Regulating Cyberwashing: False Claims, Hacker Threats, and Price Signaling

Tushan Huang[1], Tao Lu[2], Hyoduk Shin[3], Ke Fu[1]
[1]Sun Yat-sen University
[2]University of Connecticut
[3]University of California San Diego
Tushan Huang: huangtsh5@outlook.com; Tao Lu: tao.lu@uconn.edu
Hyoduk Shin: hdshin@ucsd.edu; Ke Fu: fuke@mail.sysu.edu.cn

Firms increasingly make cybersecurity claims that consumers cannot verify, creating opportunities for deceptive "cyberwashing." We develop a signaling model in which a firm privately observes its cybersecurity level and chooses both a cybersecurity claim and a price. Consumers infer cybersecurity from these signals, hackers decide whether to attack based on the expected payoff from compromising the firm's consumer base, and regulators impose expected penalties on false high-security claims. The model highlights a feedback channel absent from standard false-claim settings: cyberwashing increases demand but also increases attack exposure. Because expanding the customer base generates a smaller marginal expected breach loss for the secure firm, demand expansion through a lower price becomes a credible signal of security, a distinctive property arising from the endogenous attack channel in our model. Under the selected equilibrium, cyberwashing arises when the false-claim penalty is low and the hacker payoff-to-cost ratio is moderate, whereas a high penalty or a sufficiently low ratio sustains truthful disclosure; at high ratios, separation may fail to reemerge when breach losses and false-claim penalties are small. Consumer surplus is highest for penalties just above the level that deters cyberwashing, because a stronger penalty reduces the need for downward price signaling and raises prices. When low penalties sustain cyberwashing, welfare approaches its supremum just above the deterrence threshold under weak hacker threats; under severe threats, a strictly higher penalty uniquely maximizes welfare by curbing excessive demand expansion and breach exposure. Finally, broader cybersecurity governance has non-monotonic effects: policies that lower the hacker payoff-to-cost ratio or make vulnerabilities harder to discover may reduce welfare by weakening the disciplining effect of cyber threats and creating a cyberwashing trap. Effective anti-cyberwashing policy therefore requires coordinating false-claim regulation with cybersecurity governance, accounting for firms' pricing responses and hackers' attack incentives.

Key words: Cyberwashing, signaling game, false-claim regulation, endogenous hacker attacks, cybersecurity policy

1. Introduction

As digital services increasingly rely on consumer data, cybersecurity has become a central component of consumer trust. Data breaches can expose customers to financial fraud, identity theft, privacy loss, and the time required to restore accounts or dispute unauthorized transactions. Consumers are increasingly attentive to these risks: In PwC's 2024 global consumer survey, 83% of

respondents identified personal-data protection as crucial to earning their trust (PwC, 2024). Another recent consumer privacy survey reports that 75% of respondents would not purchase from an organization they do not trust with their data (Cisco, 2024). However, although customers can compare the price, functionality, speed, and ratings of a digital service before purchasing it, they cannot as readily inspect the controls that determine whether their personal, financial, and behavioral data will remain secure. Consumers must therefore make purchasing decisions before they can reliably assess the service's cybersecurity. They use the firm's claims and price to infer its underlying protection. This information gap creates room for cybersecurity claims to affect demand when consumers cannot fully verify whether those claims accurately reflect the firm's underlying protection.

When buyers value protection but cannot inspect the controls behind it, firms have an incentive to appear more secure than they are. Indeed, cyberwashing, whereby firms exaggerate or misrepresent their cybersecurity practices to appear more secure than they are, has become an increasingly pressing regulatory concern. A number of recent incidents have been uncovered and publicized through independent technical assessments, whistleblower reports, and regulatory investigations. For example, an independent assessment by Citizen Lab found that Zoom's actual encryption practices did not match the protections it publicly claimed to provide.[1] Zoom also faced allegations from the Federal Trade Commission (FTC) that it misrepresented the security of its services by claiming to offer end-to-end, 256-bit encryption for users' communications when, in fact, it provided a lower level of protection.[2] More recently, the FTC alleged that GoDaddy, a web-hosting company, misled consumers about the extent of its data-security protections through claims made on its websites, in email marketing, and in social media advertisements.[3] Cyberwashing is also attracting regulatory attention in other jurisdictions. For example, despite the absence of cyberwashing-specific legislation in Australia, the Australian Competition and Consumer Commission (ACCC) can address misleading cybersecurity claims under existing consumer-protection laws.[4] These developments illustrate how cyberwashing can be brought to light and, once exposed, may lead to regulatory

[1] https://citizenlab.ca/research/move-fast-roll-your-own-crypto-a-quick-look-at-the-confidentiality-of-zoom-meetings/ (accessed on August 4, 2026).

[2] https://www.ftc.gov/news-events/news/press-releases/2020/11/ftc-requires-zoom-enhance-its-security-practices-part-settlement (accessed on August 10, 2026).

[3] https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-takes-action-against-godaddy-alleged-lax-data-security-its-website-hosting-services (accessed on August 10, 2026).

[4] See https://www.corrs.com.au/insights/ai-washing-and-cyber-washing-key-legal-and-regulatory-enforcement-risks-for-australian-organisations (accessed on January 5, 2026).

action, and, in some cases, substantial monetary costs for firms.[5] Furthermore, they reveal two features of the economic problem: Consumers and buyers have limited ability to verify security claims before transacting, and regulatory correction often occurs only after the claim has already influenced market behavior. Regulation can reduce the return from overstatement, but its deterrent force depends on both the likelihood that a false claim is detected and the consequence imposed when it is.

Consumers and regulators are not the only parties affected by the market created around a firm's cybersecurity claims. Financially motivated malicious attackers—hereafter, hackers—are another player in this market. The cybercrime economy includes specialized participants that acquire access at scale, assess potential victims, and sell access to others. A joint report by the United Kingdom's National Cyber Security Centre and National Crime Agency documents that initial compromise is often opportunistic, but that access brokers subsequently filter firms for expected profitability using indicators such as network size, number of users, and company value (National Cyber Security Centre and National Crime Agency, 2023). The monetization of customer data is also visible in practice. Mandiant (2024) reports that the UNC5537 campaign systematically extracted data from Snowflake customer instances, attempted to extort affected organizations, and advertised stolen customer records for sale. Meanwhile, Verizon's 2026 Data Breach Investigations Report states that vulnerability exploitation initiated 31% of the breaches it studied and that attackers' use of AI had reduced the window for exploiting known vulnerabilities from months to hours (Verizon, 2026). This evidence suggests that a firm's installed customer and data footprint can affect the economic prize from compromise, even when the initial point of access is discovered opportunistically. The same customer base is therefore an asset to the firm and a potential prize to hackers.

This interaction raises three questions. First, when does a firm engage in cyberwashing? Second, how does attack exposure affect the credibility of security claims and the penalty required to deter cyberwashing? Third, how should false-claim enforcement be coordinated with policies that influence hackers' incentives? We address these questions using a game theoretic model in which a firm privately knows its cybersecurity quality and chooses a public security claim and price. Consumers observe both choices, infer the firm's security, and decide whether to purchase based on expected breach harm and price. Hackers decide whether to attack based on their costs and the potential return from compromising the firm's customer base; in the baseline model, the claim and

[5] Following an internal whistleblower lawsuit, Illumina agreed to pay $9.8 million to resolve allegations that it sold genomic sequencing systems with cybersecurity vulnerabilities while falsely representing compliance with cybersecurity standards; see https://www.justice.gov/opa/pr/illumina-inc-pay-98m-resolve-false-claims-act-allegations-arising-cybersecurity. LifeLock likewise agreed to pay $12 million to settle FTC and state allegations involving deceptive identity-theft-prevention and data-security claims; see https://www.ftc.gov/news-events/news/press-releases/2010/03/lifelock-will-pay-12-million-settle-charges-ftc-35-states-identity-theft-prevention-data-security.

the price therefore affect attacks through the customer base they generate, and an extension adds a second channel in which hackers also infer the likelihood of an exploitable vulnerability from the firm's claim. A low-security firm that makes a high-security claim faces an expected regulatory penalty. Because cyberwashing can expand demand while a larger customer base increases attack exposure, the model links firms' security disclosure and pricing decisions to consumer demand, hacker behavior, and welfare.

Our analysis yields three sets of insights. First, cyberwashing arises when appearing secure creates a meaningful demand advantage while the resulting attack exposure remains tolerable for the low-security firm. When hackers have little incentive to attack, security differences barely affect demand and there is little to gain from overclaiming; at intermediate threat levels, cyberwashing becomes profitable; at high levels, truthful disclosure may fail to reemerge within the admissible range when breach losses and false-claim penalties are small. Second, attack exposure gives price a signaling role. A high-security firm can make its claim credible by lowering its price and expanding its customer base, creating an exposure that a vulnerable firm is unwilling to imitate. False-claim penalties substitute for this costly price signaling, so consumer surplus is highest for penalties just above the level that deters cyberwashing: beyond that level, a stronger penalty mainly allows the secure firm to withdraw its discount. Social welfare may favor a strictly higher penalty when hacker threats are severe, because the signaling discount then expands the secure firm's customer base beyond the level that is efficient once attack exposure is taken into account. Third, false-claim regulation and cybersecurity governance interact. Measures that reduce hackers' incentives or make vulnerabilities harder to discover can reduce attack exposure, but they can also induce cyberwashing by weakening its disciplining effect. Moreover, a lower hacker payoff-to-cost ratio can reduce consumer surplus and welfare even within costly separation through the firm's pricing response. Because the signaling game admits multiple equilibria, these conclusions are stated under an explicit selection rule, which we discuss in Section 3.5.

These findings contribute to the literature in three respects. First, we introduce hackers as an additional audience in a quality-signaling problem. Unlike conventional false-claim settings, in which a favorable claim primarily affects consumer demand, cybersecurity claims also alter the value of the firm as an attack target through the customer base they create, and in an extension they also alter hackers' beliefs about the firm's vulnerability. Second, we identify a signaling role for price that is specific to this environment. Price communicates cybersecurity quality not because it is directly tied to security investment or production cost, but because the resulting market scale creates different expected attack costs for firms with different levels of vulnerability. Relative to the closest false-claim models, the intermediate optimal penalty arises here without consumer verification or investment in deceptiveness (Wu and Geylani, 2020), and the welfare maximizing penalty

can exceed the limiting penalty that yields the consumer surplus supremum. Third, we connect false-claim regulation with the broader cybersecurity-governance environment and demonstrate that, holding the firm's actual security fixed, these policy instruments should not be evaluated independently. For managers, the results imply that cybersecurity communication, pricing, customer growth, and breach-risk management should be considered jointly; for policymakers, that claim verification and penalties for misrepresentation should be coordinated with measures that affect hackers' costs and ability to discover vulnerabilities, with the caveat that our analysis takes security levels, the enforcement technology, and the costs of governance measures as given.

The rest of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 introduces the model and the solution concept. Section 4 displays the reduced form of the game and characterizes the benchmarks and the equilibrium outcomes under asymmetric information. Section 5 examines how the false-claim penalty and the hacker payoff-to-cost ratio affect profits, consumer surplus, and social welfare. Section 6 presents robustness checks and extensions, and Section 7 concludes; proofs are in Online Supplement B.

2. Literature Review

Our study is related to three streams of research: (*i*) consumer privacy, (*ii*) false claims about product quality and (*iii*) cybersecurity.

2.1. Consumer Privacy

The first stream studies consumer privacy concerns arising from firms' use of personal data. Prior work examines consumer profiling and personalized pricing (Montes et al., 2019; Li and Li, 2023; Li et al., 2024; Chen and Zou, 2026; Chen and Zhang, 2026), personalized recommendations (Zhou and Zou, 2023), GDPR and privacy regulation (Choe et al., 2025; Xu et al., 2025), privacy management in service systems (Hu et al., 2022), and privacy-preserving personalized revenue management (Lei et al., 2024). These studies show that consumer data can improve product matching, recommendations, pricing, and service quality, but can also enable surplus extraction and create privacy costs for consumers. Moreover, related operations research literature examines how data-enabled platforms strategically design the transparency of information provided to consumers. Xu et al. (2024) study a service platform that has superior information about customers' hidden preferences and uses personalized recommendations to influence service choices. Although greater transparency can improve matching, reduce congestion, and enhance aggregate customer welfare, they identify a value of opaqueness under which the platform optimally withholds recommendations from some consumers. Whereas the platform in Xu et al. (2024) communicates information about consumers' own preferences, here a firm communicates privately known information about its own security, and the signal also reaches attackers.

More closely related to our paper is the literature on consumer data breaches and firms' data protection decisions in the presence of hacker threats. Ke and Sudhir (2023) analyze how GDPR's privacy rights and data security mandates affect firms' data collection and protection decisions when consumers trade off personalization benefits against data breaches. Fainmesser et al. (2023) develop a model in which firms jointly choose data collection and data protection, while strategic hackers endogenously decide whether to attack and obtain stored user data. These studies highlight how firms' data and protection choices shape consumer privacy losses through hacker activity. Our paper differs by examining firms' strategic misrepresentation of data protection. Whereas existing studies focus on firms' actual data protection decisions, we consider a setting in which a firm's true cybersecurity level is privately known and can be strategically overstated. We show that such cyberwashing affects consumers' data breach beliefs, hackers' attack incentives, and consumer privacy losses. We thus add a strategic misrepresentation channel to this literature.

2.2. False Claims About Product Quality

Our paper is also related to the literature in which firms with private information about product quality may make false claims to influence consumers' beliefs. This literature is particularly relevant because it studies how regulatory penalties, consumer beliefs, and firms' incentives to mimic high-quality products jointly determine whether deceptive claims arise in equilibrium. Corts (2013, 2014) study firms' use of false and unsubstantiated advertising to signal product quality under regulatory enforcement of false advertising penalties: Corts (2013) shows that prohibitions on false or unsubstantiated claims can discourage firms from verifying product quality and truthfully disclosing quality information, while Corts (2014) demonstrates that socially optimal penalties for false advertising are finite rather than arbitrarily high. Piccolo et al. (2015, 2018) study false advertising in a competitive market where low-quality firms can mimic high-quality firms through false advertising, even when consumers are rational, and show that deceptive advertising can soften perceived quality differentiation and intensify price competition, so that consumers may benefit from deception through lower prices. Rhodes and Wilson (2018) show that a low-quality firm makes deceptive claims with a positive probability, rather than always mimicking or separating from the high-quality firm. Wu and Geylani (2020) study false advertising with skeptical consumers who can verify firms' claims at a cost, and show that a higher penalty for false claims can reduce consumer surplus and social welfare by discouraging consumer verification and inducing firms to invest in more sophisticated deception. Guan et al. (2025) extend the false advertising literature to a product safety setting and show that, when firms misrepresent product quality, stricter liability-based penalties can reduce consumer surplus and product safety.

Several studies also examine how nonprice information interacts with price decisions and how far nonverifiable communication can be trusted. For digital products, Xin (2020) shows that customers'

uncertainty about software value before adoption shapes a vendor's optimal licensing and pricing decisions. Zhu and Song (2024) study a platform's disclosure design when advertisers differ in whether they follow disclosed bid recommendations, and find that moderate disclosure can improve platform revenue and market efficiency; Alibeiki and Gümüş (2020) study supply competition in which a buyer privately observes suppliers' quality scores and suppliers compete on quality-adjusted prices; Avinadav et al. (2025) show that a platform privately informed about demand cannot sustain fully revealing cheap talk with competing firms, although strategic ambiguity can be partially informative; and Guan et al. (2026) study a firm's choice of whether and when to obtain third-party quality verification in the presence of biased consumer reviews. In a field experiment, Xu et al. (2026) find that disclosing a voice chatbot's identity reduces drivers' response probability, while human-like speech features can improve conversation length and order acceptance intention even when its identity is disclosed. Lai et al. (2026) show that publicly observed financing decisions signal market prospects to both lenders and competitors, creating opposing incentives to secure favorable financing terms and limit competitive pressure. In our setting, the privately informed party is the seller, the credibility of its communication is shaped by regulatory penalties and price signaling, and the same claim–price pair affects hackers through demand in the baseline and also through their security beliefs in the extension.

Relative to these models, what is new is not the cybersecurity label but the connection between two mechanisms that have been studied separately, which we organize along four dimensions. *Information*: as in the false-claim models, quality (here, security) is privately known and can be overstated, whereas firms in De Cornière and Taylor (2026), discussed below, choose security and do not misrepresent it. *Attack technology*: we adopt their demand-dependent targeting, absent from the false-claim models, under which a larger customer base is a more valuable target. *Price signaling*: combining the two gives price a role that neither strand has, because an additional customer generates a smaller expected breach loss for the secure firm, so a discount is at once a credibility device and a source of attack exposure. *Welfare*: as in Wu and Geylani (2020) and Corts (2014), a stronger penalty can reduce welfare, but its cost here is a higher price paid to a secure firm rather than less consumer verification or less informative claims, and the consumer-optimal and welfare-optimal penalties diverge exactly when the signaling discount expands the secure firm's customer base beyond the level that is efficient once attack exposure is counted (Section 5).

2.3. Cybersecurity

The last stream of cybersecurity research studies firms' actual cybersecurity investment, liability, and outsourcing. Gordon and Loeb (2002) provide a foundational model of optimal cybersecurity investment and show that optimal investment need not increase monotonically with vulnerability. Subsequent work examines cybersecurity outsourcing (Lee et al., 2013; Hui et al., 2019),

cloud architecture and cybersecurity risk (August et al., 2014), and cybersecurity investment in e-commerce supply chains (Luo and Choi, 2022). Zhu et al. (2026) study the provision of cybersecurity resources to small and midsize businesses when security outcomes are interdependent. They focus on investment incentives created by shared security dependencies, whereas we study how ex post cybersecurity disclosure interacts with the firm's pricing decision and with consumers' and hackers' responses. Avinadav et al. (2014) study a manufacturer-retailer supply chain in which demand for a legitimate digital product depends on both price and investment in digital-rights protection, characterizing the interdependence between pricing and protection decisions under different channel-leadership structures and attitudes toward demand risk. Whereas their protection decision is an observable investment intended to discourage digital piracy, we examine how cybersecurity claims affect consumers, strategic hackers, and regulatory policy.

A closely related set of studies explicitly models hacker behavior. De Cornière and Taylor (2026) are particularly relevant. They study firms' cybersecurity investment when hackers decide whether to attack, and show that firms with larger consumer bases can become more attractive targets because successful breaches generate greater hacker payoffs and larger consumer harm; this is the attack technology we adopt. Empirical research also shows that hackers respond strategically to changes in the external enforcement environment. Using distributed denial-of-service attack data from 106 countries, Hui et al. (2017) find that enforcement of the Convention on Cybercrime reduces attacks by at least 11.8 percent, although the deterrence effect disappears when enforcing countries limit their participation in international cooperation. They also document network and displacement effects, so attackers adjust both participation and target choices to enforcement conditions. Wang et al. (2024), however, show that enforcement of computer misuse laws can also discourage cybersecurity-relevant contributions to non-targeted discussions on public hack forums because of uncertainty regarding false prosecution. These studies show that cybersecurity governance affects both malicious activity and the broader behavioral environment; we identify an additional channel, through firms' claims, price signaling, and attack exposure.

3. Model Formulation

We develop a game-theoretic model involving three parties: a firm, hackers, and consumers. The firm sells a digital product at a price p. The setting we have in mind is a firm that charges its customers a price and stores economically valuable customer data, so that the customer base is both the firm's revenue source and the hackers' prize; subscription services, consumer software, web hosting, and online retail fit this description, with the price read as a subscription fee or a usage charge, whereas advertising-funded services and enterprise contracts with negotiated prices fit less well. The firm's cybersecurity level, denoted by z, is the probability that its system has no

vulnerabilities exploitable by hackers. As such, $1-z$ is the probability that the system contains an exploitable vulnerability. We assume that z is the firm's private information and takes one of two values in $\mathbf{Z}=\{z_H, z_L\}$, with $z_L < z_H$. Accordingly, we refer to the firm with z_H as the high type and the firm with z_L as the low type. For expositional brevity and analytical tractability, we set $z_L = 0$ and let $z_H = z$, where $0 < z < 1$.[6]

3.1. Cybersecurity Claims

Privately observing its type z_i, the firm discloses a public cybersecurity claim $r \in \{r_H, r_L\}$. We use r_H to denote a high-security claim, under which the firm states that its cybersecurity level is high, and r_L to denote a non-high-security claim, under which the firm does not state that its cybersecurity level is high. In practice, a firm may refrain from making a high-security claim, either by making a more limited disclosure or by remaining silent. In our model, as long as the high type claims a high cybersecurity level and the other type does not make the same claim, the market can infer the firm's security type. Thus, as in the false advertising literature (Rhodes and Wilson, 2018; Wu and Geylani, 2020), it suffices to consider a binary message space $\{r_H, r_L\}$.

3.2. Consumer Utility

There is a unit mass of consumers with valuation v for the product or service, where v is uniformly distributed on $[0,1]$. Consumers hold a prior belief that the firm's cybersecurity level is high with probability $\alpha \in (0,1)$ and low with probability $1-\alpha$. After observing the firm's cybersecurity claim r and price p, consumers update their belief about the firm's cybersecurity type. We use $\beta(z_i \mid r,p)$ to denote the posterior probability that the firm's cybersecurity type is z_i given the firm's cybersecurity claim r and price p. For notational simplicity, let $\beta = \beta(z_H \mid r,p)$. Consequently, the posterior probability that the firm is of the low type is $1-\beta$, and the firm's expected cybersecurity level from the consumers' perspective is given by $\bar{z} = \beta z_H + (1-\beta) z_L$, which equals βz.

If a consumer purchases the firm's product or service, and an exploitable vulnerability is attacked, then the consumer will incur an expected loss l. For example, this loss captures the repercussions consumers may face when their personal data are leaked to malicious parties, including identity theft, financial fraud, privacy invasion, and the time and effort required to mitigate these harms. As specified later, we incorporate the intensity of hacker attacks by assuming a continuum of hackers who decide whether to attack (De Cornière and Taylor, 2026). A greater number of hacker attacks leads to a larger expected loss for consumers. A consumer who purchases from the firm receives expected utility

$$U_c(v \mid r,p) = v - p - \bar{N}_h(r,p)(1-\bar{z})l \tag{1}$$

[6] In Online Supplement A.3, we show that the case $z_L > 0$ can be mapped to the baseline model by replacing z with $\tilde{z} = (z_H - z_L)/(1-z_L)$ and γ with $\tilde{\gamma} = \gamma(1-z_L)$, preserving the underlying equilibrium mechanism. We then conduct numerical analyses to illustrate the qualitative robustness of our key equilibrium and welfare insights.

where $\bar{N}_h(r,p)$ denotes the intensity of hacker attacks, conditional on the firm's system being vulnerable, that consumers anticipate in equilibrium. We require consumers to hold rational expectations; that is, $\bar{N}_h(r,p)$ equals the actual conditional attack intensity in equilibrium, denoted by $N_h(r,p)$. Thus, consumers' expected breach exposure is $(1-\bar{z})N_h(r,p)$. Throughout, we distinguish the *conditional attack intensity* N_h, the mass of hackers who attack when a vulnerability exists, from the *expected exposure* $(1-\bar{z})N_h$ for consumers and $(1-z_i)N_h$ for a type-i firm; a ranking of conditional intensities across firm types is not by itself a ranking of expected harm. Consumers observe only the claim and the price; they do not observe whether a vulnerability exists. We normalize $l=1$ to simplify notation. As shown in Online Supplement A.1, the relevant parameters can be scaled relative to any $l>0$, preserving the reduced-form demand and profit expressions; conclusions involving the admissible upper bound on γ require the domain qualifications stated there. Consumers receive a utility of zero from the outside option. Thus, a consumer patronizes the firm if and only if $U_c \geq 0$. Consequently, aggregate demand is

$$N_c(r,p) = \left[1-p-(1-\bar{z})N_h(r,p)\right]_+, \tag{2}$$

where $[x]_+ = \max\{x,0\}$.

