in the product market. Overall, the data linkage makes consumers better off when the insurance market is competitive and, under some conditions, even when the insurance market is monopolistic. The role of competitiveness of the product market and the data-sharing requirement are discussed.

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

JEL Codes: D4, D82, G22, L22, L41, L86

*We are grateful to Rossella Argenziano, Walter Beckert, Levent Celik, Xeni Dassiou, Vitor Farinha Luz, Marina Halac, Maxim Ivanov, Chris Jenkins, Anton Kolotilin, Stephen Lewis, Noriaki Matsushima, Romans Pancs, Steve Tadelis, Tommaso Valletti, Mike Walker for insightful comments. We also thank the seminar audiences of UL Birkbeck, HSE, Workshop on Competition and Regulation in Digital Markets in UL City 2023, XXIV Yasin (April) Conference 2023, NES30 Conference 2022, Lancaster Game Theory Conference 2022, SAET 2022, Digital Economy Network Workshop 2022. Support by the German Research Foundation (DFG) through CRC TR 224 (Project B04) is gratefully acknowledged.

†University of Bonn, Germany. Email: claudia.herresthal@uni-bonn.de.

‡HSE University, Russia. Email: tmayskaya@gmail.com.

§City, University of London, UK. Email: a.nikandrova@gmail.com.

1 Introduction

The permeating expansion of tech giants has put regulating digital markets high on the agenda of competition authorities in Europe. Thus, in the European Union the Digital Markets Act (DMA) became applicable in May 2023. Similarly, the UK government is on course to give the Digital Markets Unit (DMU) 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 paper 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 allow these companies to exploit consumers. On the other hand, companies that know more about their consumers may be able to provide 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 Google's recent acquisition of Fitbit, which sparked heated debates on whether the acquisition would benefit 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 have raised serious competition concerns. Nevertheless, the European Commission undertook an in-depth investigation and cleared the transaction, 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 failed to take into account that the Fitbit health data may give Google an informational advantage in the healthcare and health insurance markets — the markets in which Google was

[1] A summary of the European Commission's decision can be found using the following URL https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2484; see European Commission (2020) for the full decision on the case.

not active in the past.[2,3] In particular, an informational advantage may enable Google to identify low-risk individuals and offer them more attractive terms. The commentators' concern 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, through selling a Fitbit device to a consumer, to acquire the consumer's health data, thereby learning that consumer's risk profile and becoming an informed insurer. 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 an insurance market, the emergence of an informed insurer does not cause market unraveling. When insurance companies are not informed about the risk profile of their insured, they screen consumers by offering a menu of contracts, in which each contract specifies both a cover provided and a premium charged. These menus are designed so that each consumer self-selects a contract that is tailored to their risk profile. 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. Indeed, the uninformed insurance companies continue serving consumers with high-risk profiles, even if the informed insurer tempts away some of their customers with low-risk profiles. Competing in menus of price-quality contracts, as opposed to competing solely in prices, is instrumental for preserving the functioning of the insurance market.

Furthermore, we show that competition in menus of contracts ensures that the emergence of an informed insurer does not harm the insured. As concerned commentators rightfully note,

[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 "Verily, Google's health-focused sister company, is getting into insurance," The Verge, Aug 25, 2020 which can be found using the following URL https://www.theverge.com/2020/8/25/21401124/alphabet-verily-insurance-coefficient-stop-loss).

the superior information allows the informed insurer to cream-skim some low-risk consumers, thereby increasing the riskiness of the consumer pool for the uninformed insurers. However, since competition among uninformed insurers leads them to break even contract-by-contract in equilibrium, such adverse change in the consumer pool does not trigger these companies to change the menus of contracts they offer. Consequently, high-risk consumers are still able to take out insurance on the same terms as in the absence of the informed insurer, and the informed insurer is forced to offer low-risk consumers a contract that is at least as attractive as the one offered by the uninformed insurers.[4]

Despite not harming consumers, the informed insurer reaps additional profit from its superior information. This additional profit comes from the increased efficiency of the offered insurance contracts. Indeed, even a perfectly competitive insurance market is inefficient because, to prevent the high-risk consumers from choosing insurance contracts that are designed for the low-risk consumers, the uninformed insurance companies degrade the contracts designed for the low-risk consumers. The informed insurer is not subject to the same constraint — information allows the insurer to offer each consumer a tailor-made efficient contract. Our conclusion that the emergence of an informed insurer improves efficiency of the insurance market echoes the opinion of Pierre Régibeau, who was the EC Chief Competition Economist at the time of the Google/Fitbit merger decision. In his policy column, Régibeau (2021) 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.

Finally, 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. Thus, the key channel through which the consumers reap the efficiency benefit of data linkage is the emergence of a more aggressive competitor in the product market. Noteworthy, the effect of the more aggressive pricing by the merged entity is not diminished by consumer strategic behavior. Neither type of consumer has an incentive to avoid the merged

[4] Under alternative equilibrium selection, insurers may no longer break even contract-by-contract, in which case adverse change in the consumer pool interferes with the disciplinary role of competition and consumers may become worse off with the informed insurer. We elaborate more on this idea in Section 7.2.

entity in the product market, because revealing their risk type does not worsen available offers in the insurance market.

Consumer benefit from data linkage is low when the product market is already highly competitive prior to the merger. While a high number of competitors and a low degree of product differentiation both nullify consumers' gain from data linkage, they do so for different reasons. When faced with numerous competitors, the merged entity finds it too costly 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 consumers.

Additional profits in the insurance market mean that a company active in both markets has an incentive and the ability to expand its market share and, thus, may become dominant in the product market. In our model, such dominance unambiguously benefits consumers because the company achieves it through low prices. In reality, in the long run, the merged entity's aggressive pricing may drive its competitors out of the product market. If the product market is not contestable due to, for example, high barriers to entry or high technology development costs, then the exit of rivals would allow the merged entity to increase its price, to the detriment of consumers. Thus, competition authorities should weigh the consumer benefits from the merger against the danger of market monopolization in the long run.

A frequently discussed remedy for the data-induced increase in market power is the *data-sharing remedy*. We find that 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, the data-sharing remedy may hurt consumers.

The channel through which consumers reap efficiency benefits of data linkage — the merged entity becoming a more aggressive competitor in the product market — is so powerful that consumers may still benefit from data linkage even if the insurance market is not competitive. To demonstrate this point, we consider a monopolistic insurance market. Equipping a monopolistic insurer with superior information increases its ability to extract rent from the high-risk consumers. Hence, with the monopolistic insurer, data linkage introduces a non-trivial trade-off between improved efficiency and consumer exploitation. This trade-off implies that data linkage may have a negative effect on consumer welfare. Nevertheless, a more intense competition passes on sufficient efficiency gains to consumers, so that the overall effect of data linkage on consumer welfare is positive under certain conditions, for example, when the share of low-risk consumers is sufficiently low.

Our model applies beyond the Google/Fitbit example to any pair of markets in which data collected in one market is relevant for operating in another market. We refer to the former as the data collection market and to the latter as the data application market. The critical feature for our results is that in the data application market, firms offer menus of price-quality contracts. Markets like that are ubiquitous. For example, besides the insurance market, Google and Apple may enter as informed providers into the market for consumer credit[5] or the eSIM market.[6]

The novelty of our paper is to demonstrate that data linkage does not harm consumers when companies in the data application market compete in menus of price-quality contracts. Competition in menus enables uninformed companies to screen consumers, thereby limiting the informed company's ability to exploit consumers using its superior information. Furthermore, we contribute to the understanding of how the efficiency gains in the data application market are passed on to consumers in the data collection market. In particular, our modeling approach allows us to study how the number of competitors in the data collection market affects consumer benefit from data linkage.

[5]A recent trend of Big Tech companies entering into the credit market has drawn attention of the regulators — see, for example, section 5 in Financial Conduct Authority (2022). For example, after acquiring Credit Kudos, Apple launched a new *Buy Now, Pay Later* service to tailor financial contracts to the needs of their customers about whom they know a lot through monitoring their mobile device usage.

[6]Data collected through monitoring the usage of mobile devices may allow the providers of mobile operating systems to offer personalized mobile connectivity contracts in the nascent eSIMs market. The concerns that Apple and Google potential entry into eSIMs market may hurt consumers are expressed in paragraphs 4.53 and 4.54 in Ofcom (2022).

As a practical matter, our findings suggest that, to determine whether a merger that creates data linkage is indeed beneficial to consumers, competition authorities must undertake a comprehensive in-the-round assessment of both data application and data collection markets. Consumer harm in the data application market (i.e., the insurance market in our model) is unlikely to arise when this market is vertically differentiated and several rivals exert competitive constraint on each other. In the data collection market, the consumer benefit from the merger is lower when there are many competitors in the market, or when the market features a low degree of product differentiation, or when the efficiency gain in the data application market is low. Moreover, in the long run, consumers may even be harmed by the merger through monopolization of the data collection market if the potential gains for the merged entity in the data application market are high and the data collection market is not contestable.

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 the danger of product market monopolization, banning below-cost pricing and the data-sharing remedy. Section 6 considers a monopolistic insurance market. Section 7 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 (Scott 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 the 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 are pro-competitive when they increase markups, thus inducing firms to compete more fiercely to attract more consumers; and data are anti-competitive when they enable firms to extract consumer surplus in a more efficient manner. In our baseline model with a competitive insurance market, data have a pro-competitive effect, while in the model with a monopolistic insurer, data have 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 the 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); Elliott et al. (2021); Hagiu and Wright (2022); Ichihashi (2020, 2021)). Our paper stands alongside recent papers such as Condorelli and Padilla (2021), Argenziano and Bonatti (2021) and Cong and Matsushima (2023), which model data linkage between seemingly unrelated markets. In contrast to our focus, Condorelli and Padilla (2021) focus on entry deterrence, while Argenziano and Bonatti (2021) and Cong and Matsushima (2023) focus on privacy management.

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, which is relevant in the other market. Since we directly model the 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.

As 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), 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 as 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) modeled as in Rothschild and Stiglitz (1976) (RS); the other market is a differentiated product Bertrand market in which consumers have random utility (e.g., market for gadgets). Both markets serve the same unit mass of consumers. Consumers are privately informed about i) their taste for each variety offered in the product market and ii) their risk of loss, against which they can take out insurance in the insurance market. Each consumer first buys a product in the product market and then chooses a contract in the insurance market.

Figure 1 provides a schematic depiction of the model. While both the product and insurance markets have a number of competitors, exactly one company, referred to as company 0, operates in both markets. We contrast a scenario in which the markets are independent (no data linkage) with a scenario in which the markets are linked through the data flow (data linkage). When the markets are linked, whenever company 0 serves a consumer in the product market, it learns the consumer's risk of loss and can then offer a personalized contract in the insurance market.[7] The data linkage may arise as a result of a merger between a company

[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, company 0 does not encounter the same consumer in both markets, but uses the information on the consumers it serves in the product

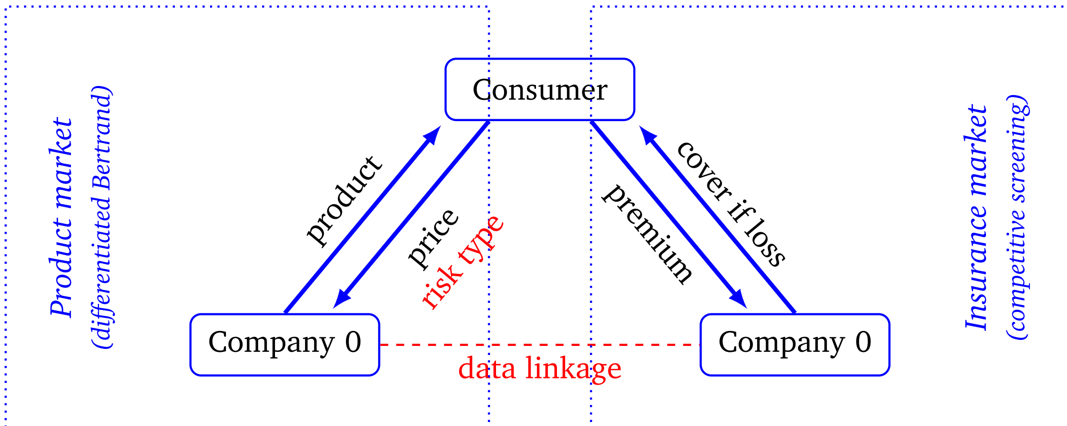


Figure 1: Sketch of the model.

in the insurance market and a company in the product market. Then, the effect of the data linkage can be interpreted as the effect of the merger.

Product market

There are $N+1$ companies in the product market, where $N \geq 1$. Each company $n = 0, 1, \ldots, N$ produces a single variety for which it sets a price t_n. The demand faced by 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.

Consumers have random utility. A consumer's 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.[8] It is assumed that μ_n are i.i.d. and follow the

market to better predict the risk of loss for its consumers in the insurance market. Under this alternative assumption, consumers would never want to hide their risk type by avoiding company 0's product. In our baseline model, there is no material difference between the two assumptions. However, in Section 6, where we consider the monopolistic insurance market, the two assumptions could lead to different results.

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

double exponential distribution:

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

Insurance market

In the insurance market, each 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 $x = y - l$. All consumers are risk-averse: given income x, the utility of the consumer is an increasing and concave function $u(x)$.

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 risk type of each consumer is independent of her taste parameter in the product market. The share of low-risk type consumers is $\gamma \in (0,1)$.

There are at least three companies, one of which is company 0. Consumer risk types are unobserved by companies, with one exception. In the scenario with data linkage, company 0 is informed about the risk type of those consumers that it served in the product market.[9] Each company offers a menu of insurance contracts to each consumer. All offers are made simultaneously. After the offers are made, each consumer chooses a contract.

An insurance contract is characterized by premium p and cover q. Each 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 pays nothing but does not receive compensation in the case of loss.

For a company, 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.

[9] We require at least three companies to ensure that in the scenario with data linkage, at least two companies that are uninformed about any consumer risk type compete with each other.

3 Equilibrium

In the scenario with data linkage, company 0 is able to use information from the product market in the insurance market. Therefore, we study the insurance market first. We consider the menu of contracts offered to a consumer in equilibrium under two alternative assumptions: when company 0 is uninformed about the consumer's risk type and when company 0 is informed. Next, we characterize 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 about the consumer type in the insurance market in the data linkage scenario.

3.1 Insurance Market

Company 0 is uninformed

If company 0 is uninformed about the consumer's type, then the equilibrium is as in the classical competitive screening model studied in RS. In the RS model, the unique pure strategy equilibrium is of the separating type.[10] We refer to this separating pure strategy equilibrium as the RS equilibrium. It is well known, however, that 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.[11] 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.[12,13]

In the RS equilibrium, 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

[10] There is also a growing literature on nonexclusive contracts whereby a pooling contract may emerge in equilibrium (Attar et al. (2011, 2022); Huang and Sandmann (2022)).

[11] 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.

[12] 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).

[13] In Section 7.2, we discuss the robustness of our results to an alternative equilibrium selection for high γ — a cross-subsidy equilibrium.

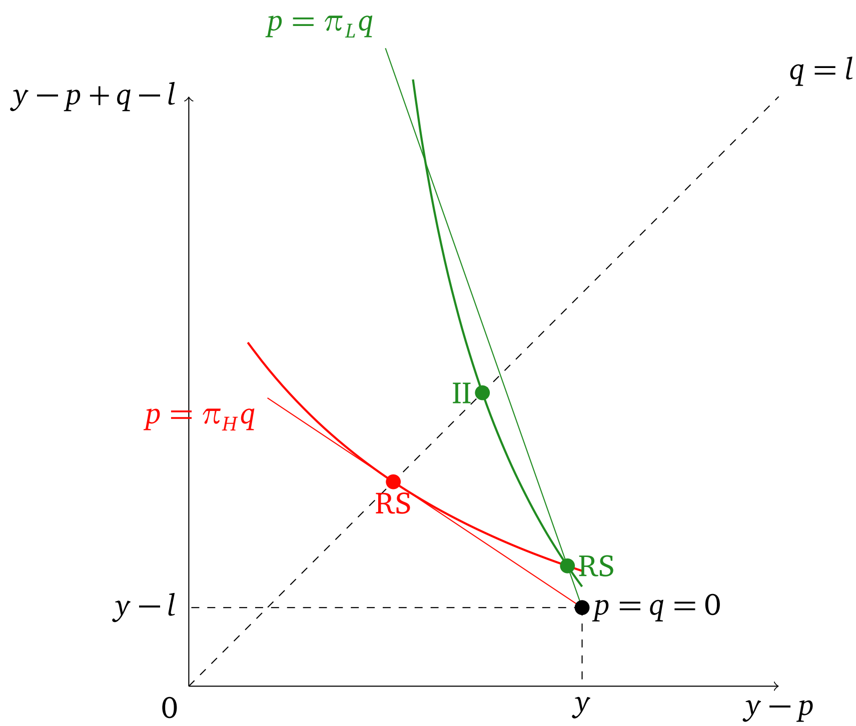


Figure 2: Equilibrium in the insurance market, drawn for the utility function $u(x)=1-e^{-x}$. Red and green RS points correspond to the RS equilibrium contract for the high- and low-risk consumer, respectively. Point II corresponds to the contract that the informed insurer offers to the low-risk consumer.