3.3. Hackers

There is a unit mass of hackers who are heterogeneous in their cost of attacking a vulnerable system, denoted by c, where c is uniformly distributed on $[0,\bar{c}]$. This assumption is consistent with prior work (De Cornière and Taylor, 2026; Fainmesser et al., 2023) and reflects that hackers may differ in their technical capabilities, access to attack tools, and the resources required to carry out an attack against a vulnerability. The attack cost c also includes expected enforcement costs, such as the risk of identification, sanctions, or prosecution.

Each hacker decides whether to attack the firm based on the attack cost and the potential gain from an attack. Attacking a firm with more active consumers yields a larger payoff to hackers, because a successful breach can expose a larger volume of consumer data; a similar assumption is used in prior literature (see, e.g., Png and Wang 2009). In particular, De Cornière and Taylor (2026) model consumers as a prize that makes larger firms more attractive targets and study security investment when security is observed, by hackers and by security-savvy consumers, rather than privately known and claimed; we take this demand-dependent attack technology as our starting point and ask how it shapes the credibility of security claims when security is privately known. Denote the hacker's gain per active consumer of the firm by h, and the hackers' rational expectations regarding the number of the firm's active consumers in equilibrium by $\bar{N}_c(r,p)$. Then, conditional

on an exploitable vulnerability being present, if a hacker with cost c chooses to attack the firm, its expected utility is given by

$$U_h(c \mid r, p) = h \cdot \bar{N}_c(r, p) - c. \tag{3}$$

A hacker chooses to attack if and only if $U_h \geq 0$. In equilibrium, we require hackers' expectation $\bar{N}_c(r, p)$ to be consistent with the actual number of active consumers, denoted by $N_c(r, p)$.

We define $\gamma = \frac{h}{\bar{c}}$ as the hacker payoff-to-cost ratio. A larger γ represents either greater value per compromised consumer or a lower scale of attack costs. External cybersecurity governance can affect both components. For example, monitoring, cross-jurisdictional cooperation, and sanctions can raise expected attack costs, whereas rapid breach response, credential invalidation, and measures that limit the monetization of stolen data can reduce h. We therefore interpret a lower γ as a less favorable environment for attackers, although distinct policies operate through different components of the ratio.

Two features of this formulation merit emphasis. First, one hacker's success does not reduce the gain available to another hacker; h is independent of the number of attackers. As discussed by De Cornière and Taylor (2026), this approximation is natural when consumers are infinitesimal and each hacker compromises only a small share of the customer base, so that overlap across attackers is limited. Online Supplement A.2 introduces hacker-side congestion and shows how the baseline results extend. Second, the baseline abstracts from vulnerability-discovery costs: hackers learn, at no cost, whether an exploitable vulnerability exists before they decide whether to incur c, whereas consumers do not observe this information. This timing approximates settings in which discovery is inexpensive relative to exploiting a known weakness, for example when attackers use automated scanning, public vulnerability disclosures, or leaked credentials; Verizon's 2026 Data Breach Investigations Report identifies vulnerability exploitation as the leading breach entry point (Verizon, 2026), and Unit 42 (2026) reports that attackers begin scanning for newly disclosed vulnerabilities within 15 minutes of a Common Vulnerabilities and Exposures (CVE) announcement. Given a vulnerability, a hacker's decision depends on the firm's claim and price only through the customer base $N_c(r, p)$ they generate, so in the baseline the claim affects attacks through demand and not through hackers' beliefs about the firm's type. Section 6.1 studies an alternative timing in which hackers incur effort before learning whether a vulnerability exists, so that the claim also affects attacks through a direct inference channel.

3.4. Firm Profit

The firm simultaneously determines its cybersecurity claim r and price p. Each attack against a vulnerable system imposes a loss F on the firm. In practice, this loss can include operational disruptions, remediation costs, and reputational damage. With the probability of an exploitable

vulnerability given by $1-z_i$, the firm's expected loss from cyberattacks is $N_h(r,p)(1-z_i)F$ where $N_h(r,p)$ is the intensity of hackers who choose to launch attacks. Hence, a firm incurs a higher expected breach loss when it faces greater hacker attack intensity $N_h(r,p)$ or has a lower cybersecurity level z_i. In §6.2, we generalize the firm's expected loss by further allowing it to increase with the firm's active customer base. The firm chooses its price from $[0,1]$, the support of consumer valuations. Throughout the analysis of the baseline model, we assume $\bar{c} > \max\{hF, hz\}$. The condition $\bar{c} > hF$ ensures positive demands for both cybersecurity types at their symmetric-information prices, while $\bar{c} > hz$ guarantees an interior hacker participation cutoff, $N_h = \gamma N_c < 1$, for every price $p \in [0,1]$ and every belief $\beta \in [0,1]$, so that the continuation outcome after any claim–price pair, on or off the equilibrium path, is given by the interior rational-expectations solution derived in Online Supplement B. Note that the condition can also be rewritten as $\gamma = h/\bar{c} < \bar{\gamma}_B = \min\{1/F, 1/z\}$. We refer to $\bar{\gamma}_B$ as the admissible upper bound on the hacker payoff-to-cost ratio; all statements about "high" hacker threats below are relative to this bound.

If the low-security firm issues a false high-security claim, i.e., chooses r_H, it incurs an expected false-claim penalty, denoted by λ. As discussed in the introduction, false cybersecurity claims may be detected and publicized through regulatory investigations, consumer-protection enforcement, independent technical assessments and whistleblower reports, and the exposed firm may face penalties imposed by regulators. More broadly, the penalty may also be interpreted as reputational loss or loss of future revenue (Wu and Geylani, 2020). Because the enforcement cases discussed in this paper primarily involve regulatory penalties, we focus on these for simplicity. The distinction matters only for welfare accounting: a monetary fine and a reputational loss of the same expected size induce the same equilibrium conduct, but the fine is a transfer to the government whereas the reputational loss is a real cost. All equilibrium results below hold under either interpretation; the welfare statements in Section 5 treat λ as a transfer, and we indicate there how they change under the alternative interpretation in Online Supplement A.9.

We note two aspects of the penalty. First, λ is an expected penalty that bundles the probability of detection with the sanction imposed upon detection; it thus reflects both monitoring stringency and punishment severity; because we do not model the enforcement technology, comparisons across levels of λ hold its cost fixed. Second, λ is a constant, independent of the firm's customer base and of realized attacks, which fits enforcement triggered by audits of the claim itself, by independent technical assessments, or by whistleblowers, as in the Zoom and Illumina cases, where the misrepresentation was identified without a breach having occurred. If instead detection were triggered mainly by breaches, the expected penalty would scale with the low type's attack exposure; Remark A2 in Online Supplement A.11 shows that such a penalty reinforces our mechanism and weakly enlarges the separating regions; the constant penalty is thus the conservative case.

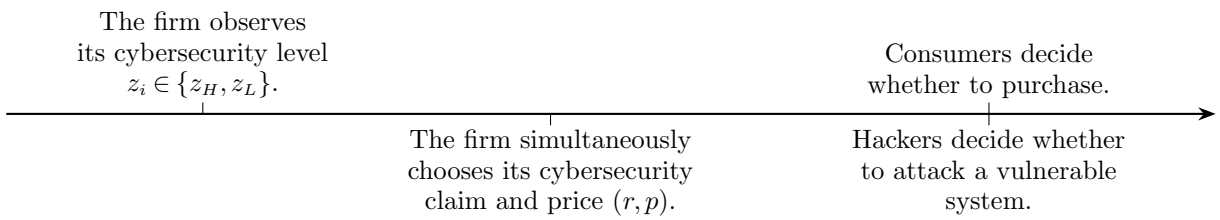


Figure 1 Sequence of Events

Similar to the false advertising literature (Rhodes and Wilson, 2018; Piccolo et al., 2018; Guan et al., 2025), we focus on equilibria in which the high-type firm truthfully makes the high cybersecurity claim r_H. Any equilibrium in which the high-type firm chooses r_L can be ruled out by restricting consumers' posterior beliefs to satisfy $\beta(z_H|r_H, p) \geq \beta(z_H|r_L, p)$ for all p and assuming the firm to make the truthful claim when indifferent between r_L and r_H at the same price. This lets us focus on regulating cyberwashing: preventing the low type from misrepresenting itself as high type.[7]

Thus, the profit functions of the high-type and low-type firms are given by

$$\pi_H(r, p) = p \cdot N_c(r, p) - N_h(r, p) \cdot (1 - z)F, \tag{4}$$

$$\pi_L(r, p) = p \cdot N_c(r, p) - N_h(r, p) \cdot F - \lambda \cdot \mathbf{1}\{r = r_H\}. \tag{5}$$

The timing of the game is shown in Figure 1. First, the firm privately observes its security level $z_i \in \{z_H, z_L\}$. Then, the firm simultaneously issues a cybersecurity claim $r \in \{r_H, r_L\}$ and sets its price p. Subsequently, after observing the firm's claim and price decisions, hackers decide whether to attack if the system is vulnerable, and consumers decide whether to purchase the product. The final-stage consumer and hacker decisions are simultaneous and governed by rational expectations about aggregate demand and attack intensity. Online Supplement C summarizes the notation.

3.5. Solution Concept

In this paper, we adopt a pure-strategy Perfect Bayesian Equilibrium (PBE) as our solution concept. The firm's cybersecurity claim r and price p serve as signals of its cybersecurity level. After observing (r, p), consumers update their beliefs about the firm's cybersecurity type according to Bayes' rule whenever applicable. Given the firm's claim and price, consumers decide whether to purchase the product, while hackers decide whether to launch an attack against an exploitable vulnerability. Anticipating these responses, the firm chooses (r, p) to maximize its profit. A PBE

[7] Because r_L can include silence, there is also a PBE in which both types pool on (r_L, p) under silence: the high type's profit, demand, and exposure are the same as under cyberwashing pooling and the low type saves the penalty. Online Supplement A.12 shows that, in the baseline model, removing the high-type claim restriction preserves the selected prices, exposure, welfare and regime boundaries under the same pure strategy lexicographic ordering, although pooling occurs through silence. For any fixed positive penalty, the high-claim outcome also arises when a sufficiently small positive fraction of consumers requires an affirmative high-security claim, subject to the conditions of Proposition A6.

consists of a claim–price pair for each type, a belief system $\beta(\cdot \mid r, p)$ defined for every claim–price pair, and the rational-expectations continuation outcome at every pair, such that each type's pair maximizes its profit against the belief system and beliefs on the equilibrium path are derived from Bayes' rule. For every candidate equilibrium analyzed below, we state the off-path beliefs that support it and verify that neither type gains from deviating to *any* claim–price pair in $\{r_H, r_L\} \times [0, 1]$, not only to the other type's equilibrium pair.

As in many signaling games, PBE leaves beliefs following off-path claim–price pairs unrestricted and can therefore support multiple outcomes. Let $\mathcal{E}$ denote the set of pure-strategy PBE outcomes admitting a consistent supporting assessment in the maintained class (equilibria in which the high type claims r_H). For the continuous signal space, consistency is understood through strictly positive, type-dependent signal densities whose distributions converge weakly to the proposed pure strategies and whose Bayes posteriors converge pointwise to the stated beliefs. The supporting assessments constructed below satisfy this requirement, as shown at the start of Online Supplement B.1; their limiting choices and continuation rules are sequentially rational by the PBE conditions above. Let $\mathcal{E}_H = \arg\max_{e \in \mathcal{E}} \pi_H(e)$. Following the Lexicographically Maximum Sequential Equilibrium (LMSE) ordering of Mailath et al. (1993), an LMSE outcome maximizes $\pi_L(e)$ over $e \in \mathcal{E}_H$. Thus, the high type's payoff is ranked first and the low type's payoff breaks ties. The uniqueness statements below refer to uniqueness of the LMSE outcome within this class of pure-strategy equilibria.

Two remarks on the scope of the selection rule are in order. First, the restriction to pure strategies excludes semi-separating equilibria in which the low type randomizes between truthful disclosure and mimicry, which are economically relevant in related advertising models (Rhodes and Wilson, 2018). Proposition A2(iii) in Online Supplement A.6 shows that such equilibria give the high type a strictly lower profit than the pure-strategy outcome selected by the LMSE, so enlarging the class in this way does not change the selected outcome. Second, the pooling outcomes selected below are supported by the pessimistic belief that an unexpected claim–price pair comes from the low type. Like the false-advertising models of Wu and Geylani (2020) and Piccolo et al. (2018), we regard such skepticism as a plausible description of consumers who cannot tell a secure firm's price cut from a vulnerable firm's attempt to attract demand. Two bodies of evidence support it: meta-analyses find a robust positive effect of price on perceived quality, so that consumers read a lower price as a sign of lower quality (Rao and Monroe, 1989; Völckner and Hofmann, 2007), and in the market for smartphone applications cheaper apps request more privacy-sensitive permissions, although permissions concern access to personal data rather than cybersecurity strength and do not identify off-path beliefs (Kummer and Schulte, 2019). We do not formalize this interpretation, and we are explicit about its cost: Proposition A2(i) shows that the selected pooling outcomes fail the Intuitive Criterion of Cho and Kreps (1987) and the D1 criterion of Banks and Sobel

(1987). The foundation we offer instead is the undefeated-equilibrium criterion of Mailath et al. (1993), which compares complete equilibria rather than isolated deviations; Proposition A2(ii) shows that the LMSE outcome is undefeated. We therefore present the LMSE as an explicit selection assumption and flag the conclusions that depend on it; Table A2 in Online Supplement A.6 lists them (cyberwashing at low penalties and the governance-induced switch into it) and those that do not (the price-signaling results within the separating regime, where the three criteria select the same outcome).

4. Equilibrium Results and Analysis

In this section, we first display the reduced form of the game, which summarizes how consumers' beliefs and the firm's price determine demand, attack intensity, and profits. We then analyze a symmetric-information benchmark, in which the firm's cybersecurity level is public information, and summarize what a benchmark with exogenous attack intensity, analyzed in Online Supplement A.5, reveals about the role of the endogenous attack channel. Finally, we examine the asymmetric-information setting, in which the firm is privately informed about its cybersecurity level and signals its type through its cybersecurity claim and price.

For a belief β that the firm is the high type, the rational-expectations continuation after any claim–price pair yields demand and conditional attack intensity:

$$N_c(p,\beta) = \frac{1-p}{1+\gamma(1-\beta z)} \quad \text{and} \quad N_h = \gamma\, N_c(p,\beta). \tag{6}$$

The profit of a type-i firm is

$$\pi_i(r,p\,|\,\beta) = \big[p-\gamma(1-z_i)F\big]\, N_c(p,\beta) - \lambda\cdot\mathbf{1}\{i=L,\ r=r_H\}. \tag{7}$$

Security therefore affects the firm's problem in two places. Through β, it raises demand at any price. Through $\gamma(1-z_i)F$, it lowers the effective marginal cost of serving a customer: each additional customer attracts γ additional attackers, and an attack costs the firm F only if its system is vulnerable, which happens with probability $1-z_i$. The second effect is a single-crossing property: for any belief and claim,

$$\pi_H(r,p\,|\,\beta) - \pi_L(r,p\,|\,\beta) = z\gamma F\, N_c(p,\beta) + \lambda\cdot\mathbf{1}\{r=r_H\}, \tag{8}$$

which is strictly decreasing in p because $N_c(p,\beta)$ is strictly decreasing in p for a given belief β. The gain from a price cut is therefore larger for the secure type. Price can serve as a signal because of this property, and much of the analysis in Section 4.2 follows from it by standard incentive-compatibility arguments. The end of Section 4.1 asks what the endogenous attack channel adds to those arguments.

4.1. Symmetric Information

Under symmetric information, the firm's cybersecurity level is publicly observable to both consumers and hackers, so that $\beta = 1$ for the high type and $\beta = 0$ for the low type in (6)–(7). The low-type firm has no incentive to make a false claim, because doing so incurs a penalty without affecting consumers' beliefs. Hence, the firm's cybersecurity claim decision is trivial: the high-type firm chooses r_H, whereas the low-type firm chooses r_L. We use the superscript m for the symmetric-information benchmark; Lemma 1 gives the optimal price and profit.

LEMMA 1. *Under symmetric information, the firm's optimal price and profit are given by* $p_i^m = \frac{1+\gamma F(1-z_i)}{2}$, *and* $\pi_i^m = \frac{[1-\gamma F(1-z_i)]^2}{4[1+\gamma(1-z_i)]}$ *for* $i \in \{H, L\}$. *As a result, the high-type firm charges a lower price and earns a higher expected profit than the low type, i.e.,* $p_H^m < p_L^m$ *and* $\pi_H^m > \pi_L^m$.

As shown in Lemma 1, the optimal price decreases with the firm's cybersecurity level z_i, and the high-type firm offers a lower price than the low type. The intuition follows from the effective marginal cost in (7): attracting more consumers increases the number of hackers willing to attack, leading to greater expected losses from cyberattacks. A higher cybersecurity level z_i, or equivalently lower vulnerability, allows the firm to accommodate greater demand expansion with a smaller increase in expected cyberattack losses. Hence, the high-type firm optimally charges a lower price than the low type in order to serve greater consumer demand, and because attacks scale with demand, it also faces more attacks (Online Supplement A.4 records the customer bases and attack intensities under symmetric information and in each equilibrium below). A similar relationship between a firm's cybersecurity level and its optimal price is documented in De Cornière and Taylor (2026). In our asymmetric-information setting, however, this result has an additional implication: because a high-security perception allows the firm to attract greater demand, the low-type firm has an incentive to mimic the high-type firm by overstating its cybersecurity level.

To see how the endogenous response of hacker attacks plays a role, we examine a benchmark in Online Supplement A.5 in which conditional attack intensity is an exogenous constant, independent of the firm's claim, price, and customer base (Proposition A1). With exogenous attacks, the profit difference between the two types does not depend on price, so price signaling is money burning: to separate from the low type, the high-type firm is indifferent between an upward and a downward price distortion. By contrast, as we show below, with endogenous attacks, the single-crossing property (8) implies a downward price distortion under costly separation, and the high-type firm receives an information rent that grows with its customer base (Lemma 2).

4.2. Asymmetric Information

We now analyze the equilibrium outcomes under asymmetric information. Two types of equilibria may arise. In a separating equilibrium, the two cybersecurity types choose different security

claims and/or prices, allowing consumers to infer the firm's true cybersecurity level. In a pooling equilibrium, the low-type firm engages in cyberwashing by making the high-security claim and charging the same price as the high type. Consequently, consumers cannot distinguish between firm types and make purchase decisions based on their prior belief. We first establish conditions under which separating and pooling equilibria exist, and then select the equilibrium outcome according to LMSE in case of coexisting equilibria.

4.2.1. The Separating Equilibrium Let p_H^{sep} and p_L^{sep} denote the prices charged by the high-type and low-type firms, respectively, in a separating equilibrium. Recall that the high type always claims r_H. We first note that the low type must claim r_L in any separating equilibrium. Suppose instead that the low type also claims r_H and separates from the high type only through its price. Then, the low type could switch its claim from r_H to r_L while charging the same price p_L^{sep}. This deviation would not change consumers' beliefs or demand, but would allow the low type to avoid the false-claim penalty λ. Hence, such a deviation is strictly profitable, implying that the two types will not separate solely through price while making the same claim. Accordingly, the separating equilibrium must satisfy the following incentive compatibility constraints:

$$\pi_H(r_H, p_H^{sep} \mid \beta = 1) \geq \pi_H(r_L, p_L^{sep} \mid \beta = 0), \tag{9}$$

$$\pi_L(r_L, p_L^{sep} \mid \beta = 0) \geq \pi_L(r_H, p_H^{sep} \mid \beta = 1). \tag{10}$$

Constraint (9) is implied by a stronger requirement that we verify in the proof: with the off-path belief $\beta(r,p) = 0$ for every pair other than (r_H, p_H^{sep}), the high type must prefer its equilibrium pair to its best deviation under that belief, $\max_p \pi_H(r,p \mid \beta = 0) = [1-\gamma F(1-z)]^2/[4(1+\gamma)]$. The proof shows that this constraint is slack in every separating equilibrium characterized below, so the low-type firm's constraint (10) is the only incentive constraint that can bind; it binds under costly separation. In what follows, we focus on the separating equilibrium that maximizes the high-type firm's profit among all feasible separating equilibria, which is consistent with the LMSE criterion. For ease of exposition, we define the following threshold values $\gamma_s = \frac{2+zF-\sqrt{z\left(zF^2+4(1+F)\right)}}{2F(1-z)}$, $\lambda_s(\gamma) = \frac{z\gamma\left[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2\right]}{4(1+\gamma)(1+\gamma(1-z))}$, and $S(\lambda) = \sqrt{\frac{z\gamma(1-\gamma F)^2 - 4(1+\gamma)(1+\gamma(1-z))\lambda}{1+\gamma}}$. Table A6 in Online Supplement C.2 collects these and the other thresholds used in the paper.

PROPOSITION 1. *A separating equilibrium always exists in which the high-type firm chooses r_H and the low-type firm chooses r_L. The equilibrium prices take one of the following two forms:*
(i) If $0 < \gamma < \gamma_s$ and $0 < \lambda < \lambda_s(\gamma)$, a costly separating equilibrium exists, in which $p_H^{sep} = \frac{1+\gamma F - S(\lambda)}{2}$ and $p_L^{sep} = \frac{1+\gamma F}{2}$.
(ii) Otherwise, if $\gamma \geq \gamma_s$ or if $\lambda \geq \lambda_s(\gamma)$, a costless separating equilibrium exists, in which $p_H^{sep} = \frac{1+\gamma F(1-z)}{2}$ and $p_L^{sep} = \frac{1+\gamma F}{2}$.

The high-type firm can use two instruments, its cybersecurity claim and price, to distinguish itself from the low type. Proposition 1 establishes that a separating equilibrium always exists, but whether the claim alone is an effective signaling instrument depends on both the hacker payoff-to-cost ratio γ and the false-claim penalty λ. Existence of a separating PBE for every $\lambda > 0$ is distinct from its selection: pooling equilibria also exist for low penalties, and which outcome the LMSE selects is the subject of Section 4.2.3.

In the costly separating equilibrium, firms make truthful claims, and the high-type firm signals its type by setting a strictly lower price than under symmetric information, $p_H^{sep} < p_H^m$, whereas the low-type firm charges its symmetric-information price, $p_L^{sep} = p_L^m$. We call this equilibrium a *costly* separating equilibrium because, in this case, the high-type firm incurs a signaling cost by deviating from its symmetric-information optimal price. In contrast, when either γ or λ is high, each firm type sets the same equilibrium price as in the symmetric-information case, that is, $p_i^{sep} = p_i^m$ for both $i \in \{H, L\}$. Thus, the high-type firm can use the cybersecurity claim alone, without any price distortion, to separate itself from the low type. We refer to this equilibrium as a *costless* separating equilibrium. Under costly separation, the high type's lower price gives it a larger customer base, and hence more attacks, than under symmetric information; the low type's outcomes are unchanged (Corollary A1, Online Supplement A.4).

The intuition is as follows. When the false-claim penalty λ is severe, or when the hacker payoff-to-cost ratio γ is high ($\gamma \geq \gamma_s$, which can occur within the admissible range only if $\gamma_s < \bar{\gamma}_B$), the cybersecurity claim alone can serve as an effective signaling instrument. The role of λ is direct: a high penalty makes misrepresenting its cybersecurity level prohibitively costly for the low-type firm. More intriguingly, a high γ can also enable separation based solely on the claim. By misrepresenting itself as the high type, the low-type firm may attract more consumers, but the larger consumer base also exposes it to a greater expected loss in the event of a breach, and a higher γ amplifies this loss, offsetting the demand-expanding benefit of mimicking. Therefore, when γ is high, the low-type firm prefers not to overstate its cybersecurity level, even if the penalty is not substantial.

On the other hand, when both γ and λ are low, the cybersecurity claim alone is insufficient, and the high-type firm must distort its price *downward* to signal its type. The distortion is downward because, for both types, the expected breach loss increases with the consumer base, and the more vulnerable low-type firm bears the greater increase as the consumer base expands. By lowering its price, the high-type firm can expand its consumer base to a point at which the low-type firm is unwilling to mimic, due to the increased cybersecurity breach risk. The single-crossing identity (8) gives a sharp expression for what the secure firm earns from this strategy.

LEMMA 2 **(Informational rent)**. *In any outcome in which the low type is indifferent between mimicking the high type's claim–price pair and truthful disclosure, $\pi_H = \pi_L^m + \lambda + z\gamma F N_{c,H}$, where $N_{c,H}$ is the high type's customer base in that outcome.*

The component of the high type's payoff above $\pi_L^m + \lambda$ equals $z\gamma F N_{c,H}$ and increases with its customer base. This is why the price distortion is downward and why it is not money burning: the discount that deters the vulnerable firm simultaneously raises the secure firm's rent. As discussed earlier, this result arises from the endogenous response of attacks in our signaling model. Online Supplement A.5 and Table A1 present a detailed comparison between our model and an exogenous-attack benchmark, clarifying which results are attributable to this feature.[8]

4.2.2. The Pooling Equilibrium A pooling equilibrium may arise when the low-type firm makes a high-security claim r_H and also mimics the high type by charging the same price, denoted by p^{pool}. Because the observed strategy (r_H, p^{pool}) does not reveal the firm's type, consumers' posterior belief coincides with the prior belief, i.e., $\beta(r_H, p^{pool}) = \alpha$. As is common in the literature (Wu et al., 2020; Wu and Geylani, 2020), we specify the off-equilibrium belief as $\beta(r, p) = 0$ when consumers observe any off-path strategy $(r, p) \neq (r_H, p^{pool})$. Formally, the pooling equilibrium must satisfy

$$\pi_H(r_H, p^{pool} \mid \beta = \alpha) \geq \max_{r,p} \pi_H(r, p \mid \beta = 0), \tag{11}$$

$$\pi_L(r_H, p^{pool} \mid \beta = \alpha) \geq \max_{r,p} \pi_L(r, p \mid \beta = 0), \tag{12}$$

where $(r, p) \neq (r_H, p^{pool})$ denotes any off-equilibrium claim-price pair.

For analytical tractability, we specialize to the prior belief of $\alpha = 1/2$.[9] In this case, the two incentive compatibility constraints jointly determine the feasible pooling price interval $[\underline{p}^{pool}, \bar{p}^{pool}]$; the closed-form bounds and the associated threshold functions γ_1, γ_p, and $\lambda_1(\gamma)$ are given in the proof in Online Supplement B. The following proposition gives the necessary and sufficient conditions for the existence of pooling equilibria.

Proposition 2. *A pooling equilibrium exists if and only if* $0 < \gamma < \gamma_p$ *and* $0 < \lambda < \lambda_1(\gamma)$*, or* $\gamma_1 < \gamma < \gamma_p$ *and* $\lambda = \lambda_1(\gamma)$*. In any such equilibrium, both firm types claim* r_H *and charge the same price* $p^{pool} \in [\underline{p}^{pool}, \bar{p}^{pool}]$*. The set of pooling prices is a nondegenerate interval if* $0 < \lambda < \lambda_1(\gamma)$ *and a single price if* $\lambda = \lambda_1(\gamma)$*.*

Proposition 2 shows that a pooling equilibrium exists only when both the hacker payoff-to-cost ratio γ and the false-claim penalty λ are low. Intuitively, the low-type firm has little incentive to cyberwash when false claims are heavily penalized or when the hacker payoff-to-cost ratio is high

[8] For example, with exogenous attacks, the breach loss F does not affect the equilibrium thresholds. Within the costly separating regime, consumer surplus and social welfare rise or fall together as the penalty changes. In contrast, with endogenous attacks, F affects the regime boundaries (Propositions 1–4), and the consumer-surplus-maximizing and welfare-maximizing penalties can differ (Proposition 8), because demand expansion now raises attack exposure.

[9] Section 6.3 relaxes this assumption: alternative prior beliefs shift the pooling-separating boundary but do not alter the qualitative equilibrium structure or the main mechanism.

enough that the additional demand generated by cyberwashing exposes it to large expected breach losses. Pooling can therefore be sustained only when the penalty is limited and the hacker threat is not severe enough to outweigh the demand benefit from mimicry.

4.2.3. Equilibrium Outcome Our results so far indicate that separating and pooling equilibria can coexist. To refine the set of equilibria, we adopt the LMSE of Section 3.5, which reflects the trade-off between revealing favorable private information and bearing the cost of doing so: a high-security firm may distinguish itself through a price concession that reduces its margin and increases attack exposure, and it need not prefer separation when these costs outweigh the benefit of being recognized as secure. An outcome-based selection allows pooling to prevail when costly differentiation is unattractive, while prioritizing the high-security type reflects the direction of imitation. Related reasoning is adopted in pricing, safety, and digital-governance models (Jiang et al., 2014; Doshi and Schmidt, 2024; Guan et al., 2025), experimental evidence documents a preference for pooling even where commonly used refinements predict costly separation (Schmidt and Buell, 2017), and Tian and Jiang (2017) demonstrate that LMSE can support economically meaningful pooling outcomes eliminated by the Intuitive Criterion.

Pooling involving cyberwashing compresses the type-specific outcomes into a common intermediate allocation: relative to symmetric information, it expands the low type's customer base and attack intensity, whereas the high type's security advantage is not recognized, reducing both (Corollary A2 in Online Supplement A.4). Propositions 3 and 4 characterize how the false-claim penalty and the hacker payoff-to-cost ratio determine which equilibrium the LMSE selects.

PROPOSITION 3. *For every admissible $\gamma \in (0, \bar{\gamma}_B)$ there is a unique LMSE outcome, characterized with respect to the expected penalty for a false claim as follows: the LMSE is costless separating if $\lambda \geq \lambda_s(\gamma)$, costly separating if $\lambda_p(\gamma) < \lambda < \lambda_s(\gamma)$, and pooling if $0 < \lambda \leq \lambda_p(\gamma)$, where $\lambda_p(\gamma) < \lambda_s(\gamma)$ whenever $\lambda_p(\gamma) > 0$. In the pooling outcome, both types charge the high type's symmetric-information price, $p^{pool} = p_H^m$.*

Proposition 3 identifies up to three enforcement regimes for any given hacker payoff-to-cost ratio (γ). When the expected false-claim penalty is low, $\lambda \leq \lambda_p(\gamma)$, deterring false claims would require the high-type firm to make an excessively costly price concession, so the high type prefers the pooling outcome, in which the low type engages in cyberwashing. Once the penalty exceeds $\lambda_p(\gamma)$, regulation and attack exposure jointly deter cyberwashing: the high type separates by charging a demand-expanding price below its full-information optimum, attracting more consumers and greater attack exposure, an outcome that is more costly for the more vulnerable low type to imitate. As the penalty increases, it bears a larger share of the deterrence burden, reducing the price

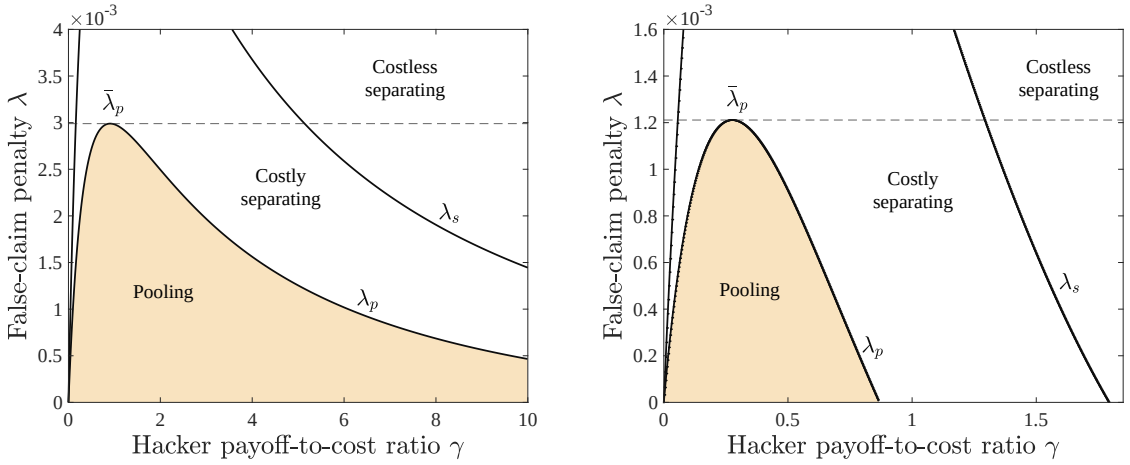
distortion required for separation. Finally, when $\lambda \geq \lambda_s(\gamma)$, the penalty alone makes a false high-security claim unprofitable at the high type's full-information price, so separation becomes costless. At $\lambda = \lambda_p(\gamma)$, the high type is indifferent between the best separating and pooling outcomes, and the LMSE tie-breaking rule selects pooling because the low type prefers it; this convention matters in Section 5, where the consumer-surplus supremum is approached as $\lambda \downarrow \lambda_p$ but not attained.

Proposition A2 in Online Supplement A.6 records three properties of the selected outcome discussed in Section 3.5. First, for every penalty at which the LMSE selects pooling, the pooling outcome fails the Intuitive Criterion and D1, under which the outcome is separating for every $\lambda > 0$; this is the price of the pessimistic off-path beliefs that sustain pooling, because a price cut that is attractive only to the secure firm is exactly the deviation that forward-induction refinements use to break pooling. Second, the LMSE outcome of Proposition 3 is undefeated in the sense of Mailath et al. (1993), the foundation we rely on instead: in the pooling region the high type weakly prefers pooling to its best separating outcome and strictly prefers the selected pooling price to any other, so no equilibrium defeats it. Third, no semi-separating equilibrium, in which the low type randomizes between truthful disclosure and the high type's claim–price pair, gives the high type a higher profit than the LMSE outcome, because the posterior attached to the high type's pair is then below one, so its customer base, and hence its rent in Lemma 2, is smaller than under pure separation; enlarging the equilibrium class in this way does not change the selected outcome.

The hacker environment affects both thresholds, $\lambda_p(\gamma)$ and $\lambda_s(\gamma)$, because γ determines how strongly an expanded customer base raises the expected cost of attack. Next, we present the equilibrium outcomes in terms of γ. Let $\bar{\gamma}$ denote the unique point at which $\lambda_p(\gamma)$ returns to zero, so that $\lambda_p(\gamma) > 0$ on $(0, \bar{\gamma})$. Online Supplement B shows that $\lambda_p(\gamma)$ is unimodal on $(0, \bar{\gamma})$, with a maximum $\bar{\lambda}_p = \max_\gamma \lambda_p(\gamma)$ attained at γ_0; for $0 < \lambda < \bar{\lambda}_p$ we denote by $\bar{\gamma}_1(\lambda) < \gamma_0 < \bar{\gamma}_2(\lambda)$ the two solutions of $\lambda_p(\gamma) = \lambda$ on $(0, \bar{\gamma})$. Whether the second crossing $\bar{\gamma}_2(\lambda)$ lies inside the admissible range $(0, \bar{\gamma}_B)$ depends on the firm's breach loss: Online Supplement B defines a threshold $F^o(z) \in (0, z)$ such that $\bar{\gamma} \leq \bar{\gamma}_B$ if and only if $F \geq F^o$.

PROPOSITION 4. *Fix $0 < \lambda < \bar{\lambda}_p$ and restrict attention to the admissible range $\gamma \in (0, \bar{\gamma}_B)$.*
(i) If $F \geq F^o$, then $\bar{\gamma}_1(\lambda) < \bar{\gamma}_2(\lambda) < \bar{\gamma}_B$, and the LMSE is separating for $\gamma < \bar{\gamma}_1(\lambda)$, pooling for $\bar{\gamma}_1(\lambda) \leq \gamma \leq \bar{\gamma}_2(\lambda)$, and separating again for $\bar{\gamma}_2(\lambda) < \gamma < \bar{\gamma}_B$.
(ii) If $F < F^o$, the pooling set is $\{\gamma \in (0, \bar{\gamma}_B) : \lambda_p(\gamma) \geq \lambda\}$, which, when nonempty, is an interval whose right endpoint is $\min\{\bar{\gamma}_2(\lambda), \bar{\gamma}_B\}$. In particular, if $\lambda < \lambda_p(\bar{\gamma}_B)$ or if $\lambda = \lambda_p(\bar{\gamma}_B)$ and $\bar{\gamma}_B > \gamma_0$, the LMSE is separating for $\gamma < \bar{\gamma}_1(\lambda)$ and pooling for all $\bar{\gamma}_1(\lambda) \leq \gamma < \bar{\gamma}_B$, so that separation does not reemerge within the admissible range; if $\lambda \geq \lambda_p(\bar{\gamma}_B)$ and $\bar{\gamma}_B \leq \gamma_0$, the LMSE is separating on the whole admissible range; and if $\lambda > \lambda_p(\bar{\gamma}_B)$ and $\bar{\gamma}_B > \gamma_0$, the pattern of part (i) obtains.

Propositions 3 and 4 are illustrated in Figure 2. When the false-claim penalty λ exceeds $\lambda_s(\gamma)$, the high-type firm separates costlessly, without any price distortion relative to the symmetric-information benchmark. When λ falls below $\lambda_s(\gamma)$, the claim alone no longer deters mimicry, and the high type must distort its price downward to sustain separation; as λ decreases further, the required distortion becomes more severe, raising the signaling cost and reducing the high type's separating profit. Once λ falls below $\lambda_p(\gamma)$, this profit is lower than the high type's pooling profit, so the unique LMSE outcome is pooling.



(a) $0 < F < F^o$ (b) $F \geq F^o$

Figure 2 Effect of hacker payoff-to-cost ratio γ and false-claim penalty λ on LMSE equilibrium outcomes within the admissible range $(0, \bar{\gamma}_B)$ (Parameters: $z = 0.1$; $F = 0.02$ in panel (a) and $F = 0.4$ in panel (b)).

As shown in Proposition 4, the pooling threshold $\lambda_p(\gamma)$ is strictly unimodal on its positive-value interval $(0, \bar{\gamma})$. This implies that the hacker payoff-to-cost ratio has a non-monotonic effect on cyberwashing. Given a fixed false-claim penalty that is not prohibitively high ($\lambda < \bar{\lambda}_p$), the pooling equilibrium, or cyberwashing, tends to emerge in equilibrium when the hacker payoff-to-cost ratio γ is moderate relative to the admissible range. The intuition is as follows. When γ is very low, either because hackers face very high attack costs or because each breach yields little gain, cybersecurity risk has little effect on consumer purchases or firm profits, so the economic difference between the two types is limited; the low type then gains little from cyberwashing, and the high type can separate at a low signaling cost and prefers separation to pooling. When γ is high, mimicking the high type substantially increases the low type's exposure to breach losses, which weakens its incentive to mimic; whether this restores separation before the admissible bound is reached depends on how costly that exposure is for the firm. Panel (b) of Figure 2 shows the case $F \geq F^o$, in which pooling is confined to an interior interval of γ for every $\lambda < \bar{\lambda}_p$. Panel (a) shows the case $F < F^o$: for penalties below $\lambda_p(\bar{\gamma}_B)$ the pooling region extends to the admissible bound, because the additional attacks induced by mimicry impose only small losses on the low type, so that no admissible threat

level is strong enough to discipline it. What is specific to the endogenous attack channel is that this second transition is governed by the firm's exposure cost F, which is why it can fall outside the admissible range.

Our equilibrium results have the following implications. A sufficiently high false-claim penalty λ can preclude cyberwashing. However, λ is an *expected* penalty that incorporates both the probability of detecting a false claim and the sanction imposed upon detection, and in practice it may be limited because verifying an individual firm's cybersecurity statements is often difficult, costly, and technically complex. When false-claim enforcement is limited, regulators may turn to broader cybersecurity governance measures that reduce the hacker payoff-to-cost ratio γ, such as cybercrime monitoring and enforcement that raise hackers' expected costs, or breach-notification and consumer-protection requirements that reduce the exploitable gains from successful breaches. Our results show, however, that such measures may have unintended consequences if they are not sufficiently strong: by reducing γ to an intermediate level, they may inadvertently induce cyberwashing rather than prevent it. This statement, like all statements about governance in this paper, holds the false-claim penalty λ, the firms' actual security levels, and the costs of the governance measures fixed; it describes the market response to a change in γ, not a regulator's joint optimization.

5. Implications for Firm Profits, Consumer Surplus, and Social Welfare

The preceding analysis identifies the conditions under which cyberwashing is either sustained or deterred, depending on the false-claim penalty λ and the hacker payoff-to-cost ratio γ. In this section, we conduct comparative statics to examine how these two parameters affect equilibrium firm profits, consumer surplus, and social welfare.

5.1. Effect of the False-Claim Penalty

To examine the effect of the false-claim penalty, we focus on the region where $\lambda_p(\gamma) > 0$ and $\lambda > 0$. In this region, as λ increases, the equilibrium shifts from pooling to costly separating and eventually to costless separating. This allows us to focus on the parameter region in which cyberwashing can potentially arise. For brevity, we hereafter write $\lambda_p(\gamma)$ and $\lambda_s(\gamma)$ simply as λ_p and λ_s.

Proposition 5 **(Effect on Equilibrium Prices)**. *As the false-claim penalty λ increases, the equilibrium price changes as follows.*

(i) The low-type firm's equilibrium price remains constant for $\lambda \le \lambda_p$, increases discontinuously at $\lambda = \lambda_p^+$ and remains constant for $\lambda > \lambda_p$.[10]

(ii) The high-type firm's equilibrium price remains constant for $\lambda \le \lambda_p$, decreases discontinuously at $\lambda = \lambda_p^+$, increases for $\lambda_p < \lambda < \lambda_s$, and remains constant for $\lambda \ge \lambda_s$.

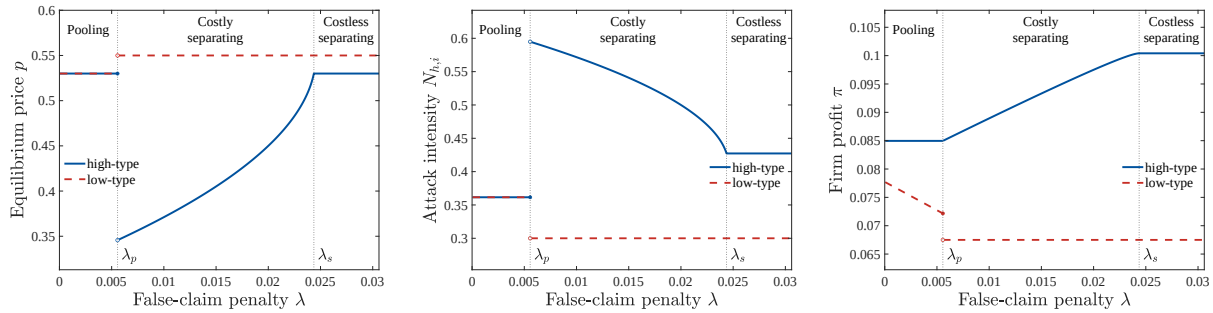


Figure 3 Effect of false-claim penalty λ on (a) equilibrium prices, (b) hacker attack intensity, and (c) firm profits, from left to right (Parameters: $z = 0.4$, $F = 0.05$, $\gamma = 2$).

Figure 3(a) depicts the comparative static result in Proposition 5. When $\lambda \le \lambda_p$, the equilibrium is pooling: both firm types make the high-security claim and charge the same price, which the penalty does not affect. Above λ_p, the equilibrium shifts to costly separation. The low-type firm truthfully reveals its cybersecurity level and charges its symmetric-information price, which is higher than the pooling price because, once its low cybersecurity level is revealed, the low type optimally serves a smaller market at a higher price. By contrast, at the tipping point $\lambda = \lambda_p^+$, the high-type firm begins to prefer separation to pooling, and its equilibrium price drops discontinuously: it distorts its price downward to attract a larger customer base, which increases attack exposure and deters the low type from mimicking. As λ continues to increase, the penalty provides greater deterrence against mimicry, a smaller price distortion is required to sustain separation, and the high type's price increases. Once $\lambda \ge \lambda_s$, the penalty alone deters cyberwashing, so the high type no longer distorts its price and both types charge symmetric-information prices.

Next, we present the resulting hacker attack intensity faced by each firm type, denoted by $N_{h,i}$ for $i \in \{H, L\}$; specifically, how it changes as the false-claim penalty increases. The result helps further explain the equilibrium pricing patterns discussed above.

PROPOSITION 6 **(Effect on Hacker Attack Intensity)**. *As the false-claim penalty λ increases, the attack intensity faced by each firm type changes as follows.*

(i) The attack intensity for the low-type firm remains constant for $\lambda \le \lambda_p$, decreases discontinuously at $\lambda = \lambda_p^+$, and remains constant for $\lambda > \lambda_p$.

(ii) The attack intensity for the high-type firm remains constant for $\lambda \le \lambda_p$, increases discontinuously at $\lambda = \lambda_p^+$, decreases for $\lambda_p < \lambda < \lambda_s$, and remains constant for $\lambda \ge \lambda_s$.