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 degrade the contract intended for the low-risk type from the efficient full insurance contract: 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 that company 0 observes the consumer's type, thus becoming an informed insurer.

Since the competition between the uninformed insurance companies leads them to break even contract-by-contract, these companies 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 for high-risk consumers pushes the insurer's profit to zero.[14]

However, the informed insurer can always *cream-skim* low-risk consumers. In particular, the informed insurer offers a full insurance contract $(p, q) = (p^I, l)$ with premium p^I to the low-risk consumer, which leaves this type of consumer just indifferent between this contract and the partial insurance contract offered 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

[14]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.

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).

The informed insurer's profit Π on each low-risk consumer is central to our analysis because it drives company 0's incentives in the product market, as we explain in Section 3.2. Proposition 1 shows how Π varies with the primitives of the insurance market.

Proposition 1. *Π does not depend on γ. Π increases in $\pi_H \in (\pi_L, 1)$. If consumers have CARA utility $u(x) = \frac{1-\exp(-\lambda x)}{\lambda}$, then Π increases in absolute risk-aversion parameter $\lambda \in (0, +\infty)$.*

The informed insurer's profit Π does not depend on the share γ of low-risk consumers in the population because the RS equilibrium does not depend on γ.

The intuition for the comparative statics with respect to π_H is as follows. For a given π_L, as π_H increases, the difference between the risk types increases. As a result, it becomes more attractive for high-risk consumers to mimic low-risk consumers. To prevent mimicry, the uninformed insurers have to degrade the contracts of the low-risk consumers more, which increases the efficiency loss due to asymmetric information. Therefore, the informed insurer can make a larger profit Π from correcting this loss.[15]

For the second part of Proposition 1, we take a specific utility function to have a single parameter that captures consumers' risk aversion. As risk aversion increases, consumers value full insurance more. Hence, the disutility of the low-risk consumer from receiving a partial

[15]In contrast, a change in π_L (for a given π_H) has an ambiguous effect on Π. On the one hand, a decrease in π_L increases the difference between the risk types, and so, Π should increase. On the other hand, as the low-risk consumer is less likely to suffer a loss, the efficiency loss associated with a partial instead of a full insurance contract decreases — in the extreme case, when $\pi_L = 0$, there is no efficiency loss at all and Π takes the lowest possible value — $\Pi = 0$. Hence, for π_L close to 0, Π increases as π_L increases.

instead of a full insurance contract is larger, which implies a higher efficiency loss due to information asymmetry. Then, as before, the informed insurer can capture a higher profit from correcting this loss.

3.2 Product Market

In the product market, each consumer chooses the variety that yields the highest utility given (1), with and without data linkage. Indeed, with data linkage, each individual consumer has no incentive to conceal her risk type by avoiding variety 0 in the product market because the informed insurer's offer leaves her no better or worse off than the best offer made by the uninformed insurers.

Utility (1) and double exponential distribution of a random taste parameter, (2), gives rise to logit demand:[16]

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

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 additional profit from the insurance market. Each low-risk consumer that company 0 serves in the product market brings company 0 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. 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, company 0's profit (7) does not contain the term $\gamma \Pi$. Hence, for ease of notation, we refer to

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

the model without data linkage as the model with $\Pi = 0$.

We are looking for a symmetric equilibrium in the product market, in which each company with $n = 1, 2, \ldots, N$ sets the same price t^* and company 0 sets price t_0^*. Proposition 2 solves for the equilibrium prices and market shares.

Proposition 2. *In equilibrium, in the product market, the prices are*

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

and

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

where the demand s^* for each variety $n = 1, 2, \ldots, N$ and the demand s_0^* for variety 0 are jointly defined in

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

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

After substituting (10) into (11), the solution $s_0^* \in (0, 1)$ to (11) always exists and is unique.

4 Results

Next, we discuss our main findings. Table A.1 in Appendix A.4 comprehensively summarizes all the results of our model.

Market consequences of data linkage

Market consequences of data linkage in the insurance market are summarized in Proposition 3, which follows directly from the discussion in Section 3.1.

Proposition 3. *In the insurance market, data linkage changes the contract that company 0 offers to the low-risk consumers it serves in the product market: this contract offers full insurance at a premium that makes the low-risk consumer just indifferent between this contract and the partial insurance contract that the low-risk consumer gets without data linkage. All other contracts* *remain unchanged.*

In the product market, the effect of data linkage directly follows from Proposition 2. 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 ($\Pi > 0$) incentivizes company 0 to compete more aggressively for consumers. As profit expression (7) shows, in the product market, $\gamma\Pi$ plays the role of a per-consumer subsidy to company 0, and this subsidy leads company 0 to lower its price, all else equal (see (8)). In response to more aggressive pricing by company 0, the other companies also lower their prices t^*, but not by as much as company 0, resulting in reduced market shares. Formally, equation (11) shows that the presence of data linkage Π introduces a wedge between s_0^* and s^*, and this wedge increases with Π.

Proposition 1 shows how Π depends on the primitives of the insurance market, such as the probability with which each high-risk consumer suffers the loss and consumers' risk aversion. Since the size of Π can be traced back to the primitives of the insurance market, the product market comparative statics with respect to Π are essentially the comparative statics with respect to the primitives of the insurance market.

In general, all the effects of data linkage on the product market intensify when Π 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.

Proposition 4. *In the product market, data linkage lowers prices t_0^* and t^*, increases the market share s_0^* of company 0, and decreases the market share s^* of each other company. Moreover, the effects of data linkage are stronger at higher Π: prices t_0^* and t^* decrease with Π; company 0's market share s_0^* increases with Π, while market shares of other companies s^* decrease with Π; at the limit $\Pi \to +\infty$, $s_0^* \to 1$ and $s^* \to 0$.*

Welfare implications of data linkage

Our main result (stated in Theorem 1) is that consumers benefit from data linkage.

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

In the product market, consumer welfare is defined as the expected utility from the best offered product; that is, $W = \mathbb{E}\left[\max_n V_n\right]$. Lemma 1 derives this welfare.

Lemma 1. *In the product market, in equilibrium, consumer welfare is*

$$W = V + \sigma \ln\left\{\exp\left(-\frac{t_0^*}{\sigma}\right) + N\exp\left(-\frac{t^*}{\sigma}\right)\right\}. \tag{12}$$

As expected, consumer welfare in the product market decreases with prices t^* and t_0^*. Since data linkage intensifies competition and, by Proposition 4, reduces the prices in the product market, it benefits consumers.[17]

The consumers are not the only ones to benefit from data linkage. According to Theorem 1, all companies also jointly gain from data linkage; that is, the total profit of all companies increases. Concurrent increases in consumer and producer welfare are a manifestation of the increase in efficiency that data linkage brings to the insurance market: the knowledge of the consumer type allows company 0 to offer a more efficient insurance contract to low-risk consumers.

Although data linkage increases the total profit of all companies, Theorem 1 shows that company 0 is the only company that benefits from data linkage. Within the insurance market, the additional surplus accrues solely to company 0 and is equal to $s_0^*\gamma\Pi$; other companies neither gain nor lose because they break even on each contract independently of whether company 0 is informed. In the product market, through lower prices, company 0 shares the efficiency gain with the consumers, at the same time lowering the competitors' profits.[18]

As with prices and markets shares in Proposition 4, the welfare effects of data linkage intensify when Π increases. Remarkably, consumers benefit from data linkage even when Π tends

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

[18]The difference in product market prices t_0^* and t^* distorts consumers' 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 $s_0^*\gamma\Pi$. However, in the setting with the outside option, which we consider in Appendix B.4, it is not at all clear that the total gain from data linkage is lower than $s_0^*\gamma\Pi$ because, in addition to distorting consumers' choice, data linkage increases market demand.

to infinity and company 0 captures the entire product market. The apparent monopolization of the product market by company 0 does not harm consumers because other product companies continue to exert a competitive constraint: company 0 sets its price very low to prevent competitors from winning consumers.[19]

Theorem 1. *Data linkage strictly increases consumer welfare in the product market and leaves consumer welfare unchanged in the insurance market.*

Data linkage increases the total profit of all companies. However, only the profit of company 0 increases; the profit of other companies in the insurance market does not change; the profit of other companies in the product market decreases.

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 following section analyzes how the consequences of data linkage change as the product market becomes more competitive. We consider two different ways of increasing competitiveness: increasing the number of varieties and decreasing the degree of taste heterogeneity.

The effect of an increase in the number of varieties N is standard: prices, market shares and profits decrease, and consumer welfare increases.[20] 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 an increase in competition in the product market lessens and eventually nullifies the welfare effects of data linkage.

Proposition 5. *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* $N \to +\infty$.

The intuition behind Proposition 5 is that the decrease in the market share of company 0 implies that fewer consumers reveal their risk type to company 0, and, thus, the efficiency gain

[19]For further discussion of the monopolization concern, see Section 5.1.
[20]For the proof, see Appendix A.4.

from data linkage is reduced. In other words, competition in the product market limits company 0's ability to collect data, thus dissipating the efficiency gain. Lower efficiency gain from data linkage implies that company 0 has less gain to pass on to consumers, which decreases consumer welfare gain from data linkage.

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 the product market prices.

Another parameter that affects market competitiveness is the degree of taste heterogeneity, or product differentiation, σ. As σ decreases, consumers in the product market become more price-sensitive, which decreases the market power of each company in the product market.

If $\sigma = 0$, then each consumer views all varieties as equivalent. Hence, the product market behaves like the homogeneous-product Bertrand market. Without data linkage, the 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 marginal cost by an arbitrary small amount, thus capturing the entire market. Hence, if $\sigma = 0$, data linkage does not change product market prices, and, thus, company 0 reaps all the insurance market efficiency gain without passing any of it onto consumers.

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, $\gamma \Pi$.[21] 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, like all the other companies in the product market, can exploit

[21] 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.

consumers' reduced price sensitivity by raising its price. Therefore, without data linkage, as σ increases, company 0 and all other companies raise their prices, while maintaining an equal market share. In contrast, with data linkage, as σ increases, for company 0, the incentive to raise the price to exploit the increased loyalty of its own consumers clashes with the incentive to lower the price to win consumers from rivals — after all, each consumer that company 0 serves in the product market brings in additional revenue in the insurance market. At lower levels of σ, the incentive to lower the price is strong because winning additional consumers is relatively easy. However, as σ increases, the incentive to lower the price weakens because it becomes harder to attract the competitors' consumers. Eventually, at higher levels of σ, the incentive to increase the price to exploit the loyalty of the existing consumers gains an upper hand. Thus, the price of company 0 decreases for low σ and increases for high σ.

In the presence of data linkage, despite non-monotonicity of company 0's price, company 0's 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, company 0 starts behaving like all other companies and captures $1/(N+1)$ of all consumers.[22]

In the presence of data linkage, as company 0's product market share decreases with σ, the efficiency gain that emerges as a result of data linkage also decreases, which, in turn, drives down company 0's gain from data linkage. In contrast, non-monotonicity of company 0's price suggests that how companies' profits and consumer welfare change due to data linkage varies with σ in a non-monotone way.

Proposition 6. *Suppose $\sigma = 0$. Then, $t_0^* = t^* = 0$, with and without data linkage. While $s_0^* = s^* = 1/(N+1)$ without data linkage, $s_0^* = 1$ and $s^* = 0$ with data linkage. Since with data linkage $s_0^* = 1$, the efficiency gain in the insurance market, $s_0^* \gamma \Pi$, is at its maximum. Company 0 captures the entire efficiency gain, while the profit of other companies in the product market does not change as a result of data linkage. Moreover, consumers do not gain from data linkage.*

As σ increases: without data linkage, $t_0^* = t^*$ increases while $s_0^* = s^* = 1/(N+1)$ remains unchanged; with data linkage, s_0^* decreases from 1 to $1/(N+1)$, s^* increases from 0 to $1/(N+1)$,

[22] While company 0 captures the entire market when $\sigma = 0$, it cannot do so for positive σ, even by setting a very low price, because there is always a positive (but vanishingly small as $\sigma \to 0$) mass of consumers who strongly dislike company 0's variety.

t^* increases, while t_0^* first decreases and then increases. Company 0's gain from data linkage decreases with σ. In the product market, the profit loss of other companies due to data linkage first increases and then decreases in σ (assuming $N \geq 3$). The consumer welfare gain from data linkage first increases and then decreases in σ (assuming $N \geq 2$).

One takeaway 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 from data linkage 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 from data linkage and σ is not monotone, as we show in Proposition 6. Yet this non-monotonicity does not necessarily refute the main takeaway that the consumer welfare gain decreases as the product market becomes more competitive. An increase in σ captures more than a decrease in the competitiveness of the product market: higher σ makes more extreme taste realizations more likely, thus directly increasing consumer welfare.

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 consumers.[23]

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,

[23] For example, the UK Government expressed such concerns in para 15 in Competition and Markets Authority (2021).

according to our model, despite the increased presence of company 0 in each market, data linkage benefits 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, the 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 suffering 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 were to compete 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 Proposition 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 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 the Google/Fitbit merger, our results imply that the merger may indeed lead to Fitbit dominance in the market for wearables.[24] Commentators agree that, while the

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

market for wearables is rapidly expanding, the available gains in the healthcare and insurance markets are so large that they dwarf the device profits.[25] Hence, the case of $\Pi \to +\infty$ is particularly relevant for the 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 become available.[26]

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 on prices or on 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 the extent to which 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 such remedy is prohibition of below-cost pricing.

The policy of prohibiting below-cost pricing is familiar from the traditional competition policy frameworks. For example, European Commission (2009) in its enforcement priorities guidance on abusive exclusionary conduct by dominant firms, while acknowledging that consumers benefit from low prices, deems pricing below one's own marginal cost anti-competitive.[27] The

[25] See, for example, Bourreau et al. (2020) or "Global Smartwatch Market: Apple 34%, Huawei 8%, Samsung 8%, Fitbit 4.2%," Forbes, May 27, 2021 which can be found using the following URL https://www.forbes.com/sites/johnkoetsier/2021/05/27/global-smartwatch-market-apple-34-huawei-8-samsung-8-fitbit-42/?sh=31556af266c7.

[26] 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 "Smartwatch Market Grows 13% YoY in Q1 2022; Apple Stays First, Samsung Solidifies Second Place," Counterpoint, May 31, 2022 which can be found using the following URL https://www.counterpointresearch.com/smartwatch-market-grows-13-yoy-q1-2022-apple-stays-first-samsung-solidifies-second-place/).

[27] 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."

concern of competition authorities hinges on its focus on a single market, where below-cost pricing is associated with negative profit and, thus, 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, 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 a negative price when, for example, σ is positive but sufficiently close to 0 (see Proposition 6).[28]

Our model unambiguously predicts that banning below-cost pricing would not benefit consumers. In contrast to the single-market reasoning of traditional competition policy, in our model with two linked markets, 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, the 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.[29]

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, with below-cost pricing, company 0 may force its competitors to 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 sacrificing profit in the short run. Since company 0's

[28]In Appendix A.4, 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).

[29]Traditionally, competition policy relies on the so-called as-efficient-competitor test ("the AECT") to detect anti-competitive low pricing conduct. For example, European Commission (2009) extensively references the application of the AECT in its enforcement priorities guidance on abusive exclusionary conduct by dominant firms, while in the UK, the role of the AECT has recently been hotly debated in the Royal Mail v. Ofcom case at Competition Appeal Tribunal (2019), Competition Appeal Tribunal (2021) and, finally, at the Supreme Court (see Ofcom press release from June 17, 2022 which can be found using the following URL https://www.ofcom.org.uk/news-centre/2022/supreme-court-rejects-royal-mail-appeal-against-ofcom-fine). Our discussion implies that in markets linked by data flows, the AECT is not a reliable indicator of consumer harm.