As illustrated in Figure 3(b), when $\lambda \le \lambda_p$, attack intensity remains constant for both types because their equilibrium prices and consumer bases do not change within the pooling regime. When the equilibrium switches from pooling to separation at $\lambda = \lambda_p^+$, the attack intensity faced by

[10] The boundary $\lambda = \lambda_p^+$ means $\lim_{\lambda \downarrow \lambda_p}$.

the high-type firm increases, whereas that faced by the low type decreases. This echoes our earlier discussion: to separate itself, the high-type firm lowers its price and expands its consumer base, which increases hackers' expected payoff and attracts more attacks, whereas the low-type firm, after revealing its type, raises its price and serves a smaller consumer base. As λ increases further, the high type reduces its price distortion and serves fewer consumers, so its attack intensity decreases until costless separation is reached, while the low type's attack intensity remains unchanged. Baksy et al. (2025) report a positive relationship between firm size and model-based cyber-exposure scores. This is consistent with the targeting channel in our model, although it is not a direct estimate of the causal effect of customer-base expansion on realized attacks; our analysis shows how this targeting channel interacts with endogenous pricing. Propositions 5 and 6 also yield the model's observable predictions, as opposed to generic distrust of security claims: a secure firm that cannot rely on verification prices below its full-information price and grows its customer base, this growth is accompanied by more attacks on the secure firm and fewer on the insecure one, and stronger enforcement against false claims raises the secure firm's price.

Two qualifications apply. First, Proposition 6 ranks conditional attack intensities, that is, the mass of hackers who attack a firm whose system is vulnerable, and does not by itself rank expected harm: the high type's expected exposure is $(1-z)N_{h,H}$, whereas the low type's is $N_{h,L}$, and under costly separation either can be larger; the welfare analysis below works with expected exposures. Second, whether the high type faces greater conditional attack intensity under separation than under pooling hinges on the assumption that hackers commit attack effort only after a vulnerability is discovered at no cost, so that the claim affects attacks only through demand; in §6.1, where hackers must exert effort first, type separation may, though not always, reduce the attack intensity faced by both types.

We then explore how pricing and cyberwashing affect the firm's profit: although the high-type firm faces greater attack intensity under separation than under pooling, its expected payoff is continuous at the pooling-to-separation transition and increases thereafter within costly separation, whereas the low-type firm is worse off because its type is revealed.

Proposition 7 **(Effect on Firm Profits)**. *As the false-claim penalty λ increases, the firm's equilibrium profits change as follows.*

(i) The low-type firm's profits decrease for $\lambda \le \lambda_p$, decrease discontinuously at $\lambda = \lambda_p^+$, and remain constant for $\lambda > \lambda_p$.

(ii) The high-type firm's profits remain constant for $\lambda \le \lambda_p$, increase for $\lambda_p < \lambda < \lambda_s$, and remain constant for $\lambda \ge \lambda_s$.

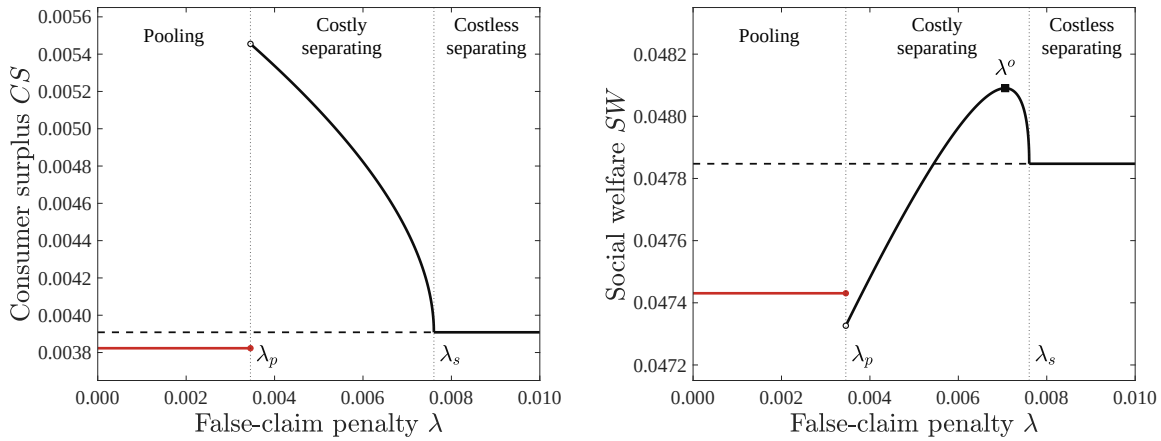
Figure 3(c) depicts the comparative static result in Proposition 7. When $\lambda \le \lambda_p$, the penalty leaves the high-type firm's profit unchanged but directly reduces the low-type firm's profit from cyberwashing. When the equilibrium shifts from pooling to costly separation at $\lambda = \lambda_p^+$, the low-type firm no longer engages in cyberwashing and therefore avoids the false-claim penalty. Nevertheless, its profit jumps downward because it loses the demand benefit from being pooled with the high type. At the tipping point, the high-type firm is indifferent between pooling and separation, so its profit is continuous. As λ continues to increase, the false-claim penalty provides greater deterrence against mimicry, allowing the high type to reduce its price distortion; its equilibrium profit therefore increases. When λ reaches or exceeds λ_s, both types achieve their symmetric-information profits.

We next explore how cyberwashing affects consumers and social welfare. Consumer surplus is the expected surplus of consumers net of their expected breach losses, and social welfare is the sum of consumer surplus, the firm's expected profit (net of its own breach losses and of the penalty), and the regulator's expected penalty revenue. The objective treats the monetary fine as a transfer and excludes hackers' surplus, which we do not regard as legitimate surplus, as well as the costs of enforcement and of any governance measure that changes γ; comparisons across λ and γ therefore describe the market response to those parameters rather than the net benefit of the interventions that change them. The formal definitions, the treatment of the penalty as a reputational loss, and a type-wise decomposition of welfare are given in Online Supplement A.9. The decomposition is what distinguishes the welfare effects of the penalty from its consumer-surplus effects: for a firm of known type i, the sum of consumer surplus and profit is maximized at a customer base q_i^W that exceeds the monopoly demand, because a monopolist serves too few customers, so the high type's demand expansion improves welfare up to q_H^W and reduces it beyond.

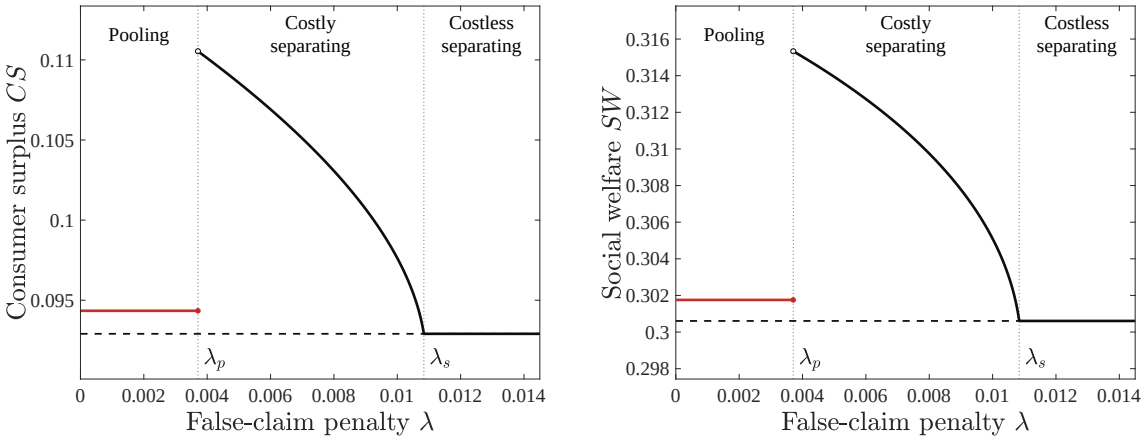
PROPOSITION 8. *For $\lambda_p(\gamma) > 0$, as the false-claim penalty λ increases, consumer surplus and social welfare change as follows.*

(i) Consumer surplus remains constant for $\lambda \le \lambda_p$, increases discontinuously at $\lambda = \lambda_p^+$, strictly decreases for $\lambda_p < \lambda < \lambda_s$, and remains constant for $\lambda \ge \lambda_s$. Consequently, $\sup_{\lambda>0} CS(\lambda) = \lim_{\lambda \downarrow \lambda_p} CS(\lambda)$, which is not attained because the LMSE selects pooling at $\lambda = \lambda_p$.

(ii) Social welfare remains constant for $\lambda \le \lambda_p$ and for $\lambda \ge \lambda_s$. At $\lambda = \lambda_p^+$, it decreases if and only if $\gamma > \hat{\gamma}$ and $\underline{F}_s < F < \overline{F}_s$, and weakly increases otherwise. On $\lambda_p < \lambda < \lambda_s$, it is strictly concave in λ. If $\gamma \le 1/\sqrt{2z(1-z)}$, social welfare jumps strictly upward at λ_p^+ and then strictly decreases throughout the costly separating regime; its global supremum is approached as $\lambda \downarrow \lambda_p$. If $\gamma > 1/\sqrt{2z(1-z)}$, social welfare first increases and then decreases within costly separation and has its unique global maximum at $\lambda^o \in (\lambda_p, \lambda_s)$, given in (18) *in Online Supplement A.9, where the high type's customer base equals its welfare-optimal level. Moreover, $SW^s(\lambda^o) > SW^p$, regardless of the sign of the welfare jump at λ_p^+.*



(a) $z = 0.19$, $F = 0.0005$, $\gamma = 5.2$



(b) $z = 0.4$, $F = 0.3$, $\gamma = 0.15$

Figure 4 Effect of false-claim penalty λ on consumer surplus and social welfare. Panel (a) is a severe-threat case; panel (b) is a weak-threat case.

Figure 4 illustrates Proposition 8 under two parameter settings. Panel (a) considers a severe hacker threat, with γ close to the admissible bound $\bar{\gamma}_B = 1/z$, whereas panel (b) considers a weak threat and a high firm loss per attack F. Although a greater false-claim penalty deters cyberwashing, an excessively severe penalty may inadvertently reduce consumer surplus: Proposition 8(i) implies that consumer surplus is highest for penalties just above λ_p, the level at which separation becomes the selected outcome, and that increasing λ further hurts consumers. Because the LMSE selects pooling at λ_p itself, consumer surplus has an unattained supremum as $\lambda \downarrow \lambda_p$. No smallest deterrent penalty exists under the maintained selection rule. Under an alternative tie-breaking convention that selects separation at λ_p, that boundary penalty would attain the same surplus value. The intuition is that, within the separating regime $\lambda > \lambda_p$, a greater penalty reduces the downward price distortion required for the high-type firm to sustain separation, leading to a higher equilibrium price by the high type, as shown in Proposition 5. Although the high-type firm's attack

intensity decreases with λ (Proposition 6), the price effect dominates in determining the overall impact of λ on consumer surplus (consumer surplus panels of Figure 4).

The welfare implications are more nuanced. At $\lambda = \lambda_p$, the high-type firm is indifferent between pooling and costly separation, and the false-claim penalty paid by the low-type firm under pooling is offset by the regulator's penalty revenue. Hence, the welfare comparison is driven by how separation reallocates consumers and attack intensity across the two firm types: the low-type firm reveals its true cybersecurity level, raises its price, serves fewer consumers, and attracts fewer attacks, whereas the high-type firm lowers its price, expands its market, and attracts more attacks.

Whether separation raises welfare at the deterrence threshold depends on how much attack exposure the secure firm's market expansion creates. When hacker threats are weak ($\gamma \leq 1/\sqrt{2z(1-z)}$), the welfare jump is strictly positive, as in panel (b) of Figure 4, and consumer surplus and welfare approach their respective suprema as $\lambda \downarrow \lambda_p$. Under more severe threats, the welfare jump is negative when $\gamma > \hat{\gamma}$ and the firm's breach loss lies in $(\underline{F}_s, \overline{F}_s)$ within the pooling region, and weakly positive otherwise. The lower threshold can equal zero, so even very small firm breach losses need not ensure a positive welfare jump. When $\gamma > 1/\sqrt{2z(1-z)}$, the signaling discount just above λ_p expands the secure firm's customer base beyond its welfare-optimal level. A higher penalty then improves welfare by allowing the firm to shrink its customer base toward that level, which it reaches at λ^o; further increases reduce welfare. This divergence between the consumer-surplus and welfare objectives reflects the attack-exposure cost of demand expansion, which is absent in the benchmark with exogenous attacks (Proposition A1(iii)).

A negative welfare jump at the deterrence threshold therefore does not imply that tolerating cyberwashing is optimal. Proposition 8(ii) establishes analytically that $SW^s(\lambda^o) > SW^p$ whenever the interior optimum exists. In panel (a) of Figure 4, a penalty just above λ_p lowers welfare, but increasing the penalty sufficiently within the costly separating regime raises welfare above its pooling level and yields the unique maximum at λ^o. The policy implication is to combine deterrence with an appropriate reduction in the high type's signaling discount, rather than to stop immediately above the deterrence threshold.

Corollary 1 compares consumer surplus and social welfare under pooling and costless separation.

COROLLARY 1. *(i) Consumer surplus is higher under pooling than under costless separation if and only if* $0 < \gamma < \gamma_c$ *and* $F > F_c$. *(ii) Social welfare is higher under pooling than under costless separation if and only if* $0 < \gamma < \gamma_w$ *and* $\underline{F}_w < F < \overline{F}_w$. *Moreover, the thresholds satisfy* $\gamma_w < \gamma_c$ *and* $F_c < \underline{F}_w < \overline{F}_w$.

Consumer surplus, and more rarely social welfare, can be higher under pooling than under costless separation when hacker threats are weak and the firm's breach loss is relatively high; see

Figure 4(b). With weak threats, the difference in breach exposure between the two types matters little to consumers, so the benefit of identifying the secure firm is modest, whereas a vulnerable firm whose type is revealed raises its price and serves fewer consumers; under pooling it keeps the lower common price. For welfare the condition is more restrictive, because separation also shifts consumers and breach losses toward the more secure firm: pooling improves welfare only when threats are weak enough and breach losses are moderate, large enough for the price advantage of pooling to matter but not so large that concentrating demand at the vulnerable firm becomes too costly. Eliminating cyberwashing through an overly large penalty may therefore worsen welfare when hacker threats are weak and breach losses are moderate.

Overall, our findings offer important policy implications for regulating cyberwashing. Recall that λ represents the *expected* penalty on false claims; it reflects regulatory stringency and can be increased by strengthening the detection of false claims or by imposing larger penalties once a false claim is detected. To best protect consumers, regulators should set this stringency level by accounting for firms' cyberwashing incentives and for their equilibrium pricing responses.[11] To approach the consumer-surplus supremum when $\lambda_p > 0$, the penalty should approach λ_p from above, deterring cyberwashing while retaining the high type's signaling discount. This conclusion differs in mechanism from the closest false-claim models. In Wu and Geylani (2020), the smallest penalty that induces truthful claims maximizes both consumer surplus and welfare, because higher penalties discourage consumers' costly investigation of claims and induce more sophisticated deception; here consumers do not investigate, the penalty interacts with the firm's price instead, and the welfare-maximizing penalty can lie strictly above the limiting consumer-optimal boundary λ_p (Proposition 8(ii)). Our analysis likewise favors bounded penalties rather than arbitrarily severe regulation, as in Corts (2014), but the cost of a larger penalty here is a higher price paid to a secure firm rather than a loss of the information that claims convey. If social welfare is the objective, regulators should also account for the external cybersecurity governance environment that shapes the hacker payoff-to-cost ratio γ. When this environment is weak (γ is high relative to $\bar{\gamma}_B$), stronger oversight of cybersecurity claims may be desirable to better balance firm profits and consumer surplus. Although our analysis abstracts from the costs of false-claim detection and enforcement, our main results imply that it may be socially undesirable to impose excessively stringent regulation even when it is costless to do so.

[11] Indeed, in practice, consumers care not only about cybersecurity but often even more about price. For example, when 1Password linked a subscription price increase to continued investment in innovation and cybersecurity, some users criticized the increase and considered switching to lower-priced alternatives. This example illustrates how a higher price can offset consumers' benefits from stronger cybersecurity. See https://www.macrumors.com/2026/02/24/1password-march-price-increase/ and https://www.1password.community/1password-at-home-31/what-justifies-the-huge-subscription-price-increase-24103.

5.2. Effect of the Hacker Payoff-to-Cost Ratio

The following proposition characterizes how the hacker payoff-to-cost ratio γ affects consumer surplus and social welfare. We focus on $0 < \lambda < \bar{\lambda}_p$ and parameter values satisfying $\bar{\gamma}_2(\lambda) < \bar{\gamma}_B$ (a sufficient condition is $F \geq F^o$; Proposition 4), under which the LMSE shifts from separation to pooling and then back to separation as γ increases, so that both equilibrium transitions induced by γ can be examined. Because pooling arises only for intermediate γ, the comparative statics in γ are more intricate and need two further thresholds. For a given λ, let $\bar{\gamma}_0(\lambda)$ denote the smaller solution of $\lambda_s(\gamma) = \lambda$, the hacker payoff-to-cost ratio at which the equilibrium moves from costless to costly separation as γ increases (so that $\bar{\gamma}_0 < \bar{\gamma}_1$; see Figure 5). The threshold $\lambda^+ > 0$ in part (ii) is a bound on the penalty, defined in the proof, below which $\bar{\gamma}_0(\lambda)$ is close enough to zero for the derivative bounds used there to hold.

PROPOSITION 9. *As the hacker payoff-to-cost ratio γ increases, consumer surplus and social welfare change as follows.*

(i) Consumer surplus increases discontinuously at $\gamma = \bar{\gamma}_2^+$, where the equilibrium shifts from pooling to costly separating. At this point, social welfare weakly increases if $\bar{\gamma}_2 \leq \hat{\gamma}$, or if $\bar{\gamma}_2 > \hat{\gamma}$ and either $F \leq \underline{F}_s(\bar{\gamma}_2)$ or $F \geq \overline{F}_s(\bar{\gamma}_2)$, and decreases if $\bar{\gamma}_2 > \hat{\gamma}$ and $\underline{F}_s(\bar{\gamma}_2) < F < \overline{F}_s(\bar{\gamma}_2)$.

(ii) Under the costly separating equilibrium, if $0 < \lambda < \lambda^+$, consumer surplus and social welfare are increasing in $\gamma \in [\bar{\gamma}_0, \bar{\gamma}_0 + \epsilon]$ for some $\epsilon > 0$.

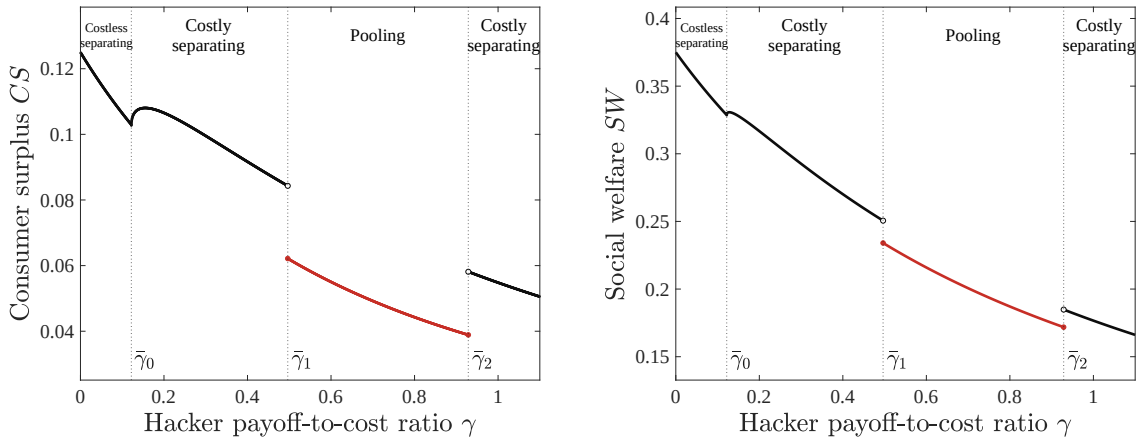


Figure 5 Effect of hacker payoff-to-cost ratio γ on consumer surplus and social welfare (Parameters: $z = 0.4$, $F = 0.05$, $\lambda = 0.01$).

As illustrated in Figure 5, consumer surplus and social welfare drop sharply when the pooling equilibrium arises, relative to the separating equilibrium.[12] An increase in the hacker payoff-to-cost

[12] Figure 5 presents the case in which separation is welfare-improving at $\gamma = \bar{\gamma}_2^+$. In the complementary parameter region, the welfare comparison between pooling and separation can be reversed.

ratio can benefit consumers by restoring separation; its welfare effect depends on the conditions in Proposition 9(i).

Within each equilibrium regime, consumer surplus and social welfare are generally decreasing in the hacker payoff-to-cost ratio, which is intuitive. An exception arises when γ increases from the costless separating regime into the costly separating regime. In this region, consumer surplus and social welfare can increase with γ, generating the inverted-U pattern shown in Figure 5. This effect is driven by the high-type firm's price distortion: just above the costless-to-costly separation boundary, the emerging downward price distortion can outweigh the direct effect of greater attack exposure and thereby raise consumer surplus and welfare.

These results suggest that, unless external cybersecurity governance is strong enough to suppress the hacker payoff-to-cost ratio, moderate governance can be welfare-reducing: by only partially reducing the ratio, it may create a cyberwashing trap in which hacker threats no longer deter low-type firms from making false cybersecurity claims. As emphasized in the previous section, this is a statement about the market response to γ with λ and the firms' security levels held fixed, under the LMSE selection, and it abstracts from the cost of the governance measure and from any effect of that measure on hackers' surplus; it does not say that hacker threats are desirable, only that anti-hacker governance and false-claim enforcement should be designed together. The possibility that cybercrime enforcement generates unintended effects is also consistent with recent empirical evidence: Wang et al. (2024) find that enforcement of computer misuse laws reduces not only potentially harmful activity but also the quantity and cybersecurity relevance of contributions to non-targeted discussions on public hack forums, a chilling effect they attribute to uncertainty about the prosecution of dual-use activities. Although the mechanism differs from ours, their findings illustrate that cybersecurity governance can alter behavior beyond its intended direct deterrence effect; our analysis identifies a distinct market-mediated consequence, in which partially effective governance weakens the signaling discipline that otherwise discourages cyberwashing.

6. Robustness and Extensions

This section examines the robustness of the baseline results and extends the model; the evidence differs across subsections, and we state its nature at the outset. Section 6.1 (costly pre-discovery effort) contains an analytical characterization of the LMSE structure and a numerical welfare analysis (details in Online Supplement A.7); Section 6.2 (demand-dependent breach loss) contains analytical existence results and a numerical selection and welfare analysis; Sections 6.3 (general prior) and 6.4 (type-dependent marginal costs and breach-triggered penalties) summarize analytical results stated in Online Supplement A.10 and A.11. Online Supplement A.8 documents the parameter grids behind every numerical claim. In addition, Online Supplement A.1, A.2, and A.3

show that our model can readily accommodate an arbitrary level of consumer loss, congestion effects among hackers, and a positive security level for the low-type firm through equivalent transformations.