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, 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 to mitigate the monopolization concern that is frequently discussed in policy circles is the data-sharing remedy. To apply 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 the 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 consumer welfare because it does not affect prices. However, in the insurance market, data linkage increases 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 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.[30]

In Appendix B.1, we show that whether the data-sharing remedy benefits 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 Monopolistic Insurance Market

Our model provides a framework for exploring the efficiency vs. consumer exploitation trade-off that is frequently discussed in relation to the extensive data collection by tech giants. Information about a consumer's risk type 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 this type of consumer exploitation. In this section, we relax the assumption of perfect competition in the insurance market. In particular, we look at an extreme case in which 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 a consumer's risk type, 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 not know the consumer's risk type, it might have to leave the high-risk consumer some rent to prevent her 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.

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

While information on risk types might allow the monopolistic insurer to exploit high-risk consumers, 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, company 0 now 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, 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.[31]

Across both markets, the overall welfare consequences for consumers from 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 offers only one contract, targeted at the high-risk consumers, thus extracting all rents from these consumers. Hence, in this case, data linkage does not affect the insurance contract offered to high-risk consumers, and, thus, all the results from our baseline model remain valid. In particular, both consumer types benefit from data linkage via reduced prices in the product market.

If the share of low-risk consumers is high, data linkage does introduce high-risk consumer

[31] The anti-competitive effect is muted under the alternative assumption that company 0, instead of encountering the same consumer in both markets, uses the information on the consumers it serves in the product market to better predict the risk of loss for its consumers in the insurance market (see footnote 7).

exploitation in the insurance market, and, hence, the overall effect from data linkage on consumer 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 high-risk consumers from concealing their type from company 0, which weakens the anti-competitive effect in the product market.

We also prove that low-risk consumers benefit from data linkage when the product market is sufficiently competitive, as measured by the number of companies N. Intuitively, when 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, high-risk — and sometimes even low-risk — consumers are made worse off by 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 provides. Hence, in the product market, the pro-competitive effect of data linkage is weak. At the same time, high-risk consumers have a lot to lose from revealing their type to company 0 because when hiding among numerous low-risk consumers, high-risk consumers get high information rent. Thus, in the product market, the anti-competitive effect of data linkage is strong.

7 Discussion and Extensions

7.1 Two Potentially Informed Insurers

In our baseline model, company 0 is the only company that operates in both markets. However, nowadays, several big tech competitors control 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 company 1 are non-overlapping. Constrained by their uninformed competitors, both companies offer 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 intensifies competition further, which increases consumers' 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.

7.2 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.[32] Formally, Miyazaki-Wilson contracts maximize the expected utility of the low-risk consumer subject to three con-

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

straints: the usual incentive-compatibility constraint for the high-risk consumer; the constraint ensuring that the insurance companies break even, on average, across the contracts for both the low- and high-risk consumers; and the constraint that prohibits cross-subsidization from high- to low-risk consumers. For sufficiently low γ, Miyazaki-Wilson contracts coincide with the RS contracts. For sufficiently high γ, Miyazaki-Wilson contracts involve cross-subsidization from low- to 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 low-risk consumers directly benefits high-risk consumers, making them better off. A positive subsidy also benefits low-risk consumers, albeit indirectly. Intuitively, the cross-subsidy relaxes the incentive-compatibility constraint of high-risk consumers, making them less willing to pretend to be low-risk. The relaxation of the constraint allows the companies to design better offers for low-risk consumers, which benefits the consumers.

In contrast to our baseline model, data linkage reduces consumer welfare in the insurance market. As in our baseline model, competition ensures that the informed insurer offers contracts that, in utility terms, are as good as the contract offered by uninformed competitors. Thus, individually, each consumer is indifferent to whether an informed insurer serves her. Nevertheless, collectively, consumers are worse off because the informed insurer tempts away only low-risk consumers and, thus, the capacity of the uniformed insurers to cross-subsidize high-risk consumers goes down, which, in turn, reduces consumer welfare in the insurance market.[33]

A priori it is not clear whether data linkage increases the overall consumer surplus 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.

[33] It is noteworthy that the loss of consumer welfare in the insurance market is, at most, the welfare difference between the cross-subsidy and RS equilibria.

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

Akerlof, George A, "The market for "lemons": quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 1970, *84* (3), 488–500. 39

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. 16

Argenziano, Rossella and Alessandro Bonatti, "Data linkages and privacy regulation," Working paper, 2021. Available at http://www.mit.edu/~bonatti/protection.pdf. 8

Attar, Andrea, Thomas Mariotti, and François Salanié, "Nonexclusive competition in the market for lemons," *Econometrica*, 2011, *79* (6), 1869–1918. 12

_ , _ , and _ , "Regulating insurance markets: multiple contracting and adverse selection," *International Economic Review*, 8 2022, *63* (3), 981–1020. 12

Australian Competition & Consumer Commission, "Digital platforms inquiry — Final report," June 2019. Retrieved from https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report. 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, 31

Bolton, Patrick and Mathias Dewatripont, *Contract theory*, Cambridge, MA; London, UK: MIT Press, 2004. 39

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 No. 107, September 2020. Retrieved from https://euagenda.eu/upload/publications/policyinsight107.pdf.pdf. 3, 24, 25

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, 9, 19, 24

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, "A new pro-competition regime for digital markets," Consultation paper No. 489, July 2021. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1003913/Digital_Competition_Consultation_v2.pdf. 23

—, "Mobile ecosystems market study final report," June 10, 2022. Retrieved from https://www.gov.uk/government/publications/mobile-ecosystems-market-study-final-report. 7

Competition Appeal Tribunal, "Royal Mail plc v Office of Communications: Judgement," Case No. 1299/1/3/18, November 12, 2019. Retrieved from https://www.catribunal.org.uk/judgments/12991318-royal-mail-plc-v-office-communications-judgment-2019-cat-27-12-nov-2019 26

— , "Royal Mail plc v Office of Communications: Judgement of the Court of Appeal," Case No. 1299/1/3/18, May 7, 2021. Retrieved from https://www.catribunal.org.uk/sites/default/files/2021-06/Royal%20Mail%20v%20Ofcom%20%5B2021%5D%20EWCA%20Civ%20669.pdf. 26

Condorelli, Daniele and Jorge Padilla, "Data-driven envelopment with privacy-policy tying," Working paper, 2021. Available at SSRN: https://www.ssrn.com/abstract=3600725. 8, 28

Cong, Jiajia and Noriaki Matsushima, "The effects of personal data management on competition and welfare," Discussion paper No. 1201, Institute of Social and Economic Research, Osaka University, March 2023. Retrieved from https://www.iser.osaka-u.ac.jp/library/dp/2023/DP1201.pdf. 8

Crémer, Jacques, Yves-Alexandre de Montjoye, and Heike Schweitzer, "Competition policy for the digital era," Report for the European Commission, 2019. Retrieved from https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf. 7

Davis, Peter, Lorenzo Coppi, Philip Kalmus, and Compass Lexecon, "The economics of secondary product markets," Discussion paper No. 1471, Office of Fair Trading, December 21, 2012. Retrieved from https://webarchive.nationalarchives.gov.uk/ukgwa/20140525130112/http:/www.oft.gov.uk/shared_oft/research/OFT1471_Waterbed.pdf. 8

de Cornière, Alexandre and Greg Taylor, "Data and competition: a simple framework, with applications to mergers and market structure," TSE Working paper No. 1076, 2021. Retrieved from https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/wp/2020/wp_tse_1076.pdf. 7, 8

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

Elliott, Matthew, Andrea Galeotti, Andrew Koh, and Wenhao Li, "Market segmentation through information," Cambridge working paper No. 2105, Janeway Institute working paper No. 2114, 2021. Retrieved from https://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe2105.pdf. 8

European Commission, "Communication from the Commission — Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings," Official Journal of the European Union, C 45/07, 2009. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52009XC0224(01). 25, 26

_ , "Case M.9660 — Google/Fitbit, Merger Procedure, Regulation (EC) 139/2004," 2020. Retrieved from https://ec.europa.eu/competition/mergers/cases1/202120/m9660_3314_3.pdf. 2

Financial Conduct Authority, "The potential competition impacts of Big Tech entry and expansion in retail financial services," Discussion Paper DP22/5, October 2022. Retrieved from https://www.fca.org.uk/publication/discussion/dp22-5.pdf. 6

Furman, Jason, Diane Coyle, Amelia Fletcher, Derek McAuley, and Philip Marsden, "Unlocking digital competition: report of the digital competition expert panel," Digital Competition Expert Panel, March 13, 2019. Retrieved from https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel. 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," *RAND Journal of Economics* forthcoming, 2022. Retrieved from http://andreihagiu.com/wp-content/uploads/2022/08/Data-enabled-learning-Final-RAND-Article.pdf. 8

Huang, Dakang and Christopher Sandmann, "Market structure and adverse selection," Working paper, 2022. Retrieved from https://drive.google.com/file/d/1H9c8C5Inff0vY6XLwDY6u64z3iPB36Qn. 12

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, Fiona Scott, Pascal Bouvier, Ariel Ezrachi, Bruno Jullien, Roberta Katz, Gene Kimmelman, Douglas Melamed, and Jamie Morgenstern, "Stigler committee on digital platforms, final report — subcommittee on market structure and antitrust," George J. Stigler Center for the Study of the Economy and the State, The University of Chicago Booth School of Business, September 2019. Retrieved from https://www.chicagobooth.edu/research/stigler/news-and-media/committee-on-digital-platforms-final-report. 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. 31, 87

Ofcom, "Ofcom's future approach to mobile markets and spectrum," Conclusions paper, December 6, 2022. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0036/248769/conclusions-mobile-spectrum-demand-and-markets.pdf. 6

Régibeau, Pierre, "Why I agree with the Google-Fitbit decision," Vox EU & CERP column, March 13, 2021. Retrieved from https://cepr.org/voxeu/columns/why-i-agree-google-fitbit-decision. 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. 72

7.3 Perfect competition

Following Akerlof (1970) adapted for the insurance market,[34] assume that quantity q is fixed to, say, $q = l$, so that only full-insurance contract is available to all consumers. Hence, only one premium p emerges in equilibrium, and consumers of either type prefer the cheapest contract.

Company 0 is uninformed

There are two cases.

Case 1 Suppose that both types buy the insurance. Then, the equilibrium premium must be $p = (\gamma\pi_L + (1-\gamma)\pi_H)l$. If the low type is willing to pay this premium, so does the high type. Hence, this equilibrium exists as long as the low type's individual rationality constraint holds:

$$u\left(y - (\gamma\pi_L + (1-\gamma)\pi_H)l\right) \geq \pi_L u(y-l) + (1-\pi_L)u(y). \tag{13}$$

The left-hand side of (13) is increasing in γ from $u(y-\pi_H l)$ to $u(y-\pi_L l)$. By assumption $u''(x) < 0$, $u(y-\pi_L l)$ is always greater than the right-hand side of (13). Define $\pi_H^* \in (\pi_L, 1)$ such that

$$u\left(y - \pi_H^* l\right) = \pi_L u(y-l) + (1-\pi_L)u(y). \tag{14}$$

Thus, condition (13) holds for all $\gamma \in (0,1)$ if $\pi_H \leq \pi_H^*$. If $\pi_H > \pi_H^*$, condition (13) holds if and only if $\gamma \geq \gamma^*(\pi_H)$ which solves

$$u\left(y - (\gamma^*\pi_L + (1-\gamma^*)\pi_H)l\right) = \pi_L u(y-l) + (1-\pi_L)u(y). \tag{15}$$

Case 2 Suppose that at most one type buys the insurance.[35] Since the high type is always willing to pay the premium that the low type is ready to pay, in the single-type-served equilibrium, it must be the high type who buys the insurance. Then, the equilibrium

[34] See also Chapter 13 in Bolton and Dewatripont (2004).

[35] Since everybody trades under full information, at least one type must be served here, and we do not have "full market breakdown," as in the original Akerlof's model in Akerlof (1970).

premium is $p = \pi_H l$, and this equilibrium exists as long as the low type consumers are not attracted by this contract:

$$u(y - \pi_H l) \leq \pi_L u(y - l) + (1 - \pi_L) u(y). \tag{16}$$

By definition (14) of π_H^*, condition (16) holds if and only if $\pi_H \geq \pi_H^*$.

Company 0 is informed

Version 1 When uninformed insurers serve both types, they make profit on the low risks and make losses on high risks. An informed insurer, which can distinguish between consumers, can profitably offer a more attractive contract to low risk but not high risk consumers. Hence, with an informed insurer, the proportion of low risks in the pool of consumers over which uniformed companies compete goes down. Denote the new fraction of low risk types by $\tilde{\gamma}$.

Case 1 If $\pi_H < \pi_H^*$, then the uninformed companies continue to serve both consumer types but at a higher price than in the absence of the informed insurer, $p = (\tilde{\gamma}\pi_L + (1 - \tilde{\gamma})\pi_H) l$. From each low risk consumer it serves, the informed insurer extracts a profit that is equal to the price offered by uninformed companies less the cost of serving the low risk type:

$$\Pi = p - \pi_L l = (1 - \tilde{\gamma})(\pi_H - \pi_L) l.$$

The informed insurer's profit per low-risk consumer served increases with the fraction of low-risk consumers it serves (decreases in $\tilde{\gamma}$), increases in π_H and decreases in π_L.

Case 2 If $\pi_H \geq \pi_H^*$ and $\tilde{\gamma} < \gamma^*(\pi_H)$, then uninformed insurers only serve the high risk type. An uninformed insurer cannot serve this type more profitably. However, it can offer the low risk type a premium that leaves this type indifferent between this contract and not taking out insurance:

$$u(y - p) = \pi_L u(y - l) + (1 - \pi_L) u(y).$$

The profit per low-risk consumer of the informed insurer is, therefore, independent of $\tilde{\gamma}$

and π_H , while it decreases in π_L .

Case 3 If $\pi_H < \pi_H^*$ and $\tilde{\gamma} \geq \gamma^*(\pi_H)$, then the uninformed companies continue to serve both consumer types but at a higher price than in the absence of the informed insurer, $p = (\tilde{\gamma}\pi_L + (1-\tilde{\gamma})\pi_H)l$ – see above.

Hence, if $\pi_H \geq \pi_H^*$ and $\gamma > \tilde{\gamma} \geq \gamma^*(\pi_H)$, then the uninformed companies serve both types with and without the informed insurer. If $\pi_H \geq \pi_H^*$ and $\tilde{\gamma} < \gamma^*(\pi_H) < \gamma$, then the uninformed companies serve both types without the informed insurer but only serve the high-risk type in the presence of the informed insurer. If $\pi_H \geq \pi_H^*$ and $\tilde{\gamma} < \gamma < \gamma^*(\pi_H)$ then the uninformed companies serve only the high-risk type with and without the informed insurer. This shows that it is possible that the entrance of an informed insurer changes the equilibrium from one in which the uninformed companies pool consumers to one in which they serve only the high type consumers.

7.3.1 Welfare

How does the entrance of an informed insurer affect welfare of each type of consumer? If $\pi_H < \pi_H^*$, or if $\pi_H \geq \pi_H^*$ and $\gamma > \tilde{\gamma} \geq \gamma^*(\pi_H)$, then the uninformed companies serve both types with and without an informed insurer, but the price is higher with the informed insurer. Therefore, both types of consumers are worse off in the presence of an informed insurer. (The informed insurer offers the same price as the uninformed insurers to those consumers it has identified as low risk. Therefore, the consumer's utility is independent of whether they are served by an informed or by an uninformed insurer). There is no efficiency gain or loss from data linkage. If $\pi_H \geq \pi_H^*$ and $\tilde{\gamma} < \gamma < \gamma^*(\pi_H)$ then the uninformed companies serve only the high-risk type with and without the informed insurer at the same price. Neither type is made better or worse off. An efficiency gain arises as the identified low-risk types are served by the informed insurer. Therefore, the entrance of the informed insurer is a Pareto improvement. If $\pi_H \geq \pi_H^*$ and $\tilde{\gamma} < \gamma^*(\pi_H) < \gamma$, then the uninformed companies serve both types without the informed insurer but only serve the high-risk type in the presence of the informed insurer. The price charged to the high-risk type increases in the presence of an informed insurer from

$$p = (\gamma\pi_L + (1-\gamma)\pi_H)\,l$$

to $p = \pi_H l$. Essentially, the contract is no longer subsidized by low-risk consumers. Further, low-risk consumers are made worse off because they get a positive utility from the pooling contract offered without the informed insurer but zero utility with the informed insurer. The price increase of the contract offered by uninformed companies leaves them unwilling to buy this contract. While both types were served without data linkage, unidentified low-risk types are no longer served with data linkage. Hence, efficiency declines with data linkage. Therefore, we find that the emergence of an informed insurer harms consumers unless low-risk consumers are not served in the absence of an informed insurer. In particular, it harms even low-risk consumers who have a "preferable" risk profile. The reason is that the informed insurer cream-skims some low-type consumers which otherwise would have helped to subsidize the insurance contract offered by uninformed companies. Each individual low-risk consumer is indifferent to which company serves him but because only low-risk are offered a contract by an uninformed insurer and the price is the same, but the choice to be insured by the informed insurer exerts a negative externality on every consumer.

7.3.2 Version 2: Alternative assumption

Suppose that the informed insurer makes personalized offers to identified consumers, but can also serve unidentified consumers. Unlike the uninformed insurers, the informed insurer still faces a share γ of low-risk types in the pool of unidentified consumers. Therefore, the informed insurer could offer a price just marginally below the price that leaves zero profit given a share of low-risk types equal to $\tilde{\gamma}$ and win over all unidentified consumers, making a positive profit on these unidentified consumers. Hence, we need to look at a different equilibrium than the one described in version 1. The difficulty is that it is also not an equilibrium for the informed insurer to serve all unidentified types at a price that leaves zero profit given a share of low-risk types equal to $\tilde{\gamma}$. Then the uninformed insurers would be making a loss, since they still would serve all identified high-risk consumers at this price. The uninformed insurers would deviate to a price that leaves zero profit for a pool consisting of only high-risk consumers. But then the informed insurer would want to raise its price for the unidentified consumers...

Conjecture: Suppose that there is only one uninformed insurer and in equilibrium, the informed and uninformed insurer mix in the support of prices $\left[\underline{p}, \bar{p}\right]$ where

$$\bar{p} = \pi_H l \tag{17}$$

(the price at which profit is zero if the consumer pool consists only of high-risk consumers - which is the case if only identified high-risk types are served) and

$$\underline{p} = (\tilde{\gamma}\pi_L + (1-\tilde{\gamma})\pi_H)\, l \tag{18}$$

(the price at which profit is zero if the consumer pool has a share $\tilde{\gamma}$ of low-risk consumers - which is the case if all unidentified consumers and all identified high-risk types are served). The informed insurer mixes according to a cdf F_0 and uninformed insurer mixes according to a cdf F_1. Then, the profit of the uniformed insurer given any price p in the support is equal to zero and so

$$p = (1-F_0(p))(\tilde{\gamma}\pi_L + (1-\tilde{\gamma})\pi_H)\, l + F_0(p)\pi_H l \tag{19}$$

If the support is as stated above, the informed insurer cannot have an atom at $\underline{p}$, because then the uninformed insurer does not win for certain when choosing $\underline{p}$.

The profit of the informed insurer given any price in the support should be equal to the profit this company would get by winning all unidentified consumers at price $\underline{p}$:

$$(1-F_1(p))(p - (\gamma\pi_L + (1-\gamma)\pi_H)\, l) \tag{20}$$

7.3.3 Version 3: Alternative assumption

Suppose that all insurance companies make offers simultaneously and suppose that the informed insurer can offer different prices to identified low-risk consumers (p_L), identified high-risk consumers (p_H) and unidentified consumers (p_U).

Suppose in an equilibrium all uninformed companies charge price p.

In any equilibrium, all uninformed insurers get zero profit in expectation.

Moreover, in any equilibrium $p_H \geq \pi_H l$, $p_L \geq \pi_L l$ and $p_U \geq (\gamma\pi_L + (1-\gamma)\pi_H)l$ as otherwise the informed insurer makes losses on the corresponding group of consumers. (That is, lower prices are weakly dominated).

In any equilibrium, the informed insurer either does not serve the identified high-risk consumers, or sets $p_H = \pi_H l$ and gets zero profit from serving them. Otherwise, the uninformed companies can undercut p_H and get positive profit. Hence, without loss of generality, we can consider only equilibria in which $p_H = \pi_H l$.

If an uninformed insurer attracts consumers at any price $p < (\gamma\pi_L + (1-\gamma)\pi_H)l$, the insurer makes a loss because at $p < \pi_H l$, the pool of consumers which are served by uninformed insurers is weakly worse than the pool which is served by the informed insurer (the informed insurer does not serve any identified high-risks at $p < \pi_H l$). Hence, we are going to restrict attention to equilibria (in undominated strategies) in which $(\gamma\pi_L + (1-\gamma)\pi_H)l \leq p \leq \pi_H l$. Hence, $(\gamma\pi_L + (1-\gamma)\pi_H)l \leq p_L \leq \min\{w_L, \pi_H l\}$ and $p_U \leq \pi_H l$.

1. If $w_L < (\gamma\pi_L + (1-\gamma)\pi_H)l$, then the unidentified low-risk consumers are not served as neither informed, nor uninformed companies find it profitable. Hence, the unique equilibrium is $p = p_H = p_U = \pi_H l$, $p_L = w_L$.

2. Suppose that $w_L \geq \pi_H l$.

   Let s be the share identified consumers. The informed insurer's profit from identified low-risk is

$$s\gamma(1-F(p_L))^K(p_L - \pi_L l)$$

   The informed insurer's profit from unidentified consumers is

$$(1-s)(1-F(p_U))^K(p_U - (\gamma\pi_L + (1-\gamma)\pi_H)l)$$

   If the supports of F_L and F_U intersect on an interval, then, on that interval,

$$\frac{\mathrm{d}}{\mathrm{d}p}\left(s\gamma(1-F(p))^K(p-\pi_L l)\right) = \frac{\mathrm{d}}{\mathrm{d}p}\left((1-s)(1-F(p))^K(p-(\gamma\pi_L + (1-\gamma)\pi_H)l)\right) = 0$$

$$\Rightarrow \quad \frac{1-F(p)}{f(p)} = K(p-\pi_L l) = K\left(p-(\gamma\pi_L+(1-\gamma)\pi_H)l\right)$$

which is a contradiction. (The informed insurer must earn the same profit at any price in the support of its mixing distribution.) Thus, the supports of F_L and F_U do not intersect on a non-degenerate interval.

Naturally, the support of F_U lies above the support of F_L.

The uninformed insurers' profit from setting price p is

$$(1-F(p))^{K-1}\left(s(1-\gamma)(p-\pi_H l)+(1-F_U(p))(1-s)(p-(\gamma\pi_L+(1-\gamma)\pi_H)l)+(1-F_L(p))s\gamma(p-\pi_L l)\right)$$

which implies

$$s(1-\gamma)(p-\pi_H l)+(1-F_U(p))(1-s)(p-(\gamma\pi_L+(1-\gamma)\pi_H)l)+(1-F_L(p))s\gamma(p-\pi_L l)=0 \tag{21}$$

Hence, substituting the upper bound of the support F_L, $p=\bar{p}_L$, $F_L(\bar{p}_L)=1$ and $F_U(\bar{p}_L)=0$, into (21) gives

$$\bar{p}_L = \frac{(1-\gamma)\pi_H+(1-s)\gamma\pi_L}{1-s\gamma}l \equiv p(\gamma');$$

substituting the lower bound of the support F_U, $p=\underline{p}_U$, $F_L(\underline{p}_U)=1$ and $F_U(\underline{p}_U)=0$, into (21) gives $\underline{p}_U=\bar{p}_L$; substituting the upper bound of the support F_U, $p=\bar{p}_U$, $F_U(\bar{p}_U)=F_L(\bar{p}_U)=1$, into (21) gives $\bar{p}_U=\pi_H l$; substituting the lower bound of the support F_L, $p=\underline{p}_L$, $F_U(\underline{p}_L)=F_L(\underline{p}_L)=0$, into (21) gives

$$\underline{p}_L=(\gamma\pi_L+(1-\gamma)\pi_H)l \equiv p(\gamma).$$

The informed insurer's profit from identified low-risk is

$$s\gamma(p(\gamma)-\pi_L l)=s\gamma(1-\gamma)(\pi_H-\pi_L)l$$

Thus, in equilibrium, the uninformed insurers mix

$$1-F(p)=\left(\frac{p(\gamma)-\pi_L l}{p-\pi_L l}\right)^{1/K}, \quad p(\gamma)\le p\le p(\gamma').$$

The informed insurer's profit from unidentified consumers is

$$(1-s)\frac{p(\gamma)-\pi_L l}{p(\gamma')-\pi_L l}\left(p(\gamma')-p(\gamma)\right)=s(1-s)\gamma(1-\gamma)(\pi_H-\pi_L)l$$

so that the total profit of the informed insurer is

$$s(2-s)\gamma(1-\gamma)(\pi_H-\pi_L)l$$

Thus, in equilibrium, the uninformed insurers mix

$$1-F(p)=\left(\frac{p(\gamma)-\pi_L l}{p(\gamma')-\pi_L l}\frac{p(\gamma')-p(\gamma)}{p-p(\gamma)}\right)^{1/K}, \quad p(\gamma')\le p\le \pi_H l.$$

The informed insurer sets p_L mixing on $[p(\gamma),p(\gamma')]$ with distribution

$$1-F_L(p)=\frac{s(1-\gamma)(\pi_H l-p)-(1-s)(p-p(\gamma))}{s\gamma(p-\pi_L l)} \quad \Longleftrightarrow$$

$$F_L(p)=\frac{p-p(\gamma)}{s\gamma(p-\pi_L l)}=\frac{p(\gamma')-\pi_L l}{p-\pi_L l}\frac{p-p(\gamma)}{p(\gamma')-p(\gamma)}$$

and sets p_U mixing on $[p(\gamma'),\pi_H l]$ with distribution

$$1-F_U(p)=\frac{s(1-\gamma)(\pi_H l-p)}{(1-s)(p-p(\gamma))} \quad \Longleftrightarrow \quad F_U(p)=\frac{1-s\gamma}{1-s}\frac{p-p(\gamma')}{p-p(\gamma)}$$

3. Suppose that $p(\gamma')\le w_L\le \pi_H l$.

   In equilibrium, the uninformed companies mix on $[p(\gamma),w_L]\cup\{\pi_H l\}$:

$$1-F(p)=\left(\frac{p(\gamma)-\pi_L l}{p-\pi_L l}\right)^{1/K}, \quad p(\gamma)\le p\le p(\gamma');$$

$$1 - F(p) = \left(\frac{p(\gamma) - \pi_L l}{p(\gamma') - \pi_L l} \frac{p(\gamma') - p(\gamma)}{p - p(\gamma)} \right)^{1/K}, \quad p(\gamma') \leq p < w_L;$$

and there will be a mass point on $\pi_H l$. (There cannot be a mass point on w_L because the informed insurer has a mass point at w_L (see below) and, therefore, uninformed companies could lower their price marginally below w_L and experience a discrete increase in the probability of offering the lowest price to unidentified consumers and, hence, enjoy a discrete increase in expected profit. Remember the share of low-risk in the pool of unidentified consumers is γ and so any price above $p(\gamma)$ means a positive profit.) The informed insurer sets p_L mixing on $[p(\gamma), p(\gamma')]$ with distribution

$$F_L(p) = \frac{p - p(\gamma)}{s\gamma(p - \pi_L l)} = \frac{p(\gamma') - \pi_L l}{p - \pi_L l} \frac{p - p(\gamma)}{p(\gamma') - p(\gamma)}$$

and sets p_U mixing on $[p(\gamma'), w_L]$ with distribution

$$1 - F_U(p) = \frac{s(1-\gamma)(\pi_H l - p)}{(1-s)(p - p(\gamma))} \quad \Leftrightarrow \quad F_U(p) = \frac{1 - s\gamma}{1 - s} \frac{p - p(\gamma')}{p - p(\gamma)},$$

so that there is a mass point on w_L. The informed insurer cannot price above w_L as any such price would attract only high-risk consumers and the insurer would make non-positive profit, which is a contradiction (at prices below w_L the informed insurer makes positive expected profit). The total profit of the informed insurer stays the same:

$$s(2-s)\gamma(1-\gamma)(\pi_H - \pi_L)l$$

4. Suppose that $p(\gamma) \leq w_L \leq p(\gamma')$.

In equilibrium, the uninformed companies mix on $[p(\gamma), w_L] \cup \{\pi_H l\}$:

$$1 - F(p) = \left(\frac{p(\gamma) - \pi_L l}{p - \pi_L l} \right)^{1/K}, \quad p(\gamma) \leq p \leq w_L;$$

and there will be a mass point on $\pi_H l$. The informed insurer sets p_L mixing on $[p(\gamma), w_L]$

with distribution

$$F_L(p) = \frac{p - p(\gamma)}{s\gamma(p - \pi_L l)} = \frac{p(\gamma') - \pi_L l}{p - \pi_L l} \frac{p - p(\gamma)}{p(\gamma') - p(\gamma)}$$

with the mass point on w_L and sets $p_U = w_L$. The informed insurer's profit from unidentified consumers is

$$(1-s)\left(\frac{p(\gamma) - \pi_L l}{w_L - \pi_L l}\right)(w_L - p(\gamma)) = (1-s)(1-\gamma)(\pi_H - \pi_L) l \frac{w_L - p(\gamma)}{w_L - \pi_L l}$$

so that the total profit of the informed insurer is

$$\left(s\gamma + (1-s)\frac{w_L - p(\gamma)}{w_L - \pi_L l}\right)(1-\gamma)(\pi_H - \pi_L) l$$

The identified high-risk consumers suffer a loss

$$\int_{p(\gamma)}^{w_L} (p - p(\gamma))\, d(1 - (1 - F(p))^K) + (1 - F(w_L))^K (\pi_H l - p(\gamma)) =$$

$$(w_L - p(\gamma))(1 - (1 - F(w_L))^K) - \int_{p(\gamma)}^{w_L} (1 - (1 - F(p))^K)\, dp + (1 - F(w_L))^K (\pi_H l - p(\gamma)) =$$

$$w_L - p(\gamma) - \int_{p(\gamma)}^{w_L} \left(1 - \frac{p(\gamma) - \pi_L l}{p - \pi_L l}\right) dp + \frac{p(\gamma) - \pi_L l}{w_L - \pi_L l}(\pi_H l - w_L) =$$

$$\left(\frac{\pi_H l - w_L}{w_L - \pi_L l} - \ln\left(\frac{(1-\gamma)(\pi_H - \pi_L) l}{w_L - \pi_L l}\right)\right)(1-\gamma)(\pi_H - \pi_L) l$$

The identified low-risk consumers suffer a loss

$$\int_{p(\gamma)}^{w_L} (p-p(\gamma))\,\mathrm{d}\big(1-(1-F(p))^K(1-F_L(p))\big)+(1-F(w_L))^K(1-F_L(w_L))(w_L-p(\gamma))=$$

$$w_L-p(\gamma)-\int_{p(\gamma)}^{w_L}\big(1-(1-F(p))^K(1-F_L(p))\big)\,\mathrm{d}p=$$

$$w_L-p(\gamma)-\int_{p(\gamma)}^{w_L}\left(1-\frac{p(\gamma)-\pi_L l}{p-\pi_L l}\left(1-\frac{p-p(\gamma)}{s\gamma(p-\pi_L l)}\right)\right)\mathrm{d}p=$$

$$\left(\frac{w_L-p(\gamma)}{w_L-\pi_L l}+(1-s\gamma)\ln\left(\frac{(1-\gamma)(\pi_H-\pi_L)l}{w_L-\pi_L l}\right)\right)\frac{(1-\gamma)(\pi_H-\pi_L)l}{s\gamma}$$

Summary The total profit of the informed insurer is

$$s(2-s)\gamma(1-\gamma)(\pi_H-\pi_L)l,\quad w_L\ge p(\gamma');$$

$$\left(s\gamma+(1-s)\frac{w_L-p(\gamma)}{w_L-\pi_L l}\right)(1-\gamma)(\pi_H-\pi_L)l,\quad p(\gamma)\le w_L\le p(\gamma')$$

$$s\gamma(w_L-\pi_L l),\quad w_L<p(\gamma).$$

The identified high-risk consumers suffer a loss

$$\left(\frac{\pi_H l-w_L}{w_L-\pi_L l}-\ln\left(\frac{(1-\gamma)(\pi_H-\pi_L)l}{w_L-\pi_L l}\right)\right)(1-\gamma)(\pi_H-\pi_L)l,\quad p(\gamma)\le w_L\le p(\gamma')$$

$$0,\quad w_L<p(\gamma).$$

The identified low-risk consumers suffer a loss

$$\left(\frac{w_L-p(\gamma)}{w_L-\pi_L l}+(1-s\gamma)\ln\left(\frac{(1-\gamma)(\pi_H-\pi_L)l}{w_L-\pi_L l}\right)\right)\frac{(1-\gamma)(\pi_H-\pi_L)l}{s\gamma},\quad p(\gamma)\le w_L\le p(\gamma')$$

$$0,\quad w_L<p(\gamma).$$

7.4 Monopoly

7.4.1 Without data linkage

Suppose one firm offers a contract with coverage $q = l$ at premium p. By the same reasoning as above, we distinguish between two cases.