6.1. Alternative Hacker Timing: Costly Pre-Discovery Effort

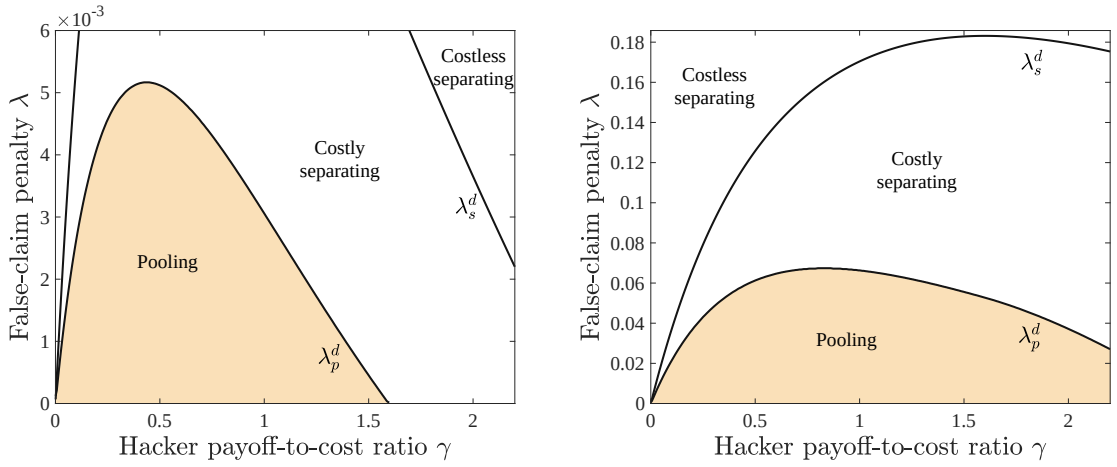
Thus far we have assumed that hackers incur the effort cost c only after identifying an exploitable vulnerability, so that the firm's claim and price affect attacks only through the customer base they generate (the *demand channel*). An alternative, and also practically relevant, setting is one in which hackers must incur the effort cost ex ante, before learning whether the firm has an exploitable vulnerability; this timing fits systems whose vulnerabilities are difficult to detect through public or automated scans, for example because access is gated by authentication or traffic is screened by firewalls. A hacker who attacks first pays c to probe the system and obtains h per active consumer only if an exploitable vulnerability exists. Hackers' expected payoff then depends not only on the firm's consumer base but also on their belief about its cybersecurity type (the *inference channel*), and this channel is what changes the conclusions.

Formally, consumers and hackers share the prior α and form a common posterior β after observing the claim and price, so the perceived probability of an exploitable vulnerability is $1-\beta z$. A hacker who probes and attacks obtains $U_h = h(1-\beta z)\bar{N}_c - c$, so the attack intensity $N_h = \gamma(1-\beta z)N_c$ is now unconditional: it counts all hackers who probe the firm, whether or not a vulnerability exists. The consumer's expected utility is $U_c = v - p - \bar{N}_h(1-\beta z)l$, both parties hold rational expectations, and we impose $\bar{c} > \max\{hF, h/4\}$, which ensures positive demands and an interior participation cutoff. To ensure the existence of a pure-strategy LMSE for every $\lambda > 0$, we focus on the case $F \le 4(1-z)$ in Proposition 10(i). This restriction is not required for the results in part (ii).

Proposition 10. *(i) If $F \le 4(1-z)$, there exist thresholds $0 \le \lambda_p^d(\gamma) < \lambda_s^d(\gamma)$ such that the unique LMSE outcome is pooling for $0 < \lambda < \lambda_p^d(\gamma)$, costly separating for $\lambda_p^d(\gamma) < \lambda < \lambda_s^d(\gamma)$, and costless separating for $\lambda \ge \lambda_s^d(\gamma)$. At $\lambda = \lambda_p^d(\gamma) > 0$, pooling is an LMSE outcome; separation is also an LMSE outcome if and only if both types are indifferent between the best pooling and separating outcomes.*

(ii) For sufficiently small $\lambda > 0$, the LMSE is costless separating for sufficiently low γ and pooling for intermediate values of γ. If $F \ge 1/4$, then for γ sufficiently close to $1/F$, the LMSE is costly separating if $0 < z < z_d(F)$ and pooling if $z_d(F) < z < 1$.

Proposition 10 and Figure 6 show that the three-regime structure in the penalty carries over under $F \le 4(1-z)$: separation is sustained when the penalty is high, and pooling can arise for intermediate hacker payoff-to-cost ratios. What changes is the role of strong threats. For sufficiently small penalties, when the breach loss is relatively high ($F \ge 1/4$), pooling may persist up to



(a) $z = 0.1$ (b) $z = 0.85$

Figure 6 LMSE equilibrium outcomes under costly pre-discovery effort (Parameters: $F = 0.4$).

the admissible bound if the high type's cybersecurity level is high enough ($z > z_d$; Figure 6(b)), whereas in the baseline separation is restored before the bound in the corresponding case $F \geq F^o$ (Proposition 4(i)). The reason is the inference channel: mimicking the high type now gives the low-type firm two benefits, better consumer beliefs and weaker hacker incentives for a given customer base, and the second benefit is strong when z is high. For the same reason, the attack intensity faced by the high type can fall rather than rise when the equilibrium moves from pooling to separation, because being identified as secure discourages hacker effort (Online Supplement A.7).

Outside the sufficient condition $F \leq 4(1 - z)$, the inference channel also creates a possibility absent from the baseline, where the secure firm never prefers to be taken for the vulnerable type. Because being believed secure now also reduces hacker participation for a given customer base, mimicry is more valuable to the vulnerable firm, and deterring it can require a price cut so deep that the secure firm would rather forgo its claim; if the penalty at the same time makes pooling infeasible, neither pooling nor separation is an equilibrium outcome in the maintained pure-strategy class. This can occur for large z and F (e.g., $z = 0.9$, $F = 2$, $\gamma = 0.4$, and $0.063 < \lambda < 0.086$); in Figure 6 and in the costly pre-discovery cases of the numerical analysis of Online Supplements A.7 and A.8, the condition $F \leq 4(1 - z)$ holds, so Proposition 10(i) guarantees an LMSE for every $\lambda > 0$.

What survives is the effect of the penalty on consumers and welfare, established numerically because the equilibrium expressions involve higher-order terms in γ (Online Supplement A.7 and A.8): on the reported penalty grids, consumer surplus, and in our examples social welfare, attain their largest evaluated values at intermediate penalties, and stricter regulation then reduces them through the firm's pricing response. Within a regime in the reported numerical comparisons, costly pre-discovery raises consumer surplus and welfare, because hackers who pay before learning

whether a vulnerability exists attack less; but for penalties that deter cyberwashing only under the baseline timing, it moves the market into pooling and can lower both for some of these penalties. Online Supplement A.7 turns this into comparative statics in discoverability, letting a fraction δ of hackers observe vulnerabilities before incurring the attack cost: lower discoverability enlarges the cyberwashing region over the reported parameter range and raises consumer surplus within each regime. Interventions that make discovery more difficult, such as limiting automated scanning and protecting login credentials, can therefore improve outcomes within each regime but also raise the penalty required to deter cyberwashing and can move the market into pooling; they should be evaluated jointly with false-claim regulation.

6.2. Demand-Dependent Breach Loss

In the baseline model, each successful attack imposes a fixed loss F on the firm. In many cybersecurity incidents, however, the firm's loss also depends on the size of the affected consumer base. For example, breach notification, customer support, remediation, and legal exposure may increase with the number of active consumers. We therefore extend the firm's profit function to

$$\pi_H(r,p) = p \cdot N_c(r,p) - N_h(r,p) \cdot (1-z)(F + \theta \cdot N_c), \tag{13}$$

$$\pi_L(r,p) = p \cdot N_c(r,p) - N_h(r,p) \cdot (F + \theta \cdot N_c) - \lambda \cdot \mathbf{1}\{r = r_H\}, \tag{14}$$

where $\theta \geq 0$ measures the degree to which firm-side breach losses scale with the firm's consumer base. The baseline model corresponds to $\theta = 0$. As in the baseline analysis, we maintain the restriction $\bar{c} > \max\{hF, hz\}$ to ensure positive demand and an interior hacker participation cutoff. To simplify the analytical existence results, we focus on $F \geq z$ in Proposition 11; the numerical analysis that follows does not impose this restriction and includes cases with $F < z$. The following proposition characterizes the existence of separating and pooling equilibria.

Proposition 11. *Under demand-dependent breach loss, (i) a costly separating equilibrium exists if $0 < \gamma \leq \gamma_s^b$ and $0 < \lambda < \lambda_s^b$; otherwise, a costless separating equilibrium exists if $\gamma > \gamma_s^b$, or if $\lambda \geq \lambda_s^b$; (ii) a pooling equilibrium exists if $0 < \gamma < \gamma_p^b$ and $0 < \lambda < \lambda_1^b$. At $\lambda = \lambda_1^b > 0$, pooling exists if the candidate price differs from the low type's symmetric-information price.*

Proposition 11 establishes that the equilibrium existence results follow the same qualitative pattern as Propositions 1 and 2 in the main model. A costless separating equilibrium can be sustained when either the hacker payoff-to-cost ratio γ or the false-claim penalty λ is sufficiently high. By contrast, costly separation can arise when both parameters are relatively low, and pooling can also be sustained when γ and λ are sufficiently low. As in the base model, multiple PBEs can arise within the same parameter ranges, and we apply the LMSE refinement; because the

associated profit comparisons are algebraically complex and do not admit a tractable closed-form characterization, we proceed numerically. Setting $z = 0.1$, $F = 0.4$, and $\theta = 0.5$, Figure 7 shows that the results in Propositions 3 and 4 for $F \geq z$ carry over when the firm's breach loss increases with its customer base: pooling emerges as the LMSE when the false-claim penalty is low and the hacker payoff-to-cost ratio is moderate.

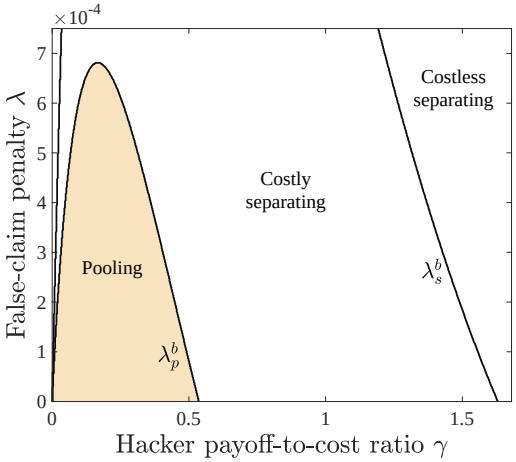


Figure 7 LMSE equilibrium outcomes under demand-dependent breach loss (Parameters: $z = 0.1$, $F = 0.4$, $\theta = 0.5$).

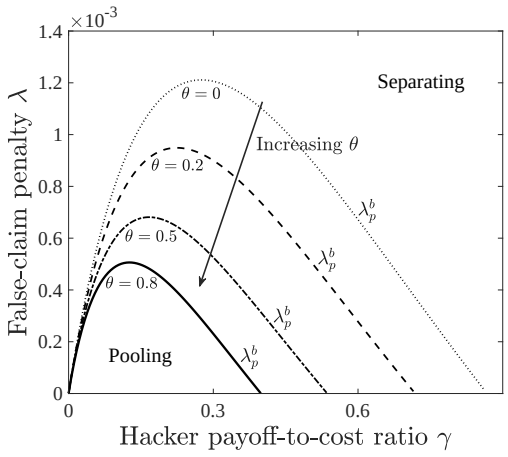


Figure 8 Effect of θ on the pooling-separating boundary λ_p^b (Parameters: $z = 0.1$, $F = 0.4$).

Furthermore, we examine how the demand-sensitive loss parameter θ affects the equilibrium outcomes by computing the pooling–separating boundary λ_p^b as θ ranges from 0 to 0.8. Figure 8 shows that λ_p^b shifts downward as θ increases, so, in these numerical comparisons, a larger θ expands the parameter region in which the LMSE selects separation and contracts the pooling region. The intuition is that a higher θ makes demand expansion more costly for the low-type firm, because attracting more consumers increases its expected breach loss; this weakens its incentive to mimic, so the high type needs less price distortion to sustain separation. Thus, the LMSE selects separation under a wider range of the examined parameters when θ is greater, and regulations that tie firms' breach-related losses more closely to the size of their customer base can help mitigate cyberwashing.

The welfare analysis for this extension is numerical and is documented in Online Supplement A.8: on the reported grid of (z, F, γ, θ) including $F < z$, consumer surplus has its penalty-grid maximum at the first separating grid point; welfare has its grid maximum either there or at a later grid point below the costless-separation boundary. No reported grid maximum occurs in the costless separating regime.

6.3. Impact of the Likelihood of the Firm Being the High Cybersecurity Type

In the main analysis, we assume equal likelihood of the firm being the high cybersecurity type, $\alpha = 1/2$. Online Supplement A.10 allows a general prior belief $\alpha \in (0, 1)$ and shows that the equilibrium

structure is unchanged. Proposition A3 establishes that the costly-to-costless separating boundary λ_s is independent of α, because the firm's type is fully revealed under separation, whereas the pooling–separating boundary λ_p increases with α, because a more optimistic prior raises the high type's pooling payoff and allows pooling to be sustained under a higher penalty. Consequently, for any penalty $0 < \lambda < \lambda_s$ there is a unique cutoff $\alpha_p(\lambda) \in (0,1)$ such that the LMSE is costly separating for $\alpha < \alpha_p(\lambda)$ and pooling for $\alpha \geq \alpha_p(\lambda)$: a higher prior probability of the high-security type expands the parameter region in which the low type cyberwashes. This concerns the low type's strategy: conditional on pooling, the ex ante incidence of false claims equals $1-\alpha$ and decreases with α (Figure A6).

Varying the prior reveals an additional consumer-surplus effect. Proposition A4 shows that both types' profits are constant in α under costly separation and increasing in α under pooling, with the low type's profit jumping upward at α_p, and that consumer surplus increases with α within each regime but falls discontinuously at α_p, where the equilibrium switches to pooling. A more favorable prior can therefore encourage cyberwashing and reduce consumer surplus at the transition, even though it raises consumer surplus within each regime. Whether consumer surplus just before the switch exceeds even the limiting pooling consumer surplus as $\alpha \uparrow 1$ depends on the penalty: it does if and only if λ exceeds a threshold λ_α, because a higher penalty raises the switching point α_p and thereby places more weight on the high type's surplus (Figure A7).

6.4. Security-Dependent Marginal Costs and Breach-Triggered Penalties

The downward price distortion rests on the fact that expanding demand is cheaper for the secure firm, and Online Supplement A.11 examines two modifications that bear directly on this property. If serving a customer also entails a type-specific marginal cost $0 \leq k_i < 1-\gamma(1-z_i)F$ for $i = H, L$ (for instance because stronger protection raises the cost of operating each account), the secure firm retains the lower effective marginal cost, and hence the downward price distortion, exactly when $k_H - k_L < \gamma z F$, that is, when the attack-exposure advantage of security outweighs any direct cost disadvantage; the analysis goes through with $z\gamma F$ replaced by $\gamma z F - (k_H - k_L)$ in the rent of Lemma 2, and when the inequality is reversed the secure firm separates, when it must, by distorting its price upward (Remark A1). If the detection of a false claim is triggered partly by breaches, so that the low type's expected penalty for claiming r_H is $\lambda_0 + \lambda_1 N_h$, the variable part of the penalty raises the low type's effective marginal cost and leaves the high type's unchanged, so the cost gap in (8) rises from $z\gamma F$ to $z\gamma F + \gamma\lambda_1$: the single-crossing arguments are reinforced, the separating regions weakly expand, and the constant penalty is the conservative case (Remark A2).

7. Conclusion

Unverifiable cybersecurity claims create a growing risk that firms may overstate their protection of consumer data. This paper develops a signaling model of cyberwashing in which a firm privately observes its cybersecurity level and chooses both a cybersecurity claim and a price, consumers infer security from these signals, and hackers decide whether to attack based on attack costs and the expected payoff from compromising the firm's consumer base. The distinctive feature of the environment, relative to standard false-claim models, is that expanding the customer base generates a smaller marginal expected breach loss for the secure firm; this single-crossing property makes a price cut a credible signal of security, gives the secure firm an informational rent that grows with its customer base, and ties the credibility of claims to the firm's breach loss and to the strength of hacker incentives. Under the selected equilibrium, stronger false-claim regulation is not always welfare-improving: when $\lambda_p > 0$, consumer surplus has an unattained supremum as $\lambda \downarrow \lambda_p$, because more stringent regulation reduces the high-type firm's need to signal through downward price distortion and thereby raises prices. Welfare has the same limiting optimum under weak hacker threats; under severe threats, its unique global maximum is attained at $\lambda^o \in (\lambda_p, \lambda_s)$, where the secure firm's customer base equals its welfare-optimal level. Although pooling may dominate costless separation, appropriately calibrated costly separation yields higher welfare whenever $\lambda_p > 0$. Similarly, broader cybersecurity governance that lowers the hacker payoff-to-cost ratio or makes vulnerabilities harder to discover may reduce welfare by weakening the disciplining effect of cyber threats and creating a cyberwashing trap. When evaluating anti-cyberwashing regulation, policymakers should therefore consider not only its direct deterrence effect but also firms' pricing responses and hacker behavior within the broader governance environment.

Our analysis also points to several extensions that the framework can accommodate. We take the firm's cybersecurity level, the enforcement technology, and the expected penalty as given; endogenizing security investment would allow false-claim penalties to affect investment incentives as well as disclosure, and endogenizing detection would let enforcement respond to breaches, which Remark A2 suggests reinforces the mechanism. We also rely on an explicit selection rule for the transitions between pooling and separation, whereas the comparative statics within the separating regimes do not depend on it. Extending the model to competing firms and to a dynamic setting, in which consumers learn from incidents and sanctions and firms build reputations, would clarify how competition and reputation concerns shape cyberwashing incentives; we view the demand-cost signaling mechanism identified here as a building block for these richer environments.

References

Alibeiki H, Gümüş M (2020) Supply competition under quality scores: Motivations, information sharing and credibility. *International Journal of Production Economics* 226:107612.

August T, Niculescu MF, Shin H (2014) Cloud implications on software network structure and security risks. *Information Systems Research* 25(3):489–510.

Avinadav T, Chernonog T, Perlman Y (2014) Analysis of protection and pricing strategies for digital products under uncertain demand. *International Journal of Production Economics* 158:54–64.

Avinadav T, Chernonog T, Shamir N (2025) Information provision from a platform to competing sellers: The role of strategic ambiguity. *Manufacturing & Service Operations Management* 27(1):269–286.

Baksy A, Caratelli D, Olson LM (2025) Cyberattacks and firm size: The vulnerability of mid-size firms. Office of Financial Research, The OFR Blog.

Banks JS, Sobel J (1987) Equilibrium selection in signaling games. *Econometrica* 55(3):647–661.

Chen G, Zou T (2026) Personalized pricing with consumers' quality uncertainty. *Production and Operations Management* 1–20.

Chen Z, Zhang M (2026) Personalized pricing in the presence of privacy concerns. *Management Science* to appear, Forthcoming.

Cho IK, Kreps DM (1987) Signaling games and stable equilibria. *Quarterly Journal of Economics* 102(2):179–221.

Choe C, Matsushima N, Shekhar S (2025) The bright side of the GDPR: Welfare-improving privacy management. *Management Science* 71(8):6836–6858.

Cisco (2024) Cisco 2024 consumer privacy survey. https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cisco-consumer-privacy-report-2024.pdf.

Corts KS (2013) Prohibitions on false and unsubstantiated claims: Inducing the acquisition and revelation of information through competition policy. *Journal of Law and Economics* 56(2):453–486.

Corts KS (2014) Finite optimal penalties for false advertising. *The Journal of Industrial Economics* 62(4):661–681.

De Cornière A, Taylor G (2026) A model of information security and competition. *Marketing Science* 45(2):318–334.

Doshi AR, Schmidt W (2024) Soft governance across digital platforms using transparency. *Strategy Science* 9(2):185–204.

Fainmesser IP, Galeotti A, Momot R (2023) Digital privacy. *Management Science* 69(6):3157–3173.

Gordon LA, Loeb MP (2002) The economics of information security investment. *ACM Transactions on Information and System Security* 5(4):438–457.

Guan X, Cao H, Li KJ, Ding Y (2025) Product safety and liability with deceptive advertising and moral hazard. *Marketing Science* 44(2):287–305.

Guan X, Guo Y, Jiang Y, Ma G, Tan Y (2026) Quality verification in the presence of review bias. *Production and Operations Management* 35(7):2871–2889.

Hu M, Momot R, Wang J (2022) Privacy management in service systems. *Manufacturing & Service Operations Management* 24(5):2761–2779.

Hui KL, Ke PF, Yao Y, Yue WT (2019) Bilateral liability-based contracts in information security outsourcing. *Information Systems Research* 30(2):411–429.

Hui KL, Kim SH, Wang QH (2017) Cybercrime deterrence and international legislation: Evidence from distributed denial of service attacks. *MIS Quarterly* 41(2):497–523.

Jiang B, Ni J, Srinivasan K (2014) Signaling through pricing by service providers with social preferences. *Marketing Science* 33(5):641–654.

Ke TT, Sudhir K (2023) Privacy rights and data security: GDPR and personal data markets. *Management Science* 69(8):4389–4412.

Kummer M, Schulte P (2019) When private information settles the bill: Money and privacy in Google's market for smartphone applications. *Management Science* 65(8):3470–3494.

Lai G, Xiao W, Zhao X (2026) Strategic financing and information revelation amid market competition. *Manufacturing & Service Operations Management* Articles in Advance.

Lee CH, Geng X, Raghunathan S (2013) Contracting information security in the presence of double moral hazard. *Information Systems Research* 24(2):295–311.

Lei YM, Miao S, Momot R (2024) Privacy-preserving personalized revenue management. *Management Science* 70(7):4875–4892.

Li X, Li KJ (2023) Beating the algorithm: Consumer manipulation, personalized pricing, and big data management. *Manufacturing & Service Operations Management* 25(1):36–49.

Li X, Wang XS, Nault BR (2024) Is personalized pricing profitable when firms can differentiate? *Management Science* 70(7):4184–4199.

Luo S, Choi TM (2022) E-commerce supply chains with considerations of cyber-security: Should governments play a role? *Production and Operations Management* 31(5):2107–2126.

Mailath GJ, Okuno-Fujiwara M, Postlewaite A (1993) Belief-based refinements in signalling games. *Journal of Economic Theory* 60(2):241–276.

Mandiant (2024) UNC5537 targets Snowflake customer instances for data theft and extortion.

Montes R, Sand-Zantman W, Valletti T (2019) The value of personal information in online markets with endogenous privacy. *Management Science* 65(3):1342–1362.

National Cyber Security Centre and National Crime Agency (2023) Ransomware, extortion and the cyber crime ecosystem.

Piccolo S, Tedeschi P, Ursino G (2015) How limiting deceptive practices harms consumers. *The RAND Journal of Economics* 46(3):611–624.

Piccolo S, Tedeschi P, Ursino G (2018) Deceptive advertising with rational buyers. *Management Science* 64(3):1291–1310.

Png IP, Wang QH (2009) Information security: Facilitating user precautions vis-à-vis enforcement against attackers. *Journal of Management Information Systems* 26(2):97–121.

PwC (2024) Voice of the consumer survey 2024.

Rao AR, Monroe KB (1989) The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research* 26(3):351–357.

Rhodes A, Wilson CM (2018) False advertising. *The RAND Journal of Economics* 49(2):348–369.

Schmidt W, Buell RW (2017) Experimental evidence of pooling outcomes under information asymmetry. *Management Science* 63(5):1586–1605.

Tian L, Jiang B (2017) Comment on "strategic information management under leakage in a supply chain". *Management Science* 63(12):4258–4260.

Unit 42 (2026) 2026 Unit 42 global incident response report.