Case 1 Suppose that both types buy the insurance. Then, the equilibrium premium must be such that the low-risk type is just willing to purchase the contract. Hence, the price must satisfy:

$$u(y - p_1) = \pi_L u(y - l) + (1 - \pi_L)u(y).$$

Then, the monopoly profit is

$$\Pi_1 = p_1 - (\gamma \pi_L + (1 - \gamma)\pi_H)\, l$$

per consumer.

Case 2 Suppose that only the high-risk type buys the insurance. Then, the equilibrium premium must be such that the high-risk type is just willing to purchase the contract. Hence, the price must satisfy:

$$u(y - p_2) = \pi_H u(y - l) + (1 - \pi_H)u(y).$$

Then the monopoly profit is

$$\Pi_2 = (1 - \gamma)\,(p_2 - \pi_H l)$$

per consumer.

The difference in profits is

$$\Pi_1 - \Pi_2 = p_1 - p_2 + \gamma\,(p_2 - \pi_L l)\,,$$

which is increasing in $\gamma \in [0,1]$ from $p_1 - p_2 < 0$ to $p_1 - \pi_L l > 0$. Hence, the monopoly serves both types if

$$\Pi_1 > \Pi_2 \quad \Leftrightarrow \quad \gamma > \hat{\gamma}$$

and serves only high-risk type if

$$\Pi_1 < \Pi_2 \quad \Leftrightarrow \quad \gamma < \hat{\gamma}$$

where

$$\hat{\gamma} = \frac{p_2 - p_1}{p_2 - \pi_L l} \in (0,1).$$

7.4.2 With data linkage

The monopoly offers p_1 to any identified low-risk consumer and p_2 to any identified high-risk consumer. Suppose the pool of unidentified consumers has a share of low-risk types equal to $\tilde{\gamma}$. Contracts to these unidentified consumers are as above with $\gamma = \tilde{\gamma}$.

Each low risk consumer is indifferent to whether the insurer identifies her as low risk because she always receives the same utility as with no insurance. The high risk consumers prefer to remain unidentified if $\tilde{\gamma} > \hat{\gamma}$ in which case she has incentives to avoid company 0 in the product market. Consequently, it should be the case that $\tilde{\gamma} \leq \gamma$.

Case 1: $\gamma < \hat{\gamma}$**.** Since $\tilde{\gamma} \leq \gamma$, in this case, we must have that $\tilde{\gamma} < \hat{\gamma}$. Hence, the monopolist offers p_2 to all unidentified consumers and only the high-risk type takes the offer. Since the monopolist always makes the high-risk type indifferent to taking out the insurance contract, the high-risk consumers have no reason to avoid company 0 in the product market. Hence, $\tilde{\gamma} = \gamma$.

Case 2: $\gamma > \hat{\gamma}$**.** If $\tilde{\gamma} \leq \hat{\gamma}$, then the high-risk consumers do not have incentives to avoid company 0 in the product market, which means that $\tilde{\gamma} = \gamma$ — a contradiction with $\gamma > \hat{\gamma}$. Hence, $\tilde{\gamma} > \hat{\gamma}$, and so, the high risk consumers prefer to remain unidentified, which means that $\tilde{\gamma} < \gamma$.

7.4.3 Welfare

As stated above, the low risk type consumer is indifferent to whether there is data linkage.

If $\gamma < \hat{\gamma}$, then, with and without data linkage, the monopolist always makes the high-risk type indifferent between taking out the insurance contract or not. Hence, the high-risk consumers are indifferent to whether there is data linkage.

If $\gamma > \hat{\gamma}$, then without data linkage, high-risk consumers have positive rent when taking out the insurance contract. With data linkage, an identified high-risk consumer has no positive rent, while an unidentified high-risk consumer has lower rent than without data linkage. This is because, in equilibrium, some high-risk consumers avoid the monopolist in the product market to have some positive rent in the insurance market, and this implies that the share of high-risk consumers in the pool of unidentified consumers is higher than in the pool of identified consumers. Therefore, the rents available to unidentified high-risk consumers are lower than without data linkage. Overall, this implies that high-risk consumers are made worse off in the insurance market.

7.4.4 Adding the product market

Case $\gamma < \hat{\gamma}$ All consumers are indifferent to revealing their risk type to company 0, and the analysis from Section 3.2 applies. However, data linkage allows the monopolist to earn additional profit per low-risk identified consumer Π:

$$\Pi = p_1 - \pi_L l$$

In the absence of data linkage, the insurer always serves all consumers in the insurance market and the profit from the insurance market is independent of outcomes in the product market. Consequently, the insurance market profit does not affect company 0's behaviour in the product market. With data linkage, the monopolist makes additional profit per identified consumer and this additional profit does affect behavior in the product market. In this case, the additional profit is only derived from identified low-risk consumers.

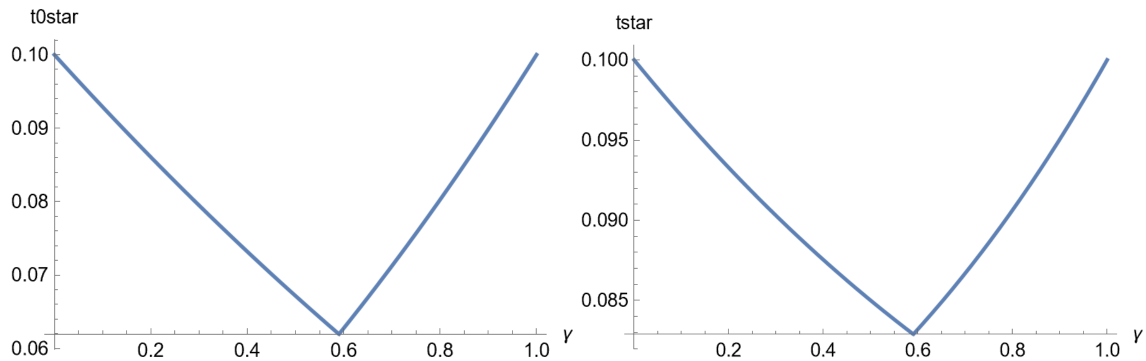


Figure 3: Prices t_0^* and t^* for myopic consumers. Parameters: $l = y = 1$, $\sigma = 0.05$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$. For these parameters, $\hat{\gamma} \approx 0.59$

Case $\gamma > \hat{\gamma}$, myopic consumers Suppose that, despite that the high-risk consumers have incentives to avoid company 0, they don't, so that $\tilde{\gamma} = \gamma$. When facing an unidentified consumer or low-risk identified consumers, the monopoly offers p_1, and it offers p_2 to high-risk identified consumers. Thus, in contrast to Case $\gamma < \hat{\gamma}$, the monopoly gets data linkage profit only from identified high-risk consumers. Its profit per identified high-risk consumer is

$$\Pi = p_2 - p_1.$$

With and without data linkage, all consumers purchase insurance — just the identified high-risk types pay a higher price (p_2 instead of p_1).

The analysis of the product market stays the same as in Section 3.2, apart from replacing $\gamma\Pi$ with $(1-\gamma)\Pi$.

Case $\gamma > \hat{\gamma}$, forward-looking consumers Suppose that now the high-risk consumers tend to avoid company 0, and so, $\tilde{\gamma} < \gamma$. As in the previous case with myopic consumers, when facing an unidentified consumer or identified low-risk consumers, the monopoly offers p_1, and it offers p_2 to identified high-risk consumers, and so, the monopoly gets data linkage profit only from identified high-risk consumers; this profit is

$$\Pi = p_2 - p_1 = w_H - w_L.$$

In the analysis of the product market, we have to take into account that the high-risk

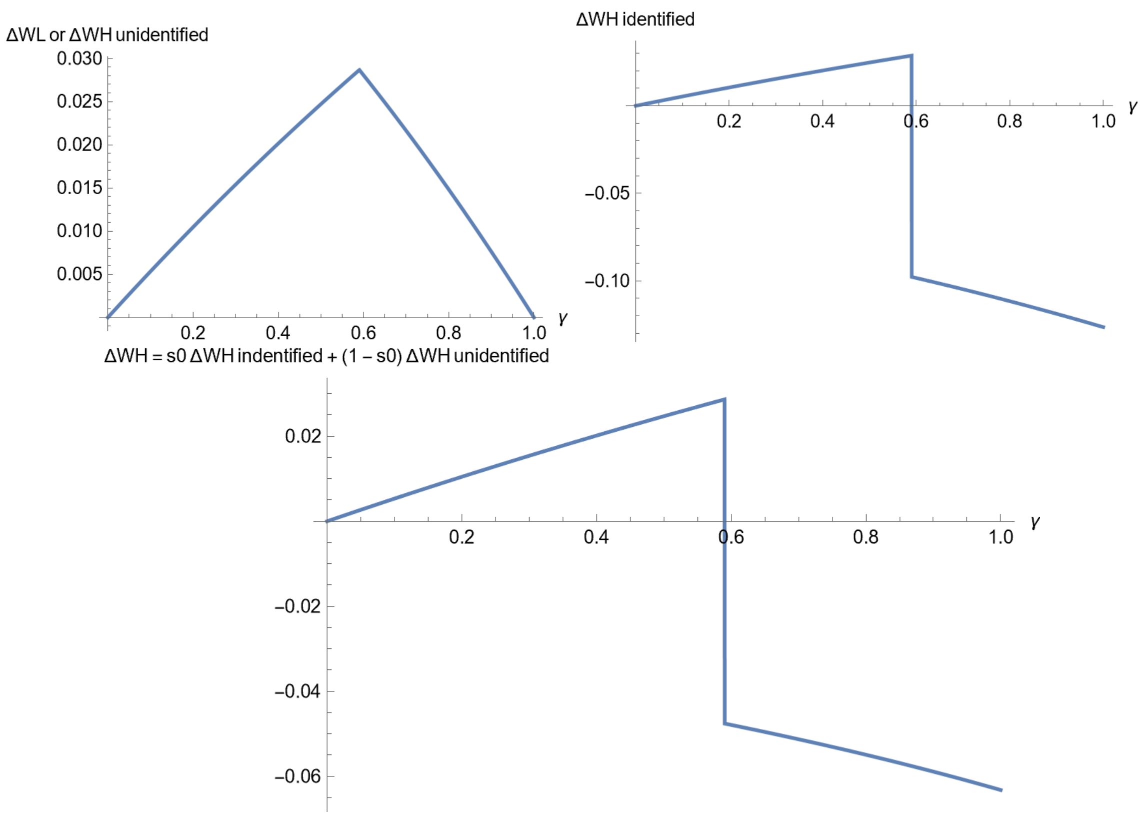


Figure 4: Welfare change Δ_W^L and Δ_W^H for myopic consumers, identified and not. Parameters: $l = y = 1$, $\sigma = 0.05$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$. For these parameters, $\hat{\gamma} \approx 0.59$

consumer gains

$$\delta V = u(y-p_1) - u(y-p_2) = (\pi_L u(y-l) + (1-\pi_L)u(y)) - (\pi_H u(y-l) + (1-\pi_H)u(y))$$
$$= (\pi_H - \pi_L)(u(y) - u(y-l))$$

(if we work with utilities), or

$$\delta V = w_H - w_L$$

(if we work with certainty equivalence) by hiding her risk type. The analysis in Proposition B.1 applies, with $\delta\Pi = (1-\gamma)s_0^H \Pi$ and δV defined above. Hence, since δV is independent of s_0^H and s_0^L, (B.14) becomes

$$t^* \left\{ \gamma(1-s_0^L)\left(1 - \frac{1-s_0^L}{N}\right) + (1-\gamma)(1-s_0^H)\left(1 - \frac{1-s_0^H}{N}\right) \right\} = \sigma\left(\gamma(1-s_0^L) + (1-\gamma)(1-s_0^H)\right)$$

and (B.15) becomes

$$s_0^L(1-s_0^L)\gamma t_0^* + s_0^H(1-s_0^H)(1-\gamma)\left(t_0^* + \Pi\right) = \sigma\left(\gamma s_0^L + (1-\gamma)s_0^H\right).$$

After we substitute t^* and t_0^* from the two equations above into (B.12) and (B.13), we get a system of two equations for s_0^H and s_0^L, which allows us then to recover the prices t^* and t_0^*:

$$\frac{\sigma\left(\gamma s_0^L + (1-\gamma)s_0^H\right) - s_0^H(1-s_0^H)(1-\gamma)\Pi}{s_0^L(1-s_0^L)\gamma + s_0^H(1-s_0^H)(1-\gamma)} = \frac{\sigma\left(\gamma(1-s_0^L) + (1-\gamma)(1-s_0^H)\right)}{\gamma(1-s_0^L)\left(1 - \frac{1-s_0^L}{N}\right) + (1-\gamma)(1-s_0^H)\left(1 - \frac{1-s_0^H}{N}\right)} + \sigma \ln \frac{1-s_0^L}{N s_0^L}$$

$$\sigma \ln \frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H} = \delta V$$

As for the change in the consumer welfare as a result of data linkage, Proposition B.2

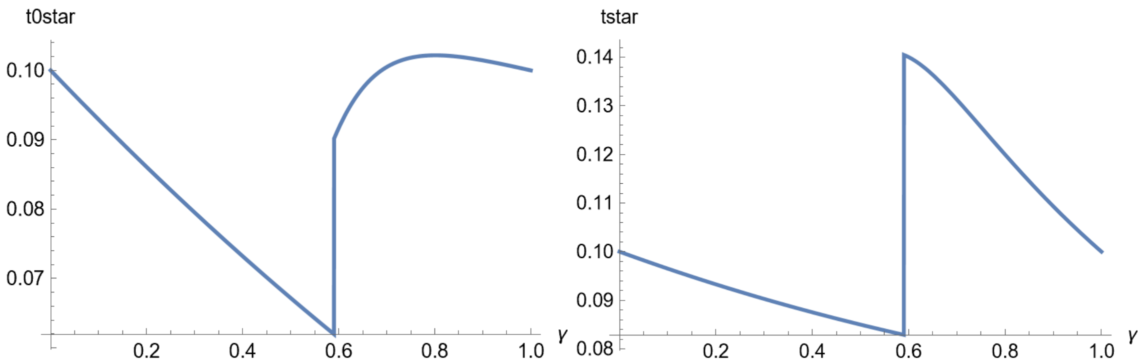


Figure 5: Prices t_0^* and t^* for forward-looking consumers. Parameters: $l = y = 1$, $\sigma = 0.05$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$. For these parameters, $\hat{\gamma} \approx 0.59$

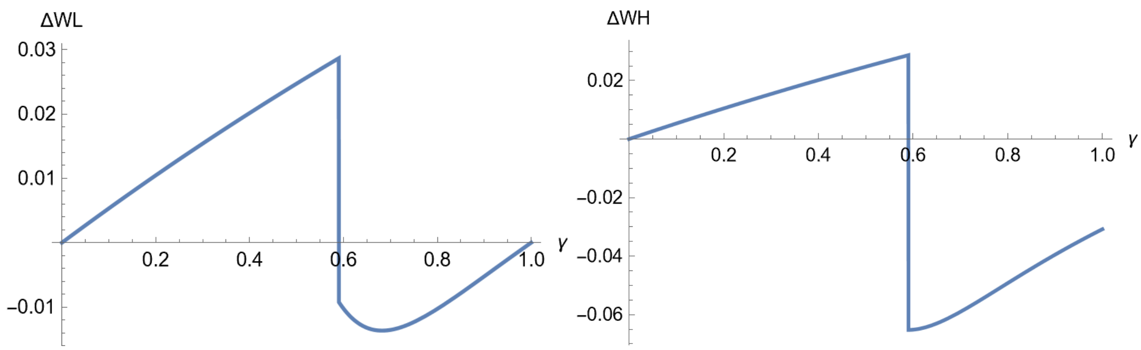


Figure 6: Welfare change Δ_W^L and Δ_W^H for forward-looking consumers. Parameters: $l = y = 1$, $\sigma = 0.05$, $N = 1$, $\pi_H = 0.8$, $\pi_L = 0.6$. For these parameters, $\hat{\gamma} \approx 0.59$

holds with no change except that (B.26) becomes

$$\Delta_W^H = \sigma\left(\ln\frac{N}{(N+1)(1-s_0^H)} + \frac{N+1}{N}\right) - t^*.$$

Graphs for the case when we work with utilities:

If we work with certainty equivalence, then note that $\delta V = \Pi (= w_H - w_L)$. Denote by

$$x \equiv \frac{w_H - w_L}{\sigma}$$

Then, the system of equations for s_0^L and s_0^H is

$$\frac{\gamma s_0^L + (1-\gamma)s_0^H - (1-\gamma)s_0^H(1-s_0^H)x}{\gamma s_0^L(1-s_0^L) + (1-\gamma)s_0^H(1-s_0^H)} = \frac{\gamma(1-s_0^L) + (1-\gamma)(1-s_0^H)}{\gamma(1-s_0^L)\left(1-\frac{1-s_0^L}{N}\right) + (1-\gamma)(1-s_0^H)\left(1-\frac{1-s_0^H}{N}\right)} + \ln\frac{1-s_0^L}{Ns_0^L}$$

$$\ln\frac{s_0^L(1-s_0^H)}{(1-s_0^L)s_0^H} = x$$

The welfare increase from data linkage for type i consumer is

$$\Delta_W^i = \sigma\left(\ln\frac{N}{(N+1)(1-s_0^i)} + \frac{N+1}{N} - \frac{\gamma(1-s_0^L) + (1-\gamma)(1-s_0^H)}{\gamma(1-s_0^L)\left(1-\frac{1-s_0^L}{N}\right) + (1-\gamma)(1-s_0^H)\left(1-\frac{1-s_0^H}{N}\right)}\right)$$

and the prices are

$$t^* = \sigma\frac{\gamma(1-s_0^L) + (1-\gamma)(1-s_0^H)}{\gamma(1-s_0^L)\left(1-\frac{1-s_0^L}{N}\right) + (1-\gamma)(1-s_0^H)\left(1-\frac{1-s_0^H}{N}\right)},$$

$$t_0^* = \sigma\frac{\gamma s_0^L + (1-\gamma)s_0^H - (1-\gamma)s_0^H(1-s_0^H)x}{\gamma s_0^L(1-s_0^L) + (1-\gamma)s_0^H(1-s_0^H)}$$

Note that the price under no data linkage is $\frac{N+1}{N}\sigma$, and the consumer welfare without data linkage is

$$W_{\text{independent}}^i = V + \sigma\left(\ln(N+1) - \frac{N+1}{N}\right) + (w_i - w_L)$$

7.4.5 Comparison: myopic and forward-looking consumers

When $\gamma < \hat{\gamma}$, prices decrease with γ starting from the price level without data linkage. In this case, we do not need to distinguish between myopic and forward-looking consumers - consumers are indifferent between being identified or not. Prices decrease with γ because the informed insurer makes additional profit from identifying and then serving low-risk consumers (uninformed insurer do not serve low-risk consumers at all). The higher the share of low-risk consumers, the higher the additional profit in the insurance market derived from a given additional market share in the product market. Therefore, company 0 has a stronger incentive to lower its price the higher γ.

Consider $\gamma > \hat{\gamma}$. When consumers are myopic, prices rise in γ until the price level without data linkage. By contrast, when consumers are forward-looking, the price level jumps up at $\gamma = \hat{\gamma}$, competitors' prices are then higher than the price level without data linkage and fall in γ, company 0's price level jumps up and then increases further until dropping back to the price level without data linkage. Both prices appear higher with forward-looking consumer than with myopic consumers at all γ.

- With myopic consumers: competitors' price t^* increases with γ and company 0's price t^* increases with γ. The reason is that company 0 makes additional profit only from identified high-risk consumers. The higher γ, the fewer additional high-risk consumers are buying from company 0 for a given decrease in its price t_0. Therefore, company 0 and its competitors increase prices in the product market with γ up to the level they would charge in the absence of data linkage.

- With forward-looking consumers, the reasons for the price changes are less clear. In contrast to the case with myopic consumers, competitors now offer the advantage of staying unidentified. This leads them to charge a higher price compared to the case with myopic consumers, all else equal. In equilibrium, the prices of company 0 and competitors increase relative to the case with myopic consumers. It turns out that competitors' price lies above the level without data linkage at all γ.

- It seems that company 0 has less reason to lower its price below the price level without data linkage at $\gamma = \hat{\gamma}$ when consumers are forward-looking rather than myopic. This could be because it is harder to attract high-risk types and these are the types from which the informed insurer derives additional profit in the insurance market. However, at some point, company 0 even increases its price above the price level without data linkage. It seems that the company is extracting more surplus from those consumers with a strong preference for their product. Competitors charge a higher price throughout. This maybe because they now offer the additional benefit of staying unidentified.

-

-

-

-

Welfare change of low-risk consumers with data linkage compared to without:

- With myopic consumers, the welfare change of low-risk consumers in positive at all γ. It is increasing for $\gamma < \hat{\gamma}$ and decreasing for $\gamma > \hat{\gamma}$. This is because low-risk consumers get the same utility in the insurance market irrespective of data linkage but they get offered lower prices in the product market with data linkage.

- With forward-looking consumers, the welfare change is positive for low γ but negative for high γ.

Appendix A Technical Results

A.1 Proof of Proposition 1

Comparative statics with respect to π_H

Let $q_L^{RS}(\pi_H)$ be the solution to (3). Applying the implicit function theorem to (3), we get

$$\frac{\mathrm{d}q_L^{RS}(\pi_H)}{\mathrm{d}\pi_H} = \frac{u(y-\pi_L q_L^{RS}) - u(y-\pi_L q_L^{RS} + q_L^{RS} - l) - lu'(y-\pi_H l)}{\pi_H(1-\pi_L)u'(y-\pi_L q_L^{RS} + q_L^{RS} - l) - \pi_L(1-\pi_H)u'(y-\pi_L q_L^{RS})}. \tag{A.1}$$

The denominator in (A.1) is positive because $\pi_H(1-\pi_L) > \pi_L(1-\pi_H)$ and

$$u'(y-\pi_L q_L^{RS} + q_L^{RS} - l) > u'(y-\pi_L q_L^{RS}) \tag{A.2}$$

since $q_L^{RS} < l$ and $u'(x)$ is decreasing. The numerator in (A.1) is negative because

$$\begin{aligned} u(y-\pi_L q_L^{RS}) - u(y-\pi_L q_L^{RS} + q_L^{RS} - l) - lu'(y-\pi_H l) &\overset{(3)}{=} \frac{u(y-\pi_L q_L^{RS}) - u(y-\pi_H l)}{\pi_H} - lu'(y-\pi_H l) \\ &\overset{q_L^{RS}>0,\ u \text{ is increasing}}{<} \frac{u(y) - u(y-\pi_H l)}{\pi_H} - lu'(y-\pi_H l) < 0; \end{aligned} \tag{A.3}$$

the last inequality follows because $\frac{u(y)-u(y-\pi_H l)}{\pi_H} - lu'(y-\pi_H l)$ is decreasing in $l > 0$:

$$\frac{\partial}{\partial l}\left\{\frac{u(y)-u(y-\pi_H l)}{\pi_H} - lu'(y-\pi_H l)\right\} = \pi_H l u''(y-\pi_H l) < 0, \tag{A.4}$$

and equal to 0 at $l = 0$. Thus, (A.1) is negative.

Π defined in (5) is increasing in π_H because p^I is increasing in π_H. Indeed, by (4), p^I depends on π_H only through q_L^{RS}; in particular, p^I is increasing in π_H because q_L^{RS} is decreasing in π_H:

$$\frac{\mathrm{d}p^I(\pi_H)}{\mathrm{d}\pi_H} \overset{(4)}{=} \frac{(1-\pi_L)\pi_L}{u'(y-p^I)} \underbrace{\left(u'(y-\pi_L q_L^{RS}) - u'(y-\pi_L q_L^{RS} + q_L^{RS} - l)\right)}_{<0 \text{ by (A.2)}} \frac{\mathrm{d}q_L^{RS}(\pi_H)}{\mathrm{d}\pi_H}, \tag{A.5}$$

which is positive because (A.1) is negative.

Comparative statics with respect to λ

To show that Π defined in (5) is increasing in λ, it is sufficient to show that p^I is increasing in λ.

For $u(x) = \frac{1-\exp(-\lambda x)}{\lambda}$, equation (3) becomes

$$\exp\left((\pi_L q_L^{RS} - \pi_H l)\lambda\right)\left\{1 - \pi_H + \pi_H \exp\left((l - q_L^{RS})\lambda\right)\right\} = 1, \tag{A.6}$$

while equation (4) gives

$$p^I = \pi_L q_L^{RS} + \frac{1}{\lambda}\ln\left(1 - \pi_L + \pi_L \exp\left((l - q_L^{RS})\lambda\right)\right). \tag{A.7}$$

Applying the implicit function theorem to (A.6), we get

$$\frac{\mathrm{d}q_L^{RS}(\lambda)}{\mathrm{d}\lambda} = \frac{1}{\lambda}\left(\frac{\left(\exp\left((l - q_L^{RS})\lambda\right) - 1\right) l(1 - \pi_H)\pi_H}{\exp\left((l - q_L^{RS})\lambda\right)\pi_H(1 - \pi_L) - \pi_L(1 - \pi_H)} - q_L^{RS}\right). \tag{A.8}$$

Differentiating (A.7) and using (A.8), we get

$$\lambda\frac{\mathrm{d}^2 p^I(\lambda)}{\mathrm{d}\lambda^2} + 2\frac{\mathrm{d}p^I(\lambda)}{\mathrm{d}\lambda} = \frac{l^2\pi_L(1-\pi_L)(\pi_H - \pi_L)^3 \exp\left((l - q_L^{RS})\lambda\right)\left(1 - \pi_H + \pi_H \exp\left((l - q_L^{RS})\lambda\right)\right)}{\left(1 - \pi_L + \pi_L \exp\left((l - q_L^{RS})\lambda\right)\right)^2} \times \frac{1 + \pi_H\left(\exp\left(2(l - q_L^{RS})\lambda\right) - 1\right)}{\left(\left(\exp\left((l - q_L^{RS})\lambda\right) - 1\right)\pi_H(1 - \pi_L) + \pi_H - \pi_L\right)^3}, \tag{A.9}$$

which is positive because $q_L^{RS} < l$. Hence, $\frac{\mathrm{d}p^I(\lambda)}{\mathrm{d}\lambda}$ is positive for all $\lambda > 0$ as long as it is positive at the limit $\lambda \to 0$.

Consider the limit $\lambda \to 0$. Applying the Taylor expansion $\exp(x) = 1 + x + O(x^2)$ to (A.6), we get

$$0 = q_L^{RS}(\lambda)(\pi_H - \pi_L)\lambda + O(\lambda^2), \tag{A.10}$$

which implies that $q_L^{RS}(\lambda)\lambda = O(\lambda^2)$; i.e., $q_L^{RS}(\lambda) = q\lambda + o(\lambda)$ for some q independent of λ. Substituting $q_L^{RS}(\lambda) = q\lambda + o(\lambda)$ into (A.7) and applying the Taylor expansion around $\lambda = 0$, we get

$$p^I(\lambda) = l\pi_L + \frac{1}{2}l^2(1 - \pi_L)\pi_L\lambda + O(\lambda^2), \tag{A.11}$$

which implies that $\frac{\mathrm{d}p^I(\lambda)}{\mathrm{d}\lambda}$ is positive at the limit $\lambda \to 0$.

A.2 Proof of Proposition 2

Company 0 chooses price t_0 to maximize

$$\max_{t_0} s_0(t_0+\gamma\Pi) = \frac{\exp\left(-\frac{t_0}{\sigma}\right)}{\exp\left(-\frac{t_0}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)}(t_0+\gamma\Pi). \tag{A.12}$$

$$\text{FOC}: \ \exp\left(-\frac{t_0}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)\left(1-\frac{t_0+\gamma\Pi}{\sigma}\right)=0. \tag{A.13}$$

SOC always holds, so that any solution t_0 to (A.13) is a local maximum.

Company $n \geq 1$ maximizes

$$\max_{t_n} s_n t_n = \frac{\exp\left(-\frac{t_n}{\sigma}\right)}{\exp\left(-\frac{t_n}{\sigma}\right)+\exp\left(-\frac{t_0^*}{\sigma}\right)+(N-1)\exp\left(-\frac{t^*}{\sigma}\right)}t_n. \tag{A.14}$$

$$\text{FOC}: \ \exp\left(-\frac{t_n}{\sigma}\right)+\left(\exp\left(-\frac{t_0^*}{\sigma}\right)+(N-1)\exp\left(-\frac{t^*}{\sigma}\right)\right)\left(1-\frac{t_n}{\sigma}\right)=0. \tag{A.15}$$

SOC always holds, so that any solution t_n to (A.15) is a local maximum.

Denote

$$s^* = \frac{\exp\left(-\frac{t^*}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)}, \quad s_0^* = \frac{\exp\left(-\frac{t_0^*}{\sigma}\right)}{\exp\left(-\frac{t_0^*}{\sigma}\right)+N\exp\left(-\frac{t^*}{\sigma}\right)} \tag{A.16}$$

the equilibrium demand for companies $n=1,\ldots,N$ and company 0, respectively. Then, (A.13) implies (8) and (A.15) implies (9). Definition (A.16) implies that $Ns^*+s_0^*=1$, which gives (10). Expressing s^*/s_0^* from (A.16) yields

$$t_0^* = t^* + \sigma \ln \frac{s^*}{s_0^*}. \tag{A.17}$$

Combining (A.17) with (8) and (9) yields equation (11). Substituting (10) into (11) yields

$$\frac{(N+1)s_0^*-1}{(1-s_0^*)(N-1+s_0^*)} - \ln\frac{1-s_0^*}{Ns_0^*} = \frac{\gamma\Pi}{\sigma}. \tag{A.18}$$

The left-hand side of (A.18) 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 (A.18) exists and is unique for any $\Pi \geq 0$.

A.3 Proof of Lemma 1

Consumer welfare is

$$W = \mathbb{E}\left[\max_n V_n\right] = \int_{-\infty}^{+\infty} v f(v)\,\mathrm{d}v, \tag{A.19}$$

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

$$\begin{aligned}
\Pr\left(\max_n V_n < v\right) &= \Pr\left(\max_n \mu_n \sigma - t_n < v - V\right) \\
&\overset{(2)}{=} \prod_{n=0}^{N} \exp\left(-\exp\left(-\frac{v - V + t_n}{\sigma} - \text{Euler's constant}\right)\right) \\
&\overset{t_n = t^*,\ t_0 = t_0^*}{=} \exp\left(-\exp\left(-\frac{v - V}{\sigma} - \text{Euler's constant}\right) S^*\right),
\end{aligned} \tag{A.20}$$

where

$$S^* = \exp\left(-\frac{t_0^*}{\sigma}\right) + N \exp\left(-\frac{t^*}{\sigma}\right). \tag{A.21}$$

Then,

$$f(v) = \frac{S^*}{\sigma} \exp\left(-\frac{v - V}{\sigma} - \text{Euler's constant}\right) \exp\left(-\exp\left(-\frac{v - V}{\sigma} - \text{Euler's constant}\right) S^*\right), \tag{A.22}$$

and so, the change of variables $x = \exp\left(-\frac{v-V}{\sigma} - \text{Euler's constant}\right) S^*$ in (A.19) yields

$$W = \int_0^{+\infty} \left(\sigma \ln \frac{S^*}{x} + V - \sigma \text{Euler's constant}\right) \exp(-x)\,\mathrm{d}x = V + \sigma \ln S^*. \tag{A.23}$$

Substituting (A.21) into (A.23) yields (12).

A.4 Comparative Statics

In this Appendix, we prove Proposition 4, Theorem 1, Propositions 5 and 6, as well as additional results presented in Table A.1.[36]

Let R_0 be company 0's equilibrium profit across both markets and R be the 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.

[36] The comparative statics with respect to γ are 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 (A.18) (for notational simplicity, we sometimes omit the star in s_0^*). Applying the implicit function theorem to (A.18), we get

$$s_0'\left(\frac{\gamma\Pi}{\sigma}\right) = \frac{(1-s_0)^2 s_0 (N-1+s_0)^2}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} > 0, \tag{A.24}$$

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 $\Pi = 0$ follows immediately from Proposition 2. Since the left-hand side of (A.18) 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.24) yield

$$t_0'(\Pi) = -\gamma \frac{(1-s_0)(N-1+s_0)^2 + N(1-s_0)^2 s_0}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} < 0. \tag{A.25}$$

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 (A.18) 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

$$R_0 = \frac{\sigma s_0^*}{1 - s_0^*}. \tag{A.26}$$

Expression (A.26) 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

$$R = \frac{\sigma s^*}{1 - s^*}. \tag{A.27}$$

Expression (A.27) 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.27), we get the expression for the joint profit as a function of s_0^*:

$$R_0 + NR = \frac{\sigma s_0^*}{1 - s_0^*} + N\frac{\sigma(1 - s_0^*)}{N - 1 + s_0^*}. \tag{A.28}$$

The right-hand side of (A.28) 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 (12), then s^* from (10), and then Π from (A.18), we get the expression for consumer welfare as a function of s_0^*:

$$W = V + \sigma\left(\ln\frac{N}{1 - s_0^*} - \frac{N}{N - 1 + s_0^*}\right). \tag{A.29}$$

Expression (A.29) increases in s_0^*. Thus, since s_0^* increases in Π from $1/(N+1)$ to 1, 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

$$\Delta_W = W(\Pi) - W(0) > 0, \tag{A.30}$$

company 0's change in profit as

$$\Delta_{R0} = R_0(\Pi) - R_0(0) > 0, \tag{A.31}$$

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

$$\Delta_{RN} = NR(\Pi) - NR(0) < 0. \tag{A.32}$$

In (A.30), (A.31) and (A.32), 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 follow from the comparative statics of R_0, R and W.