Verizon (2026) Verizon data breach investigations report 2026.

Völckner F, Hofmann J (2007) The price-perceived quality relationship: A meta-analytic review and assessment of its determinants. *Marketing Letters* 18(3):181–196.

Wang QH, Geng R, Kim SH (2024) Chilling effect of the enforcement of computer misuse act: Evidence from publicly accessible hack forums. *Information Systems Research* 35(3):1195–1215.

Wu Y, Geylani T (2020) Regulating deceptive advertising: False claims and skeptical consumers. *Marketing Science* 39(4):788–806.

Wu Y, Zhang K, Xie J (2020) Bad greenwashing, good greenwashing: Corporate social responsibility and information transparency. *Management Science* 66(7):3095–3112.

Xin M (2020) The impact of customer valuation uncertainty on software licensing. *MIS Quarterly* 44(2):561–603.

Xu A, Tan Y, He QC (2024) Information transparency with targeting technology for online service operations platform. *Production and Operations Management* 33(6):1410–1425.

Xu F, Wang X, Zhang F (2025) Consumer privacy in online retail supply chains. *Management Science* 71(10):8371–8389.

Xu Y, Dai H, Yan W (2026) Identity disclosure and anthropomorphism in voice chatbot design: A field experiment. *Management Science* 72(1):223–241.

Zhou B, Zou T (2023) Competing for recommendations: The strategic impact of personalized product recommendations in online marketplaces. *Marketing Science* 42(2):360–376.

Zhu M, August T, Niculescu MF (2026) Provision of cybersecurity resources to SMBs: The investment pitfall of security interdependence, Working Paper.

Zhu M, Song M (2024) Design information disclosure under bidder heterogeneity in online advertising auctions: Implications of bid-adherence behavior, Working Paper.

Online Supplement A. Additional Analyses

This part of the Online Supplement collects the demand and attack intensities in each equilibrium, the benchmark with exogenous attack intensity, the equilibrium refinements, the extension to a general prior belief, and the robustness checks referred to in the main text.

A.1. Normalization of Consumer Loss l

Before imposing the normalization, suppose that each attacked consumer suffers a loss $l > 0$. There is a unit mass of potential hackers. Let h_0 denote a hacker's gain per active consumer, $\gamma_0 = h_0/\bar{c}$ the corresponding hacker payoff-to-cost ratio, F_0 the firm's loss per attack, and $N_{h,0}$ the hacker attack intensity. For a firm with perceived cybersecurity level $\bar{z}$, consumer demand and hacker attack intensity satisfy

$$N_c = 1 - p - (1 - \bar{z}) l N_{h,0} \quad \text{and} \quad N_{h,0} = \gamma_0 N_c.$$

The corresponding profit functions are

$$\pi_H = pN_c - (1 - z)F_0 N_{h,0},$$
$$\pi_L = pN_c - F_0 N_{h,0} - \lambda \cdot \mathbf{1}\{r = r_H\}.$$

We now normalize the consumer loss per attack to one, i.e., $l = 1$. Define $h = lh_0$, $\gamma = h/\bar{c} = l\gamma_0$, and $F = F_0/l$. A hacker attacks whenever $c \le h_0 N_c$, so on the interior domain the participation share is $N_{h,0} = \gamma_0 N_c$, and we define the normalized attack intensity as the loss-weighted intensity $N_h \equiv l N_{h,0} = l\gamma_0 N_c = \gamma N_c$; it is not a probability or a participation share and can exceed one. Substituting $N_h = \gamma N_c$ into consumer demand $N_c = 1 - p - (1 - \bar{z})N_h = 1 - p - (1 - \bar{z})\gamma N_c$ yields

$$N_c = \frac{1 - p}{1 + \gamma(1 - \bar{z})} = \frac{1 - p}{1 + l\gamma_0(1 - \bar{z})}.$$

This is the same demand obtained before normalization. The firm's total loss from attacks is also unchanged, because $F N_h = (F_0/l)(l N_{h,0}) = F_0 N_{h,0}$. Hence the profit functions can equivalently be written as

$$\pi_H = pN_c - (1 - z)F N_h,$$
$$\pi_L = pN_c - F N_h - \lambda \cdot \mathbf{1}\{r = r_H\}.$$

These are exactly the demand and profit expressions used in the main model after normalizing $l = 1$.

One qualification concerns the domain. The interior solution of the original model requires that hacker participation be interior, $N_{h,0} = \gamma_0 N_c < 1$, at every price and belief, so that the continuation

after any claim–price pair, on or off the equilibrium path, is the one derived above. Since $N_c = (1-p)/[1+l\gamma_0(1-\beta z)]$ is largest at $p=0$ and $\beta=1$, this holds for all (p,β) if and only if

$$\gamma_0\,[1-l(1-z)]<1,$$

that is, $\gamma_0 < 1/[1-l(1-z)]$ when $l(1-z)<1$, and for every γ_0 when $l(1-z)\geq 1$. For $l=1$ this is the baseline condition $\gamma < 1/z$; in normalized units it reads $\gamma < l/[1-l(1-z)]$ when $l(1-z)<1$ and imposes no bound otherwise. The positive-demand condition $\gamma F<1$ is invariant under the transformation because $\gamma F = \gamma_0 F_0$. Therefore, for any fixed $l>0$, we can normalize l to one by redefining $h = lh_0$, $\gamma = l\gamma_0$, and $F = F_0/l$, provided that the admissible range of the normalized ratio is $0<\gamma<\min\{1/F, l/[1-l(1-z)]\}$ when $l(1-z)<1$ and $0<\gamma<1/F$ otherwise, rather than $\min\{1/F, 1/z\}$. On this domain the reduced-form demand, profit, and welfare expressions are preserved. Conclusions involving the participation bound must use the adjusted upper bound above; in particular, the cutoff $F^o(z)$ is specific to the baseline domain. For $0<\lambda<\bar{\lambda}_p$, the pooling region is truncated at this adjusted bound, and separation reemerges within the admissible range if and only if $\bar{\gamma}_2(\lambda)$ lies strictly below it. For $l=1$ the domain is the one stated in Section 3. Outside it the interior formulas overstate participation. For example, $l=0.1$, $z=0.1$, $\gamma_0=5$, and $F_0=0.01$ satisfy $\gamma_0<1/z$ and $\gamma_0 F_0<1$ but violate $\gamma_0[1-l(1-z)]<1$: at the high type's full-information price 0.523 the interior formulas give $N_c=0.329$ and $N_{h,0}=\gamma_0 N_c=1.65>1$, so the participation constraint binds and the equivalence fails. Conversely, with $l=10$, $\gamma_0=1$, $z=0.9$, and $F_0=0.1$, we have $l(1-z)=1$, so the condition imposes no bound: the high type's price is 0.505 with $N_{h,0}=N_c=0.248$, and the transformation gives $\gamma=10$ and $F=0.01$ with $N_h=2.48$, an admissible point at which the loss-weighted intensity exceeds one while participation is interior.

A.2. Congestion among Hackers

In the baseline model, we assume that each hacker earns h per active consumer upon a successful breach, independent of the number of attacking hackers. In practice, however, as more hackers breach the firm, the exploitable gain from each consumer may decrease because hackers compete for the same pool of valuable information or accounts. In what follows, we show that this congestion effect can be captured without altering our main results.

Let $\bar{N}_h(r,p)$ denote the aggregate attack intensity anticipated by hackers. For $\bar{N}_c(r,p)>0$, we assume that each hacker's gain per active consumer from a breach is

$$h-\eta\frac{\bar{N}_h(r,p)}{\bar{N}_c(r,p)} \tag{15}$$

where $\bar{N}_h/\bar{N}_c$ captures congestion through the ratio of hacker attack intensity to the firm's consumer base, and $\eta\geq 0$ measures the strength of the congestion effect. Accordingly, a successful breach

yields a lower per-consumer gain when more hackers attack (larger $\bar{N}_h$) or when the firm's consumer base is smaller (smaller $\bar{N}_c$).

Accordingly, a hacker's total payoff from attacking is adapted to

$$\begin{aligned} U_h(c \mid r,p) &= \left[h - \eta \frac{\bar{N}_h(r,p)}{\bar{N}_c(r,p)}\right] \bar{N}_c(r,p) - c \\ &= h\bar{N}_c(r,p) - \eta \bar{N}_h(r,p) - c. \end{aligned}$$

A hacker attacks if and only if $c \le h\bar{N}_c(r,p) - \eta\bar{N}_h(r,p)$. Since $c \sim U[0,\bar{c}]$, whenever the cutoff is interior, actual attack intensity satisfies

$$N_h(r,p) = \frac{h\bar{N}_c(r,p) - \eta\bar{N}_h(r,p)}{\bar{c}}.$$

In equilibrium, hackers' expectations are correct, so $\bar{N}_c(r,p) = N_c(r,p)$ and $\bar{N}_h(r,p) = N_h(r,p)$. It follows that $(\bar{c}+\eta)N_h(r,p) = hN_c(r,p)$, and hence

$$N_h(r,p) = \frac{h}{\bar{c}+\eta} N_c(r,p) = \gamma_\eta N_c(r,p), \text{ where } \gamma_\eta = \frac{h}{\bar{c}+\eta}.$$

This expression has exactly the same form as the baseline attack intensity $N_h(r,p) = \gamma N_c(r,p)$. Consequently, consumer demand becomes $N_c = (1-p)/[1+\gamma_\eta(1-\bar{z})]$, and the type-$i$ firm's payoff is

$$\pi_i = \frac{(1-p)[p - \gamma_\eta(1-z_i)F]}{1+\gamma_\eta(1-\bar{z})} - \lambda \cdot \mathbf{1}\{i = L, r = r_H\}.$$

These expressions coincide with those in the baseline model after replacing $\gamma = h/\bar{c}$ with $\gamma_\eta = h/(\bar{c}+\eta)$. Therefore, our analysis remains valid and our main results continue to hold when we incorporate the congestion effect by modifying each hacker's breach gain as in (15). Moreover, $\partial\gamma_\eta/\partial\eta = -h/(\bar{c}+\eta)^2 < 0$ implies that stronger congestion is equivalent to a lower effective hacker payoff-to-cost ratio.

A.3. Robustness to a Positive Low-Type Cybersecurity Level

We now allow $0 < z_L < z_H < 1$. For a posterior belief β that the firm is the high type, perceived cybersecurity is $\bar{z} = z_L + \beta(z_H - z_L)$. Define $\tilde{z} = \frac{z_H - z_L}{1-z_L}$, $\tilde{\gamma} = \gamma(1-z_L)$, $\tilde{z}_L = 0$, and $\tilde{z}_H = \tilde{z}$. Then $0 < \tilde{z} < 1$, $1-\bar{z} = (1-z_L)(1-\beta\tilde{z})$, and $1 - z_i = (1-z_L)(1-\tilde{z}_i)$ for $i \in \{H, L\}$. In the interior hacker-participation region, rational expectations therefore give $N_c(p \mid \beta) = \frac{1-p}{1+\tilde{\gamma}(1-\beta\tilde{z})}$ and

$$\pi_i(r, p \mid \beta) = \frac{(1-p)[p - \tilde{\gamma}F(1-\tilde{z}_i)]}{1+\tilde{\gamma}(1-\beta\tilde{z})} - \lambda \cdot \mathbf{1}\{i = L, r = r_H\}.$$

These are exactly the baseline demand and firm-profit expressions evaluated at $(\tilde{z}, \tilde{\gamma})$. Consequently, the pooling and separating incentive constraints, the LMSE payoff comparison, and the resulting prices, demands, consumer surplus, and social welfare follow from the baseline analysis under this

parameter transformation. Thus, setting $z_L = 0$ does not alter the model's underlying equilibrium mechanisms: cases with $z_L > 0$ can be mapped to the baseline model through an appropriate parameter transformation. For comparisons across values of z_L below, we impose the common sufficient restriction $0 < \gamma < \min\{1/F, 1/z_H\}$, which ensures positive demand and interior hacker participation for $0 < z_L < z_H < 1$.

To examine the equilibrium regions, we fix $F = 0.3$, $z_H = 0.4$, and $\alpha = 1/2$, and compare $z_L = 0$, 0.1, and 0.2 over the same (γ, λ) range. Figure A1 plots the pooling-to-separation boundary λ_p and the costly-to-costless separation boundary λ_s for each value of z_L. All three cases show pooling arises at low penalties for an intermediate range of hacker payoff-to-cost ratio, while higher penalties or sufficiently low or high hacker incentives sustain separation. The locations of the boundaries change with z_L, but this qualitative equilibrium pattern remains.

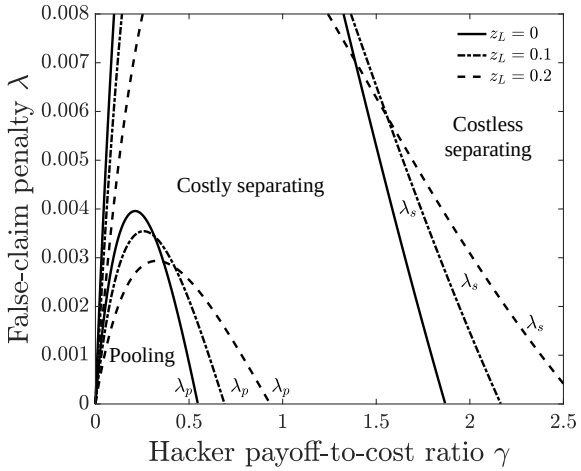
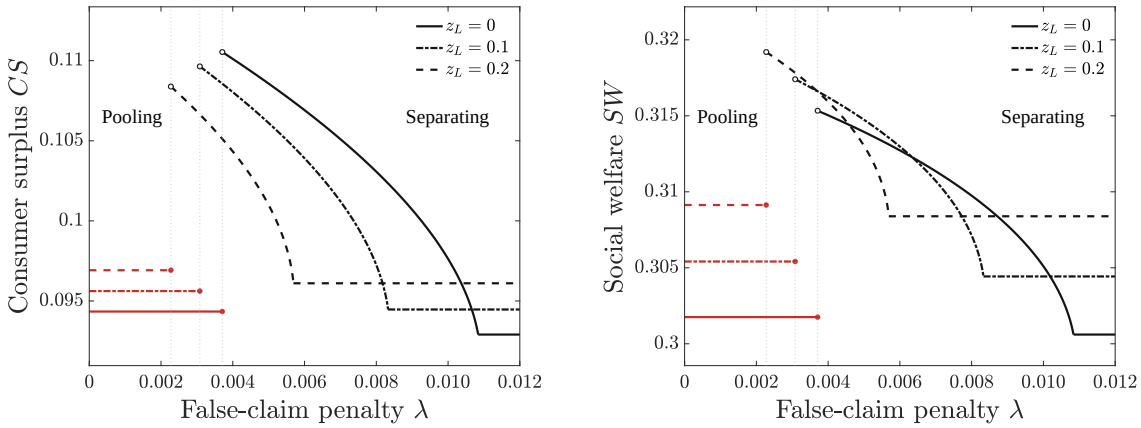


Figure A1 LMSE equilibrium outcomes with positive low-type cybersecurity level (Parameters: $z_H = 0.4$, $F = 0.3$).

We next fix $\gamma = 0.15$ and examine how consumer surplus and social welfare vary with the false-claim penalty. Figure A2 shows that, for each value of z_L, consumer surplus and social welfare have suprema approached as the penalty decreases toward the deterrence threshold from above; further increases within costly separation reduce both, while they remain constant after costless separation is reached. Thus, a positive low-type cybersecurity level changes the transition points in these examples without eliminating the non-monotone welfare effect of the penalty.



(a) Consumer surplus (b) Social welfare

Figure A2 Effect of the false-claim penalty on consumer surplus and social welfare with positive low-type cybersecurity level (Parameters: $z_H = 0.4$, $F = 0.3$, $\gamma = 0.15$).

A.4. Demand and Attack Intensity in Each Equilibrium

This section records the customer base and the attack intensity in the equilibria of Section 4; they follow from substituting the equilibrium prices into (6), and the proofs are in Online Supplement B. Under symmetric information (Lemma 1), the firm's consumer demand and attack intensity are $N^m_{c,i} = \frac{1-\gamma F(1-z_i)}{2[1+\gamma(1-z_i)]}$ and $N^m_{h,i} = \frac{\gamma[1-\gamma F(1-z_i)]}{2[1+\gamma(1-z_i)]}$ for $i \in \{H, L\}$, so $N^m_{c,H} > N^m_{c,L}$ and $N^m_{h,H} > N^m_{h,L}$: the secure firm's lower price gives it a larger customer base and, through the demand channel, more attacks.

COROLLARY A1. *(i) In a costly separating equilibrium,* $N^{sep}_{c,H} = \frac{1-\gamma F+S(\lambda)}{2[1+\gamma(1-z)]} > N^m_{c,H}$ *and* $N^{sep}_{h,H} = \frac{\gamma[1-\gamma F+S(\lambda)]}{2[1+\gamma(1-z)]} > N^m_{h,H}$. *For the low type,* $N^{sep}_{c,L} = N^m_{c,L}$ *and* $N^{sep}_{h,L} = N^m_{h,L}$.
(ii) In a costless separating equilibrium, $N^{sep}_{c,i} = N^m_{c,i}$ *and* $N^{sep}_{h,i} = N^m_{h,i}$. *In either separating equilibrium,* $N^{sep}_{c,H} > N^{sep}_{c,L}$ *and* $N^{sep}_{h,H} > N^{sep}_{h,L}$.

COROLLARY A2. *Under the LMSE pooling equilibrium,* $N^p_{c,H} = N^p_{c,L} = \frac{1-\gamma F(1-z)}{2+\gamma(2-z)}$ *and* $N^p_{h,H} = N^p_{h,L} = \frac{\gamma[1-\gamma F(1-z)]}{2+\gamma(2-z)}$. *Consequently,* $N^m_{c,L} < N^p_{c,H} = N^p_{c,L} < N^m_{c,H} \le N^{sep}_{c,H}$ *and* $N^m_{h,L} < N^p_{h,H} = N^p_{h,L} < N^m_{h,H} \le N^{sep}_{h,H}$.

Under separation, the high type serves at least as many consumers as under symmetric information, and strictly more when it lowers its price to signal its type, so it attracts weakly more hackers. Pooling involving cyberwashing compresses the type-specific outcomes into a common intermediate allocation: relative to symmetric information, it expands the low type's customer base and attack intensity, whereas the high type's security advantage is not recognized, reducing both.

A.5. A Benchmark with Exogenous Attack Intensity

This section provides the benchmark discussed in Section 4.1 and after Lemma 2. To isolate what the endogenous attack channel adds, suppose that the conditional attack intensity is an exogenous

constant $\bar{N}_h \in (0,1)$, independent of the firm's claim, price, and customer base; attacks still succeed only against a vulnerable system. Demand is then $q^e(p,\beta) = [1-p-(1-\beta z)\bar{N}_h]^+$ (the positive part matters only for off-path prices above $1-(1-\beta z)\bar{N}_h$, at which gross revenue is zero), profit is $\pi_i^e(r,p \mid \beta) = p\,q^e(p,\beta) - (1-z_i)\bar{N}_h F - \lambda \mathbf{1}\{i=L,\ r=r_H\}$, and $V(\beta) = [1-(1-\beta z)\bar{N}_h]^2/4$ denotes the maximized gross revenue under belief β. The proof of the following proposition is in Online Supplement B.1.

Proposition A1 **(Exogenous attack intensity)**. *With exogenous attack intensity: (i) $\pi_H^e(r,p \mid \beta) - \pi_L^e(r,p \mid \beta) = z\bar{N}_h F + \lambda \mathbf{1}\{r=r_H\}$ does not depend on p, so price signaling is money burning: in a costly separating equilibrium the high type is indifferent between an upward and a downward price distortion, and its informational rent relative to the low type's reservation payoff equals $\lambda + z\bar{N}_h F$, independently of its customer base. (ii) The LMSE outcome is pooling if $\lambda < V(\alpha) - V(0)$, costly separating if $V(\alpha) - V(0) < \lambda < V(1) - V(0)$, and costless separating if $\lambda \geq V(1) - V(0)$; at $\lambda = V(\alpha) - V(0)$ both types are indifferent between the pooling outcome and the best separating outcome, so both are LMSE outcomes, and we assign the boundary to pooling by convention. Neither threshold depends on F. (iii) Within the costly separating regime, consumer surplus and social welfare rise or fall together as the penalty changes.*

Two features of the main model survive in the benchmark and are therefore not attributable to the attack channel: the three regimes in the penalty, and pooling for intermediate threat levels, since the pooling threshold $V(\alpha) - V(0) = \alpha z \bar{N}_h[2-(2-\alpha z)\bar{N}_h]/4$ is inverted-U shaped in $\bar{N}_h$ because a stronger threat shrinks the market for both types. What does not survive is the type dependence of the price instrument. With exogenous attacks, a price distortion costs the secure firm exactly what it costs the vulnerable firm (part (i)), so the direction of the distortion is indeterminate, the secure firm's rent under costly separation is independent of its customer base, the firm's breach loss F plays no role in the regime boundaries (part (ii)), and the effect of deterrence on consumer surplus has no determinate sign, because the LMSE is silent on which of the two equally profitable distorted prices is played. With endogenous attacks, the single-crossing property (8) pins the distortion downward, gives the secure firm a rent that grows with its customer base (Lemma 2), makes F a determinant of the regime boundaries (Propositions 1–4), gives the consumer-surplus jump at the deterrence threshold a determinate sign, and creates a divergence, within the separating regime, between the consumer-surplus-maximizing and welfare-maximizing penalties (Proposition 8; compare part (iii)), because demand expansion now raises attack exposure. Table A1 summarizes, feature by feature, what the endogenous attack channel adds relative to the benchmark. Both rent expressions in the table are measured relative to the low type's reservation payoff.

Table A1 What the endogenous attack channel adds: Baseline model versus the exogenous-attack benchmark of Proposition A1.

Feature	Exogenous attack intensity	Endogenous attack intensity (this paper)
Equilibrium regimes in the penalty	Pooling, costly and costless separation	Same
Type dependence of price-related payoffs	None: payoff differences across types are constant in price	Effective marginal cost $\gamma(1-z_i)F$ is lower for the secure type (single crossing)
Price signaling	Money burning; upward and downward distortions are equally costly and equally effective; rent $\lambda + z\bar{N}_h F$	Downward only; the secure firm's rent is $\lambda + z\gamma F N_{c,H}$, increasing in its customer base
Price–security relation under full information	Secure firm charges more	Secure firm charges less (Lemma 1)
Role of the firm's breach loss F in regime boundaries	None	Central $(\lambda_p, \lambda_s, F^o)$
Pooling for intermediate threat levels	Yes, because a stronger threat shrinks the market for both types	Yes, because demand expansion raises attack exposure; the second separation may lie outside the admissible range when F is small
Consumer surplus at the deterrence threshold	Indeterminate (depends on the direction of the distortion)	Jumps up (Proposition 8)
Welfare- versus consumer-optimal penalty	Within costly separation, consumer surplus and welfare rise or fall together as the penalty changes (Proposition A1(iii))	Diverge when $\gamma > 1/\sqrt{2z(1-z)}$ (Proposition 8)
Attack exposure response to disclosure and pricing	None	Secure firm's exposure rises with signaling (Proposition 6)

A.6. Equilibrium Refinements

The following proposition records the properties of the LMSE outcome that are discussed in Sections 3.5 and 4.2.3; the formal definition of an undefeated equilibrium (Mailath et al., 1993) is given with the proof in Online Supplement B.1.