Comparative statics with respect to N

Let $s_0(N)$ be the solution to (A.18). Applying the implicit function theorem to (A.18), we get

$$s_0'(N) = -\frac{(1-s_0)^2 s_0 \left(N(1-s_0)+(N-1+s_0)^2\right)}{N\left(N(1-s_0)^2 s_0+(N-1+s_0)^2\right)} < 0, \tag{A.33}$$

so that s_0^* is decreasing in N. Since for any fixed $s_0^* \in (0,1)$, the left-hand side of (A.18) 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):

$$s'(N) = -\frac{1-s_0+Ns_0'(N)}{N^2} \overset{(A.33)}{=} -\frac{(1-s_0)(N-1+s_0)^2(1-s_0+s_0^2)}{N^2\left(N(1-s_0)^2 s_0+(N-1+s_0)^2\right)} < 0, \tag{A.34}$$

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.26) and (A.27), respectively, and their limits are $\sigma-\gamma\Pi$, σ, 0 and 0.

Differentiating the joint profit (A.28) with respect to N, we get

$$\frac{\mathrm{d}}{\mathrm{d}N}\left(R_0(N)+NR(N)\right) = -\frac{\sigma\left((1-s_0)^4-((N-1)(1-s_0)+N)((N+1)s_0-1)s_0'(N)\right)}{(1-s_0)^2(N-1+s_0)^2}, \tag{A.35}$$

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.28) goes to σ.

Differentiating consumer welfare (A.29) with respect to N, we get

$$W'(N) = \frac{\sigma\left(N(1-s_0)+(N-1+s_0)^2\right)\left(1-s_0+Ns_0'(N)\right)}{N(1-s_0)(N-1+s_0)^2}, \tag{A.36}$$

which is positive by (A.34). At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.29) goes to $+\infty$.

Substituting R_0 from (A.26) into (A.31) and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_{R0} = \frac{\sigma s_0^*}{1-s_0^*} - \frac{\sigma}{N}. \tag{A.37}$$

Differentiating (A.37) with respect to N and using (A.33), we get

$$\Delta'_{R0}(N) = -\frac{(N-1)(N-1+s_0)((N+1)s_0-1)\sigma}{N^2\,(N(1-s_0)^2 s_0 + (N-1+s_0)^2)} < 0, \tag{A.38}$$

so that Δ_{R0} is decreasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.37) goes to 0.

Substituting R from (A.27) into (A.32), then s^* from (10), and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_{RN} = \frac{\sigma(1-s_0^*)}{1-(1-s_0^*)/N} - \sigma. \tag{A.39}$$

Differentiating (A.39) with respect to N and using (A.33), we get

$$\Delta'_{RN}(N) = \frac{(N-1)(1-s_0)^2((N+1)s_0-1)\sigma}{(N-1+s_0)\,(N(1-s_0)^2 s_0 + (N-1+s_0)^2)} > 0, \tag{A.40}$$

so that Δ_{RN} is increasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.39) goes to 0.

Substituting W from (A.29) into (A.30) and using the result that $s_0^* = 1/(N+1)$ if $\Pi = 0$, we get

$$\Delta_W = \sigma\left(\ln\frac{1-1/(N+1)}{1-s_0^*} + \frac{N}{N-1+1/(N+1)} - \frac{N}{N-1+s_0^*}\right). \tag{A.41}$$

Differentiating (A.41) with respect to N and using (A.33), we get

$$\Delta'_W(N) = -\frac{\big((N^3-1+N(1-s_0)^2)(1-s_0)+N\big)((N+1)s_0-1)\sigma}{N^2(N+1)\,(N(1-s_0)^2 s_0 + (N-1+s_0)^2)} < 0, \tag{A.42}$$

so that Δ_W is decreasing in N. At the limit $N \to +\infty$, $s_0^* \to 0$, and so, (A.41) goes to 0.

Comparative statics with respect to σ

By (A.24), s_0^* is decreasing in σ. Hence, by (10), s^* is increasing in σ. Since by (A.18), 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.27), 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.27) converge to 0. As $\sigma \to +\infty$, $s^* \to 1/(N+1)$, and so, (9) and (A.27) go to $+\infty$.

Differentiating (8) with respect to σ, using (A.24) and substituting Π from (A.18) yield

$$t_0'(\sigma) = \frac{\frac{N^2}{(N-1+s_0)^2} + \frac{1}{s_0} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \tag{A.43}$$

The numerator in (A.43) 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.43) 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 (A.18), it follows that

$$\lim_{\sigma \to 0} \frac{\sigma}{1 - s_0^*(\sigma)} = \gamma \Pi. \tag{A.44}$$

If $\Pi = 0$, then $s_0^* = 1/(N+1)$, and so (A.44) also holds. Then, (A.44) implies that (8) converges to 0 as $\sigma \to 0$.

Differentiating (A.26) with respect to σ, using (A.24) and substituting Π from (A.18) yield

$$R_0'(\sigma) = \frac{\frac{N\left(N-(1-s_0)^2\right)}{(N-1+s_0)^2} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \tag{A.45}$$

The numerator in (A.45) 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.45) is negative and then positive. If $\sigma \to +\infty$, then $s_0^* \to 1/(N+1)$, and so, (A.26) goes to $+\infty$. If $\Pi > 0$ and $\sigma \to 0$, then (A.26) converges to $\gamma \Pi$ because $s_0^* \to 1$ and (A.44) holds. If $\Pi = 0$ and $\sigma \to 0$, then $s_0^* = 1/(N+1)$ and so (A.26) goes to 0.

Differentiating the joint profit (A.28) with respect to σ, using (A.24) and substituting Π from (A.18) yield

$$\begin{aligned} R_0'(\sigma) + NR'(\sigma) = {} & \frac{s_0\left((N-1)(1-s_0)+N\right)\left((N+1)s_0-1\right)}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} \times \\ & \left(\frac{N\left((N+1)\left(1+(1-s_0)s_0^2\right)-2+s_0\right)}{s_0\left((N-1)(1-s_0)+N\right)\left((N+1)s_0-1\right)} + \ln\left(\frac{1-s_0}{Ns_0}\right)\right). \end{aligned} \tag{A.46}$$

The numerator in (A.46) 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.46) is negative and then positive.

Differentiating consumer welfare (A.29) with respect to σ, using (A.24) and substituting Π from

(A.18) yield $W'(\sigma) = w(N, s_0)$, where

$$w(N,s_0) = \frac{(N-1+s_0)^2\,(1-(1-s_0)s_0)\ln\left(\frac{Ns_0}{1-s_0}\right) - (1-s_0)^2 s_0 - N(N-1+s_0)\left(1+s_0^2\right)}{N(1-s_0)^2 s_0 + (N-1+s_0)^2} - \ln s_0. \quad \text{(A.47)}$$

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

$$\frac{\partial w(N,s_0)}{\partial N} = \frac{1-(1-s_0)s_0}{(N(1-s_0)^2 s_0 + (N-1+s_0)^2)^2}\left\{N^2\,(N-(3-s_0)(1-s_0)) + \frac{(1-s_0)^4}{N}\right.$$
$$+ 2N(1-s_0)^2\,(1-(1-s_0)s_0) + \left(2N-(1-s_0)(3+s_0^2)\right)(1-s_0)^2$$
$$\left. + (1-s_0+N)\,(1-s_0)^2 s_0 (N-1+s_0)\ln\left(\frac{Ns_0}{1-s_0}\right)\right\} > 0, \quad \text{(A.48)}$$

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.29) converges to V because (A.44) holds, $s_0^* \to 1$ 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.29) goes to $+\infty$ if $N \geq 3$.

Differentiating (A.31) with respect to σ, using (A.45) and substituting $s_0^* = 1/(N+1)$ for $\Pi = 0$, we get

$$\Delta'_{R0}(\sigma) = \frac{1 + \frac{1}{N} - \frac{((N+1)s_0-1)^2}{N(N-1+s_0)^2} - \frac{1}{Ns_0} + \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{N(1-s_0)^2}{(N-1+s_0)^2} + \frac{1}{s_0}}. \quad \text{(A.49)}$$

The numerator in (A.49) is decreasing in $s_0 \in (1/(N+1), 1)$ and equal to 0 at $s_0 = 1/(N+1)$. Hence, (A.49) 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 (A.18), it follows that

$$\lim_{\sigma\to+\infty} \Delta_{R0}(\sigma) \overset{\text{(A.37)}}{=} \lim_{\sigma\to+\infty} \frac{\sigma\left((N+1)s_0^*(\sigma)-1\right)}{N\left(1-s_0^*(\sigma)\right)} = \frac{\gamma\Pi N}{N^2+N+1}. \quad \text{(A.50)}$$

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

Differentiating (A.39) with respect to σ, using (A.24) and substituting Π from (A.18) yield

$$\Delta'_{RN}(\sigma) = \frac{\frac{1}{s_0}\left(\frac{s_0^2}{1-s_0} - \frac{1}{N}\right)\left(\frac{1}{N} - \frac{1}{1-s_0}\right) - 1 - \ln\left(\frac{1-s_0}{Ns_0}\right)}{\frac{1}{N} + \frac{(N-1+s_0)^2}{N^2(1-s_0)^2 s_0}}. \quad \text{(A.51)}$$

The numerator in (A.51) 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

$$\frac{\left(s_0-\frac{1}{N+1}\right)\left(s_0+\frac{3N-2+\sqrt{(9N-8)N}}{2}\right)}{N^2(1-s_0)^3s_0^2}\left(1+\underbrace{\frac{\frac{2N}{N+1}\left((N-1)^2-3\right)}{\sqrt{(9N-8)N}+\frac{N(1+3N)}{N+1}}}_{>0\text{ if }N\geq 3}-(N+1)s_0\right). \tag{A.52}$$

Hence, if $N \geq 3$, then, as a function of $s_0 \in (1/(N+1), 1)$, the numerator in (A.51) is positive and then negative. Since s_0^* is decreasing in $\sigma \in (0, +\infty)$ from 1 to $1/(N+1)$, (A.51) is negative and then positive as a function of $\sigma \in (0, +\infty)$. To find the limit $\sigma \to +\infty$, observe that from (A.37) and (A.39), we get

$$\frac{\Delta_{RN}}{\Delta_{R0}} = -\frac{N(1-s_0^*)}{N-1+s_0^*}. \tag{A.53}$$

As $\sigma \to +\infty$, $s_0^* \to 1/(N+1)$, and so, (A.53) 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.41) with respect to σ twice and using (A.24) yield

$$\Delta_W''(\sigma) = \frac{\gamma^2\Pi^2(N-1+s_0)^3(1-s_0)^2 s_0 \tilde{w}(N,s_0)}{\sigma^3\left(N(1-s_0)^2 s_0 + (N-1+s_0)^2\right)^3}, \tag{A.54}$$

where

$$\tilde{w}(N,s_0) = N(1-s_0)\big((1-s_0)\big(N(1+s_0^2)-(1-s_0)^2(1+s_0)\big)-2s_0N\big)+(1-2s_0)(N-1+s_0)^3. \tag{A.55}$$

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

$$\begin{aligned}\frac{\partial \tilde{w}(N,s_0)}{\partial s_0} = &-\frac{5+6(1-2s_0)^2+5(1-2s_0)^4}{16}-(N-1)\left(2s_0+5(1-s_0)^4\right)\\ &-(N-1)^2\left(2s_0(1-s_0)(1+2s_0)+(1-2s_0)^2\right)-2\left(N-1+s_0\right)^3<0,\end{aligned} \tag{A.56}$$

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

$$\tilde{w}(N,1) = -N^3, \quad \tilde{w}\left(N, \frac{1}{N+1}\right) = \frac{N^3(1+N+N^2)\left((N-2)(N^2+2)+2(N-1)^2\right)}{(N+1)^5}. \tag{A.57}$$

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.41) with respect to σ, using (A.24) and substituting Π from (A.18) yield

$$\Delta_W'(\sigma) = 1 + \frac{1}{N} - \frac{N}{N-1+s_0} + \ln\left(\frac{N}{(N+1)(1-s_0)}\right) - \frac{N(1-s_0)+(N-1+s_0)^2}{N(1-s_0)+\frac{(N-1+s_0)^2}{(1-s_0)s_0}}\left(\frac{(N+1)s_0-1}{(1-s_0)(N-1+s_0)} - \ln\left(\frac{1-s_0}{Ns_0}\right)\right). \quad \text{(A.58)}$$

At $s_0 = 1/(N+1)$, (A.58) is equal to 0. At the limit $s_0 \to 1$, (A.58) goes to $+\infty$. Hence, $\Delta_W(\sigma)$ is increasing and then decreasing. To find the limit $\sigma \to +\infty$, observe that from (A.37) and (A.41), 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 For Online Publication: 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, 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 consumers, we need to compare (B.1) with Δ_W defined in (A.30), the consumer welfare gain from data linkage without data sharing. Theorem B.1 shows that, under an additional restriction $u'(0) \le 1$,[37] 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 $\sigma > \hat{\sigma}$.*

[37] 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 the 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 a monopolistic insurance market was derived in Stiglitz (1977). Figure B.1 illustrates the equilibrium in the same space as in Figure 2.

If the monopolist knows its consumer's risk type, then it 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 her risk type, indifferent to buying the insurance; that is, the monopolist fully extracts consumer surplus.

If the monopolist does not know its consumer's risk type, 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 low-risk consumers in the population, is below a certain threshold γ^*, the monopolist does not serve the low-risk consumer at all and offers the high-risk consumer a full-insurance contract at a premium that makes her 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 contracts for the low-risk and high-risk consumers correspond to the black point and red point II, respectively.

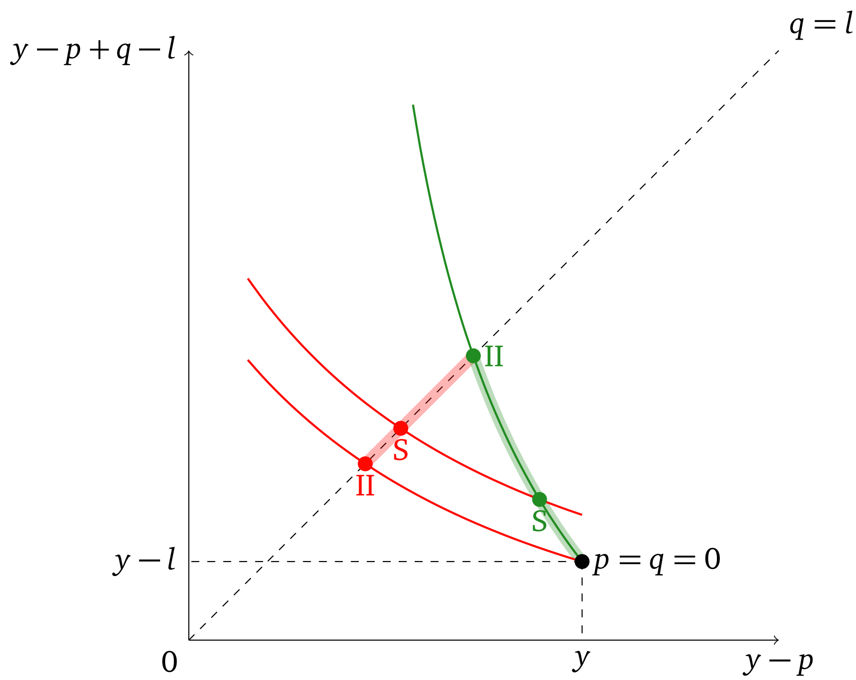


Figure B.1: Equilibrium in the monopolistic insurance market, drawn for the utility function $u(x) = 1-e^{-x}$. Red and green S points correspond to the Stiglitz equilibrium contract for the high- and low-risk consumer, respectively. Red and green points II correspond to the contract that the informed 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 her 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 high-risk consumers; that is, the incentive-compatibility constraint of 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 offered contracts maximizes the monopolist's expected profit, and so satisfies the additional optimality condition:[38]

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

[38]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 the 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 the green point II. At the same time, the red point S, which corresponds to the high-risk contract, moves up along the highlighted segment of the 45-degree line from the red point II to the 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, 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, 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 her type does not affect the offered contract — she receives full insurance at a premium that makes her indifferent to buying insurance (see the 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 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 the green point II and to the 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 the competitive insurance market. Data linkage promotes contract efficiency for 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 pronounced than in a competitive market, because without data linkage, the monopolistic insurer does not serve low-risk consumers at all, and so data linkage opens the insurance market to new consumers.