PROPOSITION A2 **(Refinements)**. *(i) For every admissible (z, F, γ) and every $0 < \lambda \le \lambda_p(\gamma)$, the pooling outcome selected by the LMSE fails the Intuitive Criterion and D1; under either refinement the equilibrium outcome is separating for every $\lambda > 0$. (ii) Within the maintained class of pure-strategy equilibria, the LMSE outcome of Proposition 3 is undefeated. (iii) No semi-separating equilibrium, in which the low type randomizes between truthful disclosure and the high type's claim–price pair, gives the high type a higher profit than the LMSE outcome, so enlarging the equilibrium class to include such equilibria does not change the selected outcome.*

Part (i) is the price of the pessimistic off-path beliefs that sustain pooling: a price cut that is attractive only to the secure firm is exactly the deviation that forward-induction refinements use to break pooling. Part (ii) is the foundation we rely on instead: in the pooling region the high type

weakly prefers pooling to its best separating outcome and strictly prefers the selected pooling price to any other, so no equilibrium defeats it. Part (iii) follows from Lemma 2, because the posterior attached to the high type's pair in a semi-separating equilibrium is below one, so its customer base, and hence its rent, is smaller than under pure separation.

Table A2 makes explicit which conclusions of the baseline model depend on the LMSE selection and which survive under the Intuitive Criterion and D1. Because the high-type-optimal separating equilibrium of Proposition 1 is the outcome selected by both forward-induction criteria for every $\lambda > 0$ (proof of part (i)), everything that happens within the separating regime is common to the three criteria: the downward price distortion, its decrease in λ, and the resulting attack-intensity and profit patterns (Propositions 5–7), the costly-to-costless boundary $\lambda_s(\gamma)$, and the fact that the separating welfare increases in λ exactly while the secure firm's customer base exceeds the efficient level q_H^W (Online Supplement A.9). What depends on the selection is the existence of a pooling regime at low penalties and everything built on the pooling-to-separation transition: the deterrence threshold $\lambda_p(\gamma)$, the consumer-surplus jump at λ_p^+, the governance-induced switch into cyberwashing as γ falls (Proposition 9 and Section 5), and the location of the consumer-optimal penalty. Under the Intuitive Criterion or D1 the penalty has no deterrence role. If $\lambda_s > 0$, the consumer-surplus supremum is approached as $\lambda \downarrow 0$; welfare is maximized at λ^o if $\lambda^o > 0$, and otherwise has its supremum in the same limit. If $\lambda_s \leq 0$, both measures are constant for all $\lambda > 0$.

Table A2 Conclusions of the baseline model under the LMSE selection and under the Intuitive Criterion or D1.

Conclusion	LMSE	Intuitive Criterion / D1
Pooling with false claims (cyberwashing)	Selected for $0 < \lambda \leq \lambda_p(\gamma)$ (Propositions 3–4)	Eliminated: the outcome is separating for every $\lambda > 0$ (Proposition A2(i))
Enforcement regimes in λ	Three: pooling, costly separating, costless separating	Up to two: costly separating for $\lambda < \lambda_s(\gamma)$ and costless for $\lambda \geq \lambda_s(\gamma)$, with the same boundary λ_s
Governance-induced cyberwashing trap (switch into pooling as γ falls to intermediate levels; Proposition 9)	Present, through the regime switch at $\bar{\gamma}_2(\lambda)$	Absent; only the dependence of the separating price distortion on γ remains
Costly price signaling and its effects within separation (Propositions 5–7)	Present for $\lambda_p < \lambda < \lambda_s$	Present for $0 < \lambda < \lambda_s$: the same separating outcome is selected
Consumer-surplus optimum	Unattained supremum as $\lambda \downarrow \lambda_p$ when $\lambda_p > 0$ (Proposition 8(i))	Unattained supremum as $\lambda \downarrow 0$ if $\lambda_s > 0$; otherwise any $\lambda > 0$
Welfare optimum	For $\lambda_p > 0$: supremum as $\lambda \downarrow \lambda_p$ if $\gamma \leq 1/\sqrt{2z(1-z)}$; otherwise unique maximum at λ^o (Proposition 8(ii))	If $\lambda_s > 0$, λ^o when positive, otherwise supremum as $\lambda \downarrow 0$; if $\lambda_s \leq 0$, any $\lambda > 0$

A.7. Costly Pre-Discovery Effort: Additional Results

This section complements Section 6.1. As in the baseline, every candidate equilibrium is supported by the off-path belief $\beta = 0$, and the proof of Proposition 10 verifies the high type's incentive constraint against its best deviation under that belief. Unlike in the baseline, this constraint can fail in the costly separating candidate when λ is small and z is large, because the separating price can then be so low that the high type would rather be taken for the low type at its optimal price under $\beta = 0$. If this happens at a penalty for which a pooling equilibrium exists, pooling is the LMSE outcome and the characterization is unaffected; if it happens at a penalty above the largest one that sustains pooling, no pure-strategy equilibrium exists in the maintained class outside the sufficient condition $F \leq 4(1-z)$ of Proposition 10(i). The second case requires large z and F and does not arise for any parameter combination used in this section or in Online Supplement A.8.

Prices and attack intensity. Because hackers now make a strategic discovery-and-attack decision, the expressions for equilibrium attack intensity and welfare involve higher-order terms in the hacker payoff-to-cost ratio γ, which makes a complete analytical welfare characterization less tractable. We therefore conduct a numerical study. We consider two settings: $z = 0.4$, $F = 0.05$, and $\gamma = 2$, and the same parameters with the high-type firm's cybersecurity level raised to $z = 0.85$. Figure A3(a) shows that, when $z = 0.4$, the effects of the false-claim penalty on the equilibrium price and firm attack intensity are the same as in the baseline model: when λ is just high enough for the equilibrium to switch from pooling to separating, the high-type firm distorts its price downward, which attracts more consumers but also leads to greater attack intensity, and this distortion and its effects diminish as λ increases. However, when $z = 0.85$, the attack intensity associated with the high-type firm decreases when the equilibrium transitions from pooling to separating, in contrast to the baseline model, where Proposition 6 holds for all parameters. The reason is the inference channel: identifying the firm as the high type in the separating equilibrium discourages hacker effort, and when z is sufficiently high this deterrence effect dominates the attack-attraction effect generated by the high type's larger consumer base.

Consumer surplus and welfare. Figure A4 plots consumer surplus and social welfare under costly pre-discovery effort and compares them with the baseline model under the same parameter values. In the reported penalty-grid comparisons, consumer surplus, and in these examples social welfare, attain their largest evaluated values at intermediate penalties, and excessively stringent regulation reduces them because it changes the firm's equilibrium pricing response; Online Supplement A.8 reports the corresponding grid maxima for the tested values of the timing parameter introduced below. Within the same equilibrium regime, the costly pre-discovery timing yields higher consumer surplus and welfare, because hackers who must pay before learning whether a vulnerability exists attack less. For penalties between λ_p and λ_p^d, costly pre-discovery changes the equilibrium from

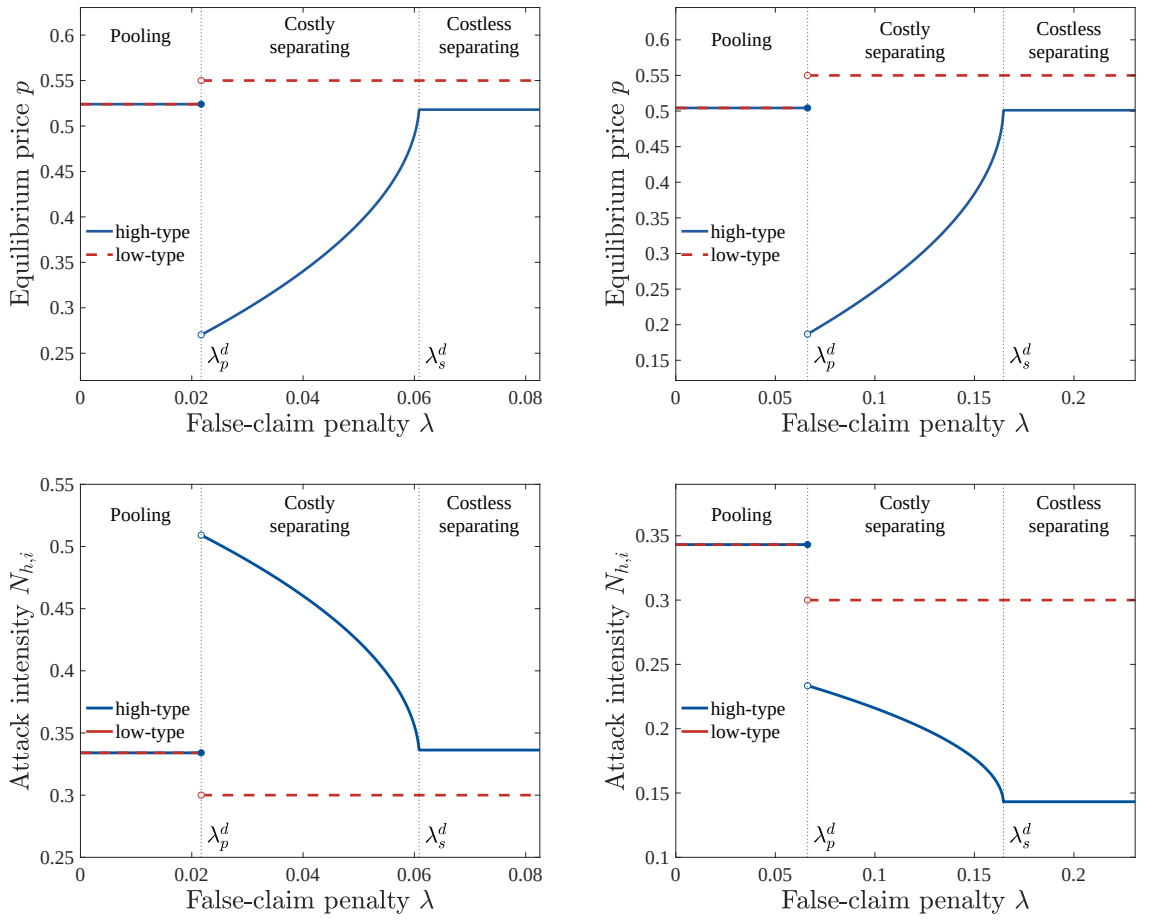


Figure A3 **Effect of false-claim penalty λ on equilibrium price and (unconditional) attack intensity under costly pre-discovery effort (Parameters: $F = 0.05$, $\gamma = 2$).**

separating to pooling. Both consumer surplus and social welfare are lower than under the baseline over only part of this interval. Their net changes also reflect the benefit from reduced attack exposure.

A discoverability parameter. The comparison above is between two timings and does not, by itself, deliver comparative statics in the difficulty of discovering vulnerabilities. To make such statements, suppose that a fraction $\delta \in [0,1]$ of hackers observe at no cost whether an exploitable vulnerability exists before deciding whether to incur c (the baseline timing), whereas the remaining fraction $1-\delta$ must incur c before learning this (the timing of Section 6.1). A higher δ thus corresponds to more easily discoverable vulnerabilities. Conditional on a vulnerability, the two groups participate with shares $\min\{\gamma N_c, 1\}$ and $\min\{\gamma(1-\beta z)N_c, 1\}$, so total participation is $\gamma N_c[\delta + (1-\delta)(1-\beta z)]$ whenever both cutoffs are interior; consumers' expected exposure is $(1-\beta z)$ times this amount, and the firm's expected loss is $(1-z_i)F$ times this amount; $\delta = 1$ recovers the baseline and $\delta = 0$ recovers the model of Section 6.1. The reduced form then has the same structure as (6)–(7) with

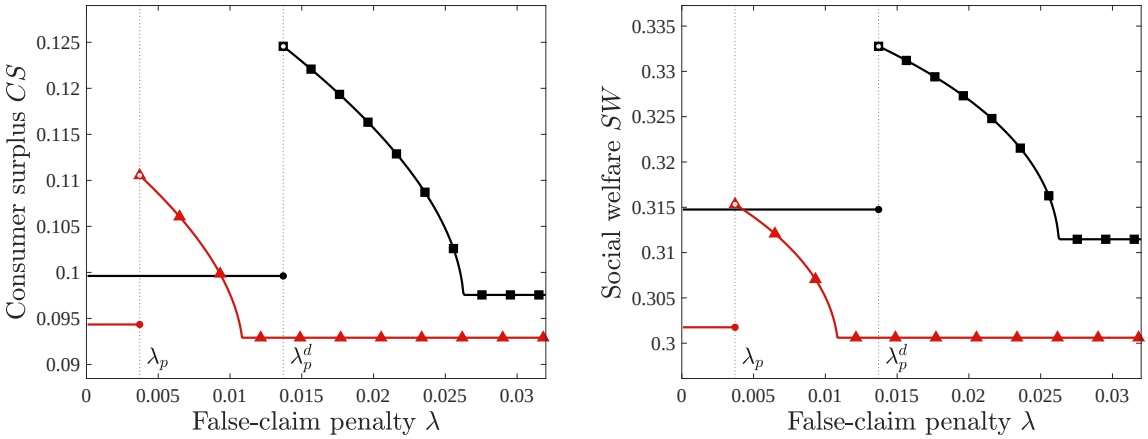
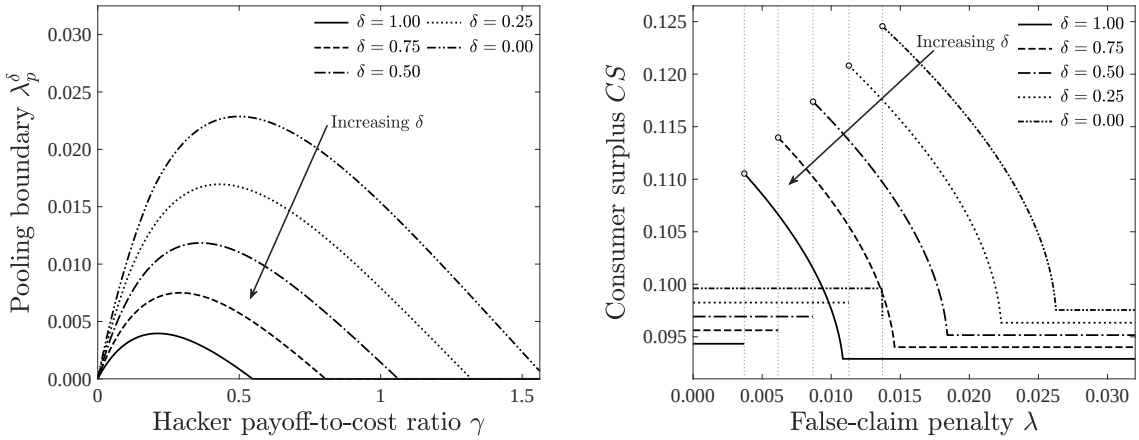


Figure A4 **Comparison of CS and SW under free and costly pre-discovery effort (Parameters: $z = 0.4$, $F = 0.3$, $\gamma = 0.15$). Red lines with triangle markers represent the baseline model; black lines with square markers represent costly pre-discovery.**



(a) Pooling boundary $\lambda_p^\delta(\gamma)$ (b) Consumer surplus against λ at $\gamma = 0.15$

Figure A5 **Discoverability family: a fraction δ of hackers observe vulnerabilities before incurring the attack cost (Parameters: $z = 0.4$, $F = 0.3$). Panel (a) spans the domain $\gamma < \min\{1/F, 1/[z(2-z)], 4\} \approx 1.56$ on which the interior participation condition holds for every δ. Open circles in panel (b) mark the separating-side limits at the numerically located pooling boundaries for each δ.**

γ replaced by a belief-dependent multiplier, and the LMSE can be computed exactly for each δ. Because each group must be interior at every price and belief, including off-path pairs, the linear form requires its own participation bound: for $\delta > 0$ the informed group binds, and since $N_c = (1-p)/\{1+\gamma(1-\beta z)[\delta+(1-\delta)(1-\beta z)]\}$ is largest at $p = 0$ and $\beta = 1$, the condition is $\gamma < \bar{\gamma}(\delta) \equiv 1/\{z[2-z-\delta(1-z)]\}$, which equals $1/z$ at $\delta = 1$ and decreases to $1/[z(2-z)]$ as $\delta \downarrow 0$; for $\delta = 0$ only the uninformed group is present and the condition is $\gamma(1-\beta z)N_c < 1$, for which $\gamma < 4$ suffices (Section 6.1). Together with $\gamma F < 1$, the timings can therefore be compared on the common domain $\gamma < \min\{1/F, 1/[z(2-z)], 4\}$, on which the interior continuation is valid for every $\delta \in [0,1]$; for $z = 0.4$ and $F = 0.3$ this is $\gamma < 1.56$, and Figure A5(a) is drawn on it. Figure A5 shows the result

for the parameters of Figure A4. The pooling boundary $\lambda_p^{\delta}(\gamma)$ shifts up monotonically as δ falls, so lower discoverability enlarges the cyberwashing region for every tested γ; and consumer surplus as a function of λ shifts up within each regime as δ falls, while the penalty required to leave the pooling regime rises. Interventions that make vulnerability discovery more difficult therefore raise consumer surplus and welfare within each equilibrium regime in the reported comparisons, but they also raise the false-claim penalty required to deter cyberwashing, and for penalties between λ_p^1 and λ_p^{δ} they move the market into pooling and can lower both for some of these penalties.

A.8. Documented Numerical Robustness

Several statements in Sections 6.1 and 6.2 of the main text are numerical rather than analytical. To make clear which claims hold where, we report the grids on which they were checked and the procedure used.

Table A3 Consumer-surplus and welfare suprema and pooling-to-separation jumps on the parameter grids.

Demand-dependent breach loss (Section 6.2)

Case	n	Location of $\sup_\lambda CS$			Location of $\sup_\lambda SW$			CS	SW
		λ_p^+	interior	$\geq \lambda_s$	λ_p^+	interior	$\geq \lambda_s$	jump up	jump up
$\theta = 0.0$	135	1.00	0.00	0.00	0.74	0.26	0.00	1.00	0.96
$\theta = 0.2$	77	1.00	0.00	0.00	0.77	0.23	0.00	1.00	0.94
$\theta = 0.5$	36	1.00	0.00	0.00	0.78	0.22	0.00	1.00	0.89
$\theta = 0.8$	11	1.00	0.00	0.00	0.82	0.18	0.00	1.00	1.00
$\theta > 0$, $F \geq z$	24	1.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00
$\theta > 0$, $F < z$	100	1.00	0.00	0.00	0.72	0.28	0.00	1.00	0.91

Discoverability family (Section 6.1)

Case	n	Location of $\sup_\lambda CS$			Location of $\sup_\lambda SW$			CS	SW
		λ_p^+	interior	$\geq \lambda_s$	λ_p^+	interior	$\geq \lambda_s$	jump up	jump up
Each δ on its own interior domain $\gamma < \min\{1/F, \bar{\gamma}(\delta)\}$; $\delta = 1$ is the baseline									
$\delta = 0.00$	199	1.00	0.00	0.00	0.65	0.35	0.00	1.00	0.99
$\delta = 0.25$	212	1.00	0.00	0.00	0.79	0.21	0.00	1.00	0.96
$\delta = 0.50$	190	1.00	0.00	0.00	0.80	0.20	0.00	1.00	0.96
$\delta = 0.75$	164	1.00	0.00	0.00	0.77	0.23	0.00	1.00	0.96
$\delta = 1.00$	135	1.00	0.00	0.00	0.74	0.26	0.00	1.00	0.96
All δ on the common domain $\gamma < \min\{1/F, 1/[z(2-z)], 4\}$									
$\delta = 0.00$	223	1.00	0.00	0.00	0.82	0.18	0.00	1.00	1.00
$\delta = 0.25$	216	1.00	0.00	0.00	0.82	0.18	0.00	1.00	1.00
$\delta = 0.50$	200	1.00	0.00	0.00	0.85	0.15	0.00	1.00	1.00
$\delta = 0.75$	179	1.00	0.00	0.00	0.87	0.13	0.00	1.00	1.00
$\delta = 1.00$	158	1.00	0.00	0.00	0.89	0.11	0.00	1.00	1.00

Notes. Suprema are taken over the 160-point penalty grids. Shares are based on the n cases with a pooling region and rounded to two decimals; 1.00 need not mean 100%. Maxima are classified as "λ_p^+" at the first separating grid point, "interior" at a later point below λ_s, and "$\geq \lambda_s$" in costless separation. Here λ_p^+ is a grid label, not an exact right-hand limit; "jump up" compares the first separating and last pooling grid values.

Grid. Each grid uses $z \in \{0.1, 0.25, 0.4, 0.6, 0.85\}$, $F \in \{0.02, 0.05, 0.1, 0.2, 0.3, 0.5\}$, and eight values of γ, $\gamma_k = f_k \bar{\gamma}(z, F)$ with $f_k = 0.08 + 0.89(k-1)/7$ for $k = 1, \ldots, 8$, where $\bar{\gamma}(z, F)$ is the upper bound of the domain on which the interior continuation is valid at every price and belief: $\min\{1/F, 1/z\}$ for the baseline and for the demand-dependent breach-loss model, whose attack technology is the baseline's; for the discoverability family, either each δ's own bound $\min\{1/F, \bar{\gamma}(\delta)\}$ of Online Supplement A.7, with $\bar{\gamma}(1) = 1/z$ and $\bar{\gamma}(0) = 4$, so that the $\delta = 1$ row is the baseline on the baseline domain, or the common bound $\min\{1/F, 1/[z(2-z)], 4\}$, on which all values of δ are compared on the same parameter combinations; Table A3 reports both. The additional grid values are $\theta \in \{0, 0.2, 0.5, 0.8\}$ and $\delta \in \{0, 0.25, 0.5, 0.75, 1\}$.

Procedure. A parameter combination enters a row only if pooling is the LMSE outcome for some penalty ($\lambda_p > 0$); the others are excluded, and n is the number retained. For each retained combination we evaluate the LMSE outcome at 160 equally spaced penalties, $\lambda_j = 10^{-8} + (1.3\lambda_s - 10^{-8})(j-1)/159$, $j = 1, \ldots, 160$, where λ_s denotes the corresponding model's costless-separation threshold, report pooling whenever both pooling and separation are LMSE at a cutoff (a numerical reporting convention), record the first grid penalty at which separation is selected, and locate the maxima of consumer surplus and social welfare over this finite grid, resolving numerical ties toward the smaller penalty: "λ_p^+" denotes the first grid penalty at which separation is selected, "interior" a later grid penalty below λ_s, and "$\geq \lambda_s$" a grid penalty at or beyond λ_s. The grid-rise columns compare the first separating grid penalty with the last pooling grid penalty; these finite differences are not exact discontinuities at the continuous selection boundary. Replication uses absolute tolerances of 10^{-12} for the high type's separating payoff constraint and 10^{-14} for overlap of the pooling demand intervals. Pooling boundaries are located by 90 bisection steps and retained if at least 10^{-13}. Objective values within 10^{-15} of the largest evaluated value are treated as tied and assigned to the smallest grid penalty. By construction every grid point satisfies the interior participation bound, so hacker participation is below one at every price and belief, on and off the equilibrium path, and the high type's incentive constraint is checked at every selected separating outcome using the stated tolerance. The two implementations of the baseline, $\theta = 0$ and $\delta = 1$ on the baseline domain, use the same grid and procedure and produce identical rows, as they should.