High γ

When $\gamma > \gamma^*$, high-risk consumers get a disutility from revealing their type — instead of getting a contract marked by the red point S, they get one marked by the red point II in Figure B.1. In equilibrium, 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 low-risk and with probability s_0^H if the consumer is high-risk. Then, conditional on company 0 not knowing the consumer's risk type, the probability that the consumer is 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 her risk type, the high-risk consumer pays premium $p_H(\gamma')$ and, relative to revealing her 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 that 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 rearrangements of equations in (B.17) when $t_0 = t_0^*$. □

Equilibrium, as 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, $\gamma' < \gamma$

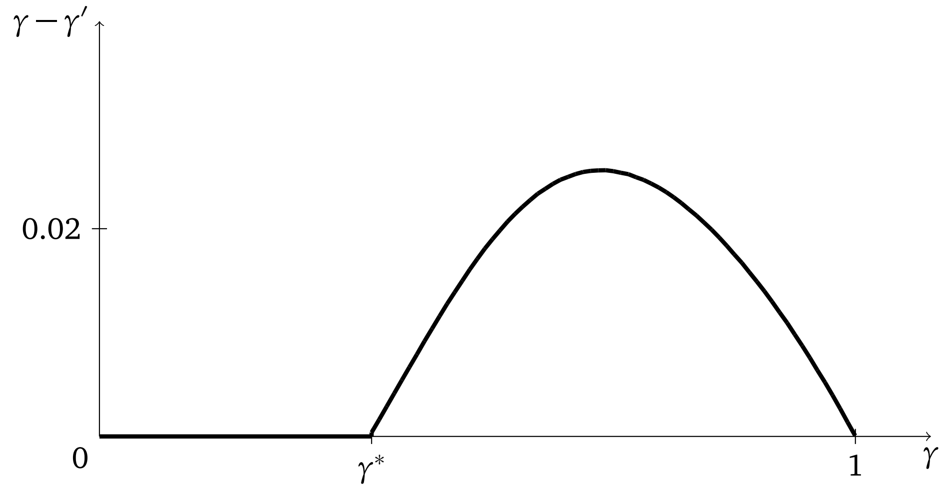


Figure B.2: The difference between γ, the share of low-risk consumers, and γ', the share of low-risk consumers among those consumers that company 0 does not serve in the product market. Parameters: $u(x) = 1 - e^{-x}$, $l = y = 1$, $\sigma = 0.1$, $N = 5$, $\pi_H = 0.8$, $\pi_L = 0.6$.

indicates that low-risk consumers, who have no incentives to hide their type from company 0, pose a negative externality to high-risk consumers, who suffer a disutility from revealing their type. Indeed, when the high-risk consumer does not reveal her risk type to company 0, data linkage changes her 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 consumer welfare as a result of data linkage.

Proposition B.2. *As a result of data linkage, 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 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 of Lemma 1, we derive consumer welfare in the product market with data linkage:

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

where

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

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

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

Without data linkage, 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, the consumer welfare gain in the product market is

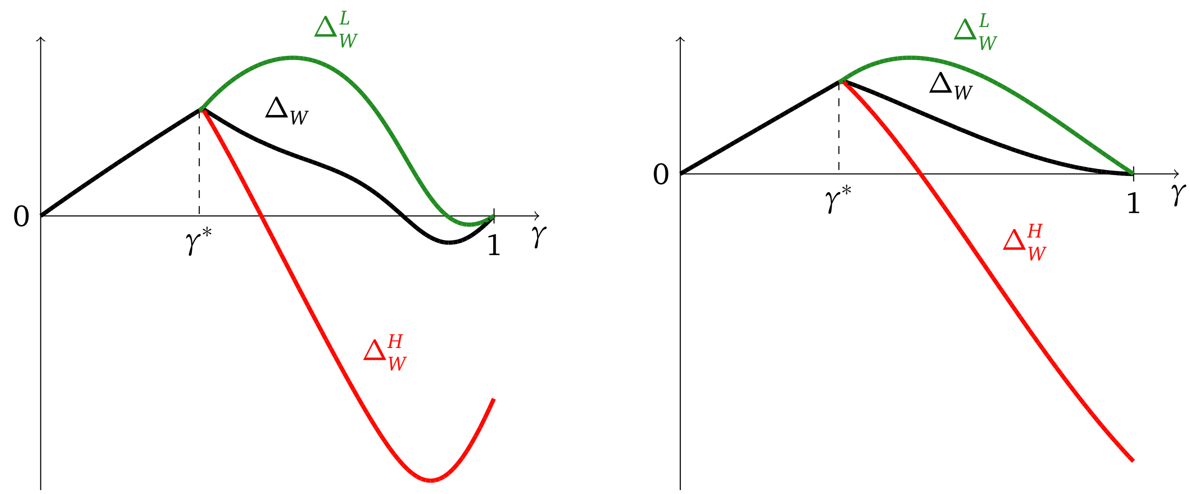
$$W^i_{\text{linked}} - W^i_{\text{independent}} = \sigma\left(\ln \frac{N}{(N+1)(1-s_0^i)} + \frac{N+1}{N}\right) - t^*. \tag{B.32}$$

In the insurance market, as a result of data linkage, low-risk consumers' welfare does not change, while high-risk consumers' welfare changes from $u(y - p_H(\gamma))$ to $u(y - p_H(\gamma'))$.[39] Together with (B.32), it gives us (B.25) and (B.26). □

Proposition B.2 implies that low-risk consumers gain from data linkage more than high-risk consumers do; that is, $\Delta_W^L > \Delta_W^H$. 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, low-risk consumers demand more variety 0 than high-risk consumers demand. Hence, the direct comparison of (B.25) and (B.26) gives $\Delta_W^L > \Delta_W^H$.

A priori, it is not clear whether consumers benefit from data linkage; that is, the signs of Δ_W^L and Δ_W^H are ambiguous. Figure B.3 shows that both low- and high-risk consumers may benefit from data linkage; that is, both Δ_W^L and Δ_W^H may be positive. According to the figure, both Δ_W^L and Δ_W^H are positive if γ is sufficiently close to γ^*. If γ is high, high-risk consumers are worse off from data linkage, while whether the welfare gain of low-risk consumers is positive depends on other parameters. Figure B.3a demonstrates that low-risk consumers' 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 is sufficient for making the low-risk consumer's, and even the average consumer's, welfare gain positive for all γ. All

[39] Data linkage changes the utility of the 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 her risk type to company 0. If the high-risk consumer reveals her risk type to company 0, her 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 a 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.

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

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

Substituting (B.43) into

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

we get

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

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

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

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,

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

Therefore,

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

The asymptotics (B.47) and (B.48) imply 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

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

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

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

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

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

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

we get

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

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

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,

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

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

$$N\Delta_W^L = t^*\left(\frac{N}{s_0^L}\left(\frac{\sigma}{t^*} - 1 + \frac{1}{N}\right)s_0^L\left(\ln\frac{N}{(N+1)(1-s_0^L)} + \frac{N+1}{N}\right) - \frac{1}{N} - (N-1)\ln\left(\left(1+\frac{1}{N}\right)\left(1-\frac{Ns_0^L}{N}\right)\right)\right), \tag{B.67}$$

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 that $s_0^L \to 1$. 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

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

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

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

□

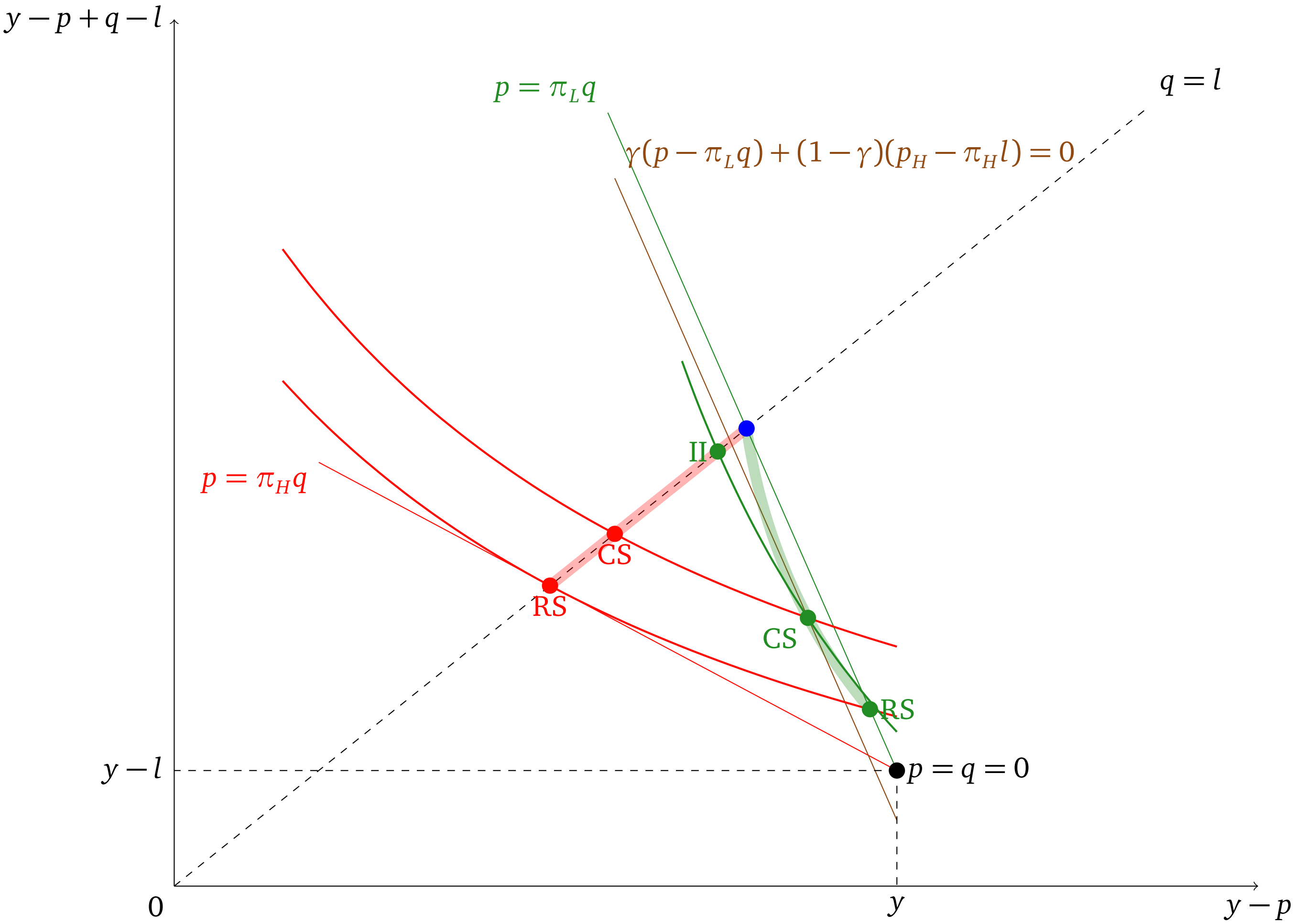


Figure B.4: Cross-subsidy equilibrium in the insurance market, drawn for the utility function $u(x) = 1-e^{-x}$. Red and green RS points correspond to the RS equilibrium contract for the high- and low-risk consumer, respectively. For sufficiently high γ, red and green CS points correspond to the cross-subsidy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

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

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

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

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

At the optimum, the incentive-compatibility constraint of 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 high-risk consumers, and not the other way round. Whether this constraint binds depends on the proportion of low-risk consumers, γ.

If γ is below a 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 low-risk consumers subsidize 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 the 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 that company 0 observes the consumer's type and, thus, becomes an informed insurer. If $\gamma \leq \hat{\gamma}$, all the analysis from the baseline model remains valid. Suppose that $\gamma > \hat{\gamma}$. To avoid making losses on a contract, the informed insurer does not serve high-risk consumers. To 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 uninformed 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 a 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 that $\gamma > \hat{\gamma}$.

Let s_0 be the demand for variety 0. Then, the uninformed insurance companies face a population with a proportion of low-risk consumers equal to[40]

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

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 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 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

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

[40] 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 is immaterial to our analysis because the informed insurer can never make positive profit on 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.2. In particular, company n's optimization problem is the same as in Section A.2. Combining (A.17) with (B.85), (B.86) and (B.87) yields equation (B.88) for equilibrium s_0^*. □

Welfare implications 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

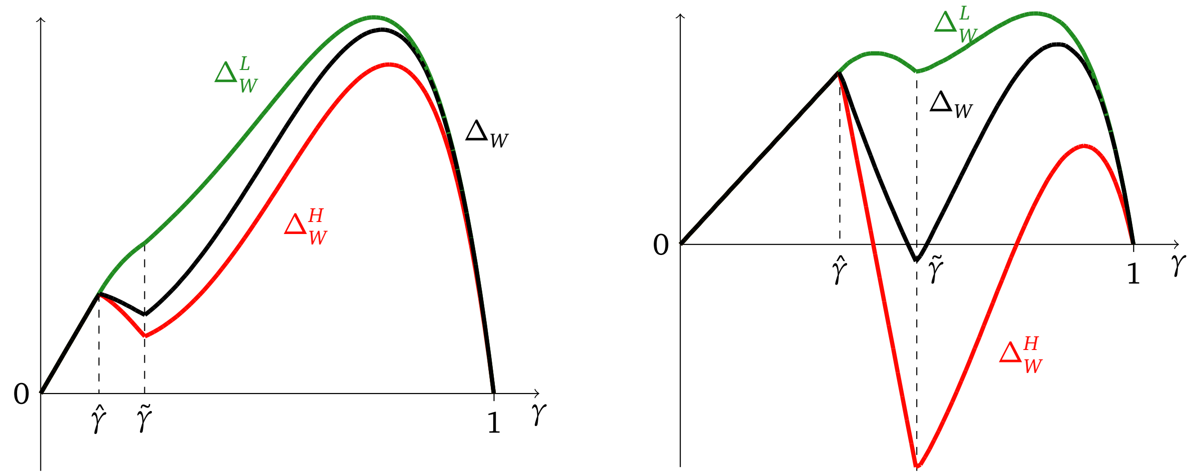
$$H(\gamma)=-\frac{\gamma^2(1-\gamma)(1-\pi_L)\pi_Lu'(y-p_L)}{\pi_H-\pi_L}\left(\pi_H\frac{u''(y-p_L)}{u'(y-p_L)^3}+(1-\pi_H)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^3}\right)$$
$$-\frac{\gamma^3(1-\pi_L)\pi_Lu'(y-p_H)}{(\pi_H-\pi_L)u'(y-p_L+q_L-l)}\left(\pi_L\frac{u''(y-p_L)}{u'(y-p_L)^2}+(1-\pi_L)\frac{u''(y-p_L+q_L-l)}{u'(y-p_L+q_L-l)^2}\right)$$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

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 consumer welfare in different markets.

In the insurance market, the welfare of consumers is determined by the offers of the uninformed companies. Hence, as a result of data linkage, 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 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, consumer welfare is given in (A.29); that is, the consumer welfare expression from the baseline model remains valid. Hence, data linkage increases consumer welfare of both types equally and by amount (A.41), 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 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 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 insurers, 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

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

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

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 $n = 1, \dots, N$ 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 (11) 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), $s_0^* = \bar{s}$. □

Lemma B.4 gives the expression for consumer welfare.

Lemma B.4. *In the product market, in equilibrium, 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 Lemma 1. □

Table B.1 summarizes the comparative statics results for the model with the outside option. We focus on those results that are similar to our benchmark model without the outside option, omitting the comparative statics exercises which are not robust to the introduction of the outside option. According to Table B.1, the results of Theorem 1 and Proposition 4 are fully robust, while Propositions 5 and 6 are partially robust to the introduction of the outside option.

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.

	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 the 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.

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.26) and (A.27). 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.26), 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.

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.26) and (A.27), respectively, and their limits are $\sigma-\gamma\Pi$, σ, 0 and 0.

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.26), company 0's change in profit, defined in (A.31), 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.30), 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.32). 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.27), the joint profit of companies $n=1,\dots,N$ 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 that $V>0$.

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

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

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.