Table A3 reports the results for the demand-dependent breach-loss model (rows indexed by θ, with $\theta = 0$ the baseline) and for the discoverability family of Online Supplement A.7 (rows indexed by δ, with $\delta = 1$ the baseline and $\delta = 0$ the costly pre-discovery model). Three regularities hold on every reported grid: consumer surplus has its grid maximum at the first separating grid penalty; welfare has its grid maximum either there or at a later grid penalty below λ_s, never in the costless separating regime; and consumer surplus rises between the paired grid points, whereas welfare falls in a minority of these comparisons, all of which feature small F and large γ, in line with

Proposition 8(ii) (on the common domain, which excludes the largest values of γ, welfare rises in almost all reported comparisons, but a few downward changes remain). The restriction $F \geq z$ used for the analytical existence results of Section 6.2 does not drive the welfare conclusions: the two rows that split the $\theta > 0$ cases by whether $F \geq z$ show the same pattern. For the baseline rows ($\theta = 0$ and $\delta = 1$), the pattern is also what Proposition 8 proves. Finally, the high type's incentive constraint against its best deviation under the pessimistic belief passes the stated numerical feasibility check at every selected separating outcome on every reported grid; its only violations occur in costly separating candidates for the discoverability family with $\delta \leq 0.5$ at penalties below $0.64\lambda_p^\delta$, where pooling exists and is selected, so no gap in existence within the maintained pure-strategy class is found on the reported grids. For the costly pre-discovery model, a gap can arise outside the sufficient condition $F \leq 4(1-z)$ of Proposition 10(i); at $(z, F, \gamma) = (0.9, 2, 0.4)$, it contains $(0.063, 0.086)$.

A.9. Welfare Objective and Type-Wise Decomposition

Consumer surplus is $CS = \mathbb{E}_i[N_{c,i}^2/2]$, the expected surplus of consumers net of their expected breach losses, where the expectation is over the firm's type with the prior $\alpha = 1/2$. Social welfare is

$$SW = CS + \mathbb{E}_i[\pi_i] + R, \tag{16}$$

where π_i is the firm's expected profit, net of its own breach losses and of the penalty, and $R = (1-\alpha)\lambda \cdot \mathbf{1}\{\text{pooling}\}$ is the regulator's expected penalty revenue. The objective counts the surplus of consumers, the firm, and the government, treats the monetary fine as a transfer, and excludes hackers' surplus, the costs of running the enforcement technology, and the costs of any governance measure that changes γ. The exclusion of hackers' surplus is deliberate, since the gains from compromising consumer data are not legitimate surplus, but it means that the objective is a legitimate-stakeholder objective rather than a Kaldor–Hicks one. The exclusion of enforcement and governance costs means that comparisons across values of λ and γ describe the market response to those parameters, not the net benefit of the interventions that change them; in particular, equal reductions in $\gamma = h/\bar{c}$ achieved by raising attack costs (monitoring, sanctions) or by lowering the value of stolen data (credential invalidation, breach response) have the same market effects but need not have the same social costs. If the penalty is instead interpreted as a reputational loss, R is dropped from (16); all statements in Section 5 about the separating regimes are unaffected, and within the pooling regime social welfare then decreases in λ instead of being constant.

For a firm of known type i serving a customer base q, price and demand are linked by $p = 1 - q[1 + \gamma(1 - z_i)]$, and the sum of consumer surplus and the firm's profit is

$$W_i(q) = \big[1 - \gamma(1 - z_i)F\big]q - \Big[\tfrac{1}{2} + \gamma(1 - z_i)\Big]q^2. \tag{17}$$

The first term is the gross value of serving q customers net of the firm's breach losses, which scale with q because attacks scale with demand; the second term collects the foregone valuations of the marginal consumers and consumers' own breach losses, which grow with the square of the customer base. W_i is maximized at $q_i^W = [1-\gamma(1-z_i)F]/[1+2\gamma(1-z_i)]$, which exceeds the symmetric-information monopoly demand $N_{c,i}^m = [1-\gamma(1-z_i)F]/[2(1+\gamma(1-z_i))]$. Demand expansion by the high type is therefore welfare improving up to q_H^W and welfare reducing beyond it. Just above λ_p, the high type's signaling discount expands its customer base to $N_{c,H}(\lambda_p^+) = \frac{1}{2}[1-\gamma F(1-z)](1+\rho)/a$, where $a = 1+\gamma(1-z)$ and $\rho = \sqrt{z\gamma/(2+(2-z)\gamma)}$, and this exceeds q_H^W exactly when $\rho(2a-1) > 1$, that is, when $\gamma > 1/\sqrt{2z(1-z)}$. In that case the interior welfare maximum in Proposition 8(ii) is attained at the penalty at which the high type's customer base equals q_H^W,

$$\lambda^o = \frac{1}{4a}\left[\frac{z\gamma(1-\gamma F)^2}{1+\gamma} - \left(\frac{1+\gamma F(2za-1)}{2a-1}\right)^2\right]. \tag{18}$$

A.10. General Prior Belief: Additional Results

This section provides the results on a general prior belief $\alpha \in (0,1)$ that are summarized in Section 6.3; the proofs are in Online Supplement B.1.

Proposition A3. *(i) The costly-to-costless separating boundary λ_s is independent of α, whereas the pooling-separating boundary λ_p increases with α. (ii) For $0 < \lambda < \lambda_s$, there exists a unique $\alpha_p(\lambda) \in (0,1)$ such that the LMSE is costly separating for $\alpha < \alpha_p(\lambda)$ and pooling for $\alpha \geq \alpha_p(\lambda)$.*

Under separation, the firm's type is fully revealed, so λ_s is unaffected by the prior belief. A higher α makes consumers more optimistic about the firm's cybersecurity under pooling, increasing the high type's pooling payoff and allowing pooling to be sustained under a higher false-claim penalty. This creates a unique cutoff $\alpha_p(\lambda)$ at which the high type's preferred equilibrium switches from costly separation to pooling. As illustrated in Figure A6, we numerically characterize the LMSE outcome for $\alpha = 0.3$, 0.5, and 0.8. The qualitative equilibrium structure remains unchanged. As α increases, the pooling region expands, indicating that a higher prior probability of the firm being the high-security type makes cyberwashing more likely to arise in equilibrium.

Next, focusing on $0 < \lambda < \lambda_s$, we examine how α affects firm profits and consumer surplus.

Proposition A4. *(i) Both types' profits remain constant for $\alpha < \alpha_p$ and increase for $\alpha > \alpha_p$. At $\alpha = \alpha_p$, the high-type firm's profit is continuous, whereas the low-type firm's profit increases discontinuously.*

(ii) Consumer surplus increases for $0 < \alpha < \alpha_p$ and for $\alpha > \alpha_p$, but decreases discontinuously at $\alpha = \alpha_p$. Moreover, $\lim_{\alpha\uparrow 1} CS^p(\alpha) < \lim_{\alpha\uparrow\alpha_p} CS^s(\alpha)$ if and only if $\lambda_\alpha < \lambda < \lambda_s$.

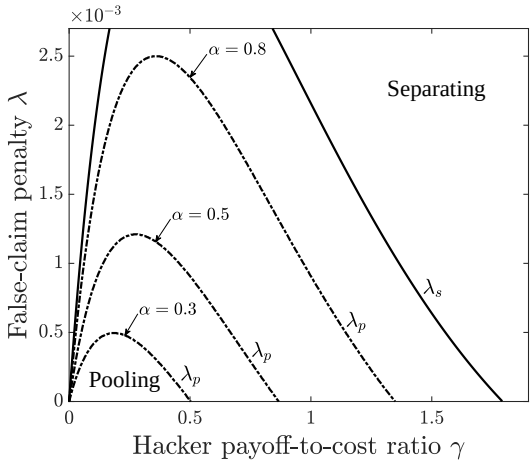
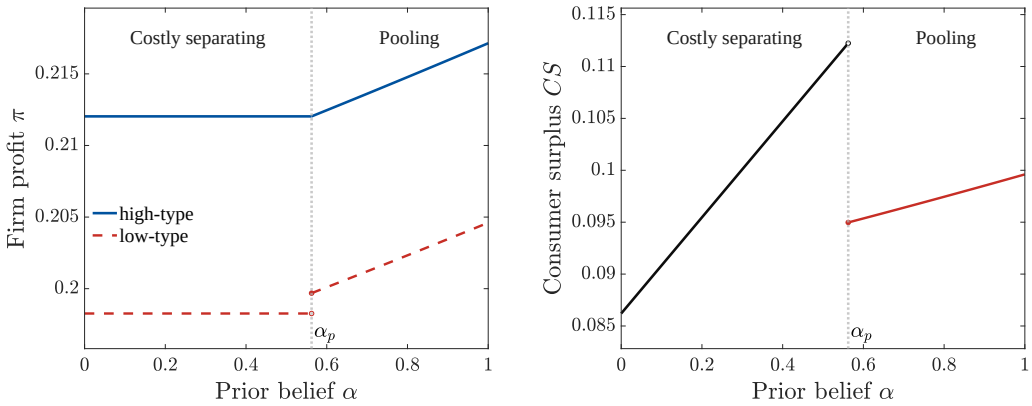


Figure A6 Effect of prior belief α on the pooling-separating boundary λ_p (Parameters: $F = 0.4$, $z = 0.1$).



(a) Firm profits (b) Consumer surplus

Figure A7 Effect of prior belief α on firm profits and consumer surplus (Parameters: $z = 0.4$, $F = 0.3$, $\gamma = 0.15$, $\lambda = 0.0045$).

Figure A7 shows that, as α increases, the LMSE switches from costly separation to pooling at $\alpha_p \approx 0.562$. The high type's profit is continuous at the transition because it is indifferent between the two equilibria, whereas the low type's profit jumps upward. Under separation, type-specific profits do not depend on α; under pooling, both types' profits increase as consumers become more likely to believe the firm is highly secure. Within each regime, consumer surplus increases with α. However, it falls discontinuously at α_p when the equilibrium switches to pooling. Thus, a more favorable prior belief can encourage cyberwashing and reduce consumer surplus at the equilibrium transition, even though it raises consumer surplus within each regime.

Moreover, the comparison between consumer surplus as $\alpha \uparrow \alpha_p$ and as $\alpha \uparrow 1$ depends on the false-claim penalty. At the switching point, separating consumer surplus is a weighted average of the type-specific surpluses, with $CS^s_H > \lim_{\alpha \uparrow 1} CS^p(\alpha) > CS^s_L$. A higher penalty increases the switching threshold α_p, thereby placing greater weight on the high-type surplus. Hence, when $\lambda > \lambda_\alpha$,

consumer surplus immediately before the switch remains higher than even the limiting pooling consumer surplus as $\alpha \uparrow 1$. The reverse holds when $\lambda \leq \lambda_\alpha$.

A.11. Security-Dependent Marginal Costs and Breach-Triggered Penalties

This section states the two modifications of the single-crossing property summarized in Section 6.4.

Remark A1 (Security-dependent marginal costs). Suppose that serving a customer also entails a type-specific marginal cost $0 \leq k_i < 1 - \gamma(1 - z_i)F$ for $i = H, L$, for instance because stronger protection raises the cost of operating each account. The reduced form (7) becomes $\pi_i = [p - k_i - \gamma(1 - z_i)F]\, N_c(p, \beta)$ net of the penalty, and the symmetric-information price is $p_i^m = [1 + k_i + \gamma(1 - z_i)F]/2$. The secure firm has the lower effective marginal cost, and hence the lower full-information price and the downward price distortion, exactly when $k_H - k_L < \gamma z F$: the attack-exposure advantage of security must outweigh any direct cost disadvantage. The single-crossing identity (8) then reads $\pi_H - \pi_L = [\gamma z F - (k_H - k_L)] N_c(p, \beta) + \lambda \mathbf{1}\{r = r_H\}$, and the qualitative analysis goes through with $z\gamma F$ replaced by $\gamma z F - (k_H - k_L)$ in the rent of Lemma 2, with the thresholds recomputed from the modified reduced form. When $k_H - k_L > \gamma z F$ the ordering reverses: the secure firm is the high-cost type, its full-information price is higher, and it separates, when it must, by distorting its price upward. At $k_H - k_L = \gamma z F$, the customer-base term in $\pi_H - \pi_L$ vanishes, so the single-crossing argument does not determine the direction of the price distortion. Table A4 illustrates with the parameters of Figure 3; the sign of the distortion changes at $k_H - k_L = \gamma z F$ and the pooling region remains an interval of low penalties on either side.

Remark A2 (Breach-triggered detection). If the detection of a false claim is triggered partly by breaches, so that the low type's expected penalty for claiming r_H is $\lambda_0 + \lambda_1 N_h$ with $\lambda_1 \geq 0$, then by (6) the penalty equals $\lambda_0 + \lambda_1 \gamma N_c(p, \beta)$ and the low type's reduced-form profit is $[p - \gamma(F + \lambda_1)] N_c(p, \beta) - \lambda_0$. The variable part of the penalty raises the low type's effective marginal cost of serving customers and leaves the high type's unchanged, so the cost gap in (8) rises from $z\gamma F$ to $z\gamma F + \gamma\lambda_1$. Every argument in Sections 4 and 5 that relies on the single-crossing property is therefore reinforced; the separating regions weakly expand, and the secure firm's informational rent weakly increases. The constant-penalty baseline is thus the conservative case for the credibility of security claims.

A.12. Silence, Claim-Monotone Beliefs, and the Unrestricted Message Game

We have restricted attention to equilibria in which the high type claims r_H, supported by the belief restriction $\beta(z_H \mid r_H, p) \geq \beta(z_H \mid r_L, p)$ for all p (*claim-monotone beliefs*, CM) and a truth-telling tie-break. Because r_L includes silence, the restriction removes an equilibrium in which neither type claims and no penalty is paid. The *unrestricted game* makes both claims available to both types and restricts beliefs only by Bayes' rule on the path; the LMSE ordering of Section 3.5 is applied

Table A4 Direction of the high type's price distortion with security-dependent marginal costs (Parameters: $z = 0.4$, $F = 0.05$, $\gamma = 2$, $k_L = 0$).

$(k_H - k_L)/(\gamma z F)$	p_H^m	p_H^{sep} at $\lambda = \lambda_p^+$	distortion	λ_p
0.0	0.5300	0.3457	downward	0.00556
0.5	0.5400	0.3596	downward	0.00806
0.9	0.5480	0.3707	downward	0.00993
1.1	0.5520	0.7277	upward	0.01019
1.5	0.5600	0.7326	upward	0.00939

to its PBE outcomes unchanged. Notation follows Tables A5 and A6 in Online Supplement C and the proof of Proposition 2. The proofs of the results in this section are available from the authors upon request; numerical statements refer to the parameter combinations described in Online Supplement A.8.

Silent pooling and an equivalence result: A *silent-pooling* PBE at price p has both types choose (r_L, p), the prior at that pair and $\beta = 0$ elsewhere. Whenever silent pooling and cyberwashing pooling are both sustainable at the same price, silent pooling has the same customer base, attack intensity, consumer surplus and high-type profit as cyberwashing pooling, and the low type's profit is higher by λ.

PROPOSITION A5. *For every admissible (γ, λ), $\lambda > 0$:*
(i) the LMSE outcome is pooling iff $0 < \gamma < \bar{\gamma}$ and $\lambda \le \lambda_p(\gamma)$, costly separating iff $\max\{\lambda_p(\gamma), 0\} < \lambda < \lambda_s(\gamma)$, and costless separating otherwise; all thresholds of Propositions 3, 4 and 9 are unchanged.
(ii) In the pooling regime both types choose (r_L, p_H^m): the high type's profit, customer bases, attack intensities and consumer surplus coincide with the LMSE pooling outcome of Proposition 3, the low type's profit is higher by λ (no penalty is paid); in the separating regimes, prices, customer bases, attack intensities, payoffs, and welfare coincide with those of Proposition 1, although the high type's claim need not be unique.
(iii) Under the monetary-transfer interpretation of the false-claim penalty, consumer surplus and social welfare coincide with $CS(\gamma, \lambda)$ and $SW(\gamma, \lambda)$, so Propositions 5–9 and Corollary 1 retain their allocation and welfare conclusions. The low type's pooling profit in Proposition 7(i) is instead constant in λ. Under the reputational-loss interpretation, silent-pooling welfare is higher by $(1-\alpha)\lambda$ than welfare at the corresponding cyberwashing-pooling outcome.

The restriction determines claim choices, including the form of pooling, while preserving the allocations, welfare, and regime boundaries in Sections 4–5 and the separating payoffs: in both games the penalty works by reducing the low type's temptation to mimic the high type's separating pair, not through any penalty paid on the path.

Claim-monotone beliefs: CM restricts how consumers read claims.

LEMMA A1. *In any cyberwashing pooling PBE of the baseline model with $\lambda > 0$, both types choose (r_H, p^{pool}), and beliefs satisfy $\beta(z_H \mid r_L, p^{pool}) < \beta(z_H \mid r_H, p^{pool}) = \alpha$.*

LEMMA A2. *Let (r, p^*) be the pair chosen by both types in a pooling PBE. Because the high claim costs the low type λ and the high type nothing, the beliefs at which the low type weakly gains from (r_H, p^*) are a subset of those at which the high type strictly gains, and conversely at (r_L, p^*). The D1 criterion therefore assigns $\beta(z_H \mid r_H, p^*) = 1$ in a silent-pooling PBE, which the high type then breaks, and $\beta(z_H \mid r_L, p^*) = 0$ in a cyberwashing pooling PBE, as in Proposition 2. Both beliefs are claim-monotone.*

At the high type's symmetric-information price, silent pooling also fails the Intuitive Criterion and D1. The price-cut deviation in the proof of Proposition A2(i) remains profitable for the high type, whose equilibrium payoff is unchanged. The low type is even less willing to deviate because its equilibrium payoff is higher by the false-claim penalty under silent pooling, whereas its deviation payoff is unchanged. The undefeated criterion does not discriminate between the two forms of pooling, so Proposition A2(ii) is a statement within the maintained equilibrium class. Pooling does not survive D1 on the full signal, as shown in Proposition A2(i); the selected outcome rests on CM, which we maintain as an additional belief restriction, and on the pessimistic price beliefs that Section 3.5 defends.

CM is the reading of claims used in the relevant literatures: silence is the worst news in verifiable-disclosure models (Grossman, 1981; Milgrom, 1981; Milgrom and Roberts, 1986), and messages that are costlier to overstate are read monotonically in costly-talk models (Kartik et al., 2007; Kartik, 2009). Section 3's binary message space with the high type claiming high quality follows Corts (2014), Rhodes and Wilson (2018), Piccolo et al. (2018) and Wu and Geylani (2020). Cyberwashing pooling requires the high-security claim to be interpreted more favorably than silence at the equilibrium price (Lemma A1). Evidence on security and privacy assurances finds positive effects on initial trust (Hu et al., 2010) or weak and null effects (McKnight et al., 2004; Kim et al., 2008), providing evidence consistent with nonnegative trust effects of assurances, although these studies do not directly identify the off-path beliefs assumed here.

A foundation without belief restrictions: Suppose a fraction $\sigma \in (0,1)$ of consumers buy only from a firm that makes the high-security claim (e.g., procurement rules or platform filters), without the claim-monotone belief restriction; the same-price truth-telling tie-break of Section 3.4 is retained. Under r_H the reduced form is unchanged; under r_L demand is $q_\sigma(p, \beta) = (1-\sigma)(1-p)/[1+(1-\sigma)\gamma(1-\beta z)]$ with $N_h = \gamma q_\sigma$. Let $\Lambda(\sigma) = \sigma\pi_L^m/[1+(1-\sigma)\gamma]$, the value of the claim to a firm *known* to be the low type, and assume $\lambda \geq \Lambda(\sigma)$.

PROPOSITION A6. *Let $\lambda^{\mathrm{e}} = \lambda - \Lambda(\sigma) \geq 0$. (i) At any common belief β, the high type strictly prefers r_H to r_L for $\gamma(1-z)F < p < 1$. Every PBE in which it chooses r_L gives it strictly less than the best PBE with claim r_H. (ii) The LMSE outcome is unique and has the high type claim r_H; its price, customer base, attack intensity and profit, the low type's claim and price, and the regime are those of Proposition 3 at the penalty λ^{e} (the shift is exact because λ enters only the low type's constraints), so the boundaries in λ are $\lambda_p(\gamma) + \Lambda(\sigma)$ and $\lambda_s(\gamma) + \Lambda(\sigma)$; in the pooling regime both types claim r_H and the penalty is paid; under separation the revealed low type earns $\pi_L^m - \Lambda(\sigma)$. As $\sigma \to 0$, $\Lambda(\sigma) \to 0$ and the outcome converges to that of Proposition 3.*

At $\lambda^{\mathrm{e}} = 0$, extend the candidate prices and payoffs of Proposition 3 continuously to zero: the LMSE is pooling if $\lambda_p(\gamma) \geq 0$, costly separating if $\lambda_p(\gamma) < 0 < \lambda_s(\gamma)$, and costless separating if $\lambda_s(\gamma) \leq 0$. At $\lambda_p(\gamma) = 0$, the high type is indifferent and the low type strictly prefers pooling, so the LMSE selects pooling. Prices outside the interval in part (i) yield nonpositive high-type payoffs under either claim, whereas claiming r_H at $p = [1 + \gamma(1-z)F]/2$ guarantees a positive payoff under any belief. The restriction of Section 3.4 is thus the $\sigma \downarrow 0$ limit of a model in which the claim has direct demand value without imposing claim-monotone beliefs; the stated truth-telling tie-break resolves same-price indifference.

Extensions: The reported numerical checks support the allocation equivalence in Proposition A5 for the examined prior beliefs $\alpha \in \{0.3, 0.5, 0.8\}$ (Online Supplement A.10) and demand-dependent loss parameters $\theta \in \{0, 0.2, 0.5, 0.8\}$ (Section 6.2). The timing of Section 6.1 is the one exception: because the inference channel makes pooling worth more to the high type, the maintained-class pooling price is constrained above the high type's optimum for $\lambda \in (\lambda_u^d, \lambda_p^d]$, where λ_u^d is the penalty at which the low type's pooling constraint binds at that unconstrained price, while silence is not, so the unrestricted game pools on a weakly larger set of penalties; the difference is negligible for moderate z (Figure A4: 0.0137 versus 0.0139) and material for large z (Figure 6(b) at $\gamma = 1$: 0.066 versus 0.093). Outside the sufficient condition $F \leq 4(1-z)$ of Proposition 10(i), a gap can occur above the largest penalty supporting pooling and below λ_e^d, the threshold for the existence of separation defined in its proof. Silent pooling fills this gap in the example of Section 6.1. The separating-regime comparative statics remain unchanged. However, removing CM can alter both the pooling boundary and within-pooling responses when the low type's incentive constraint binds.

References

Grossman SJ (1981) The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics* 24(3):461–483.

Hu X, Wu G, Wu Y, Zhang H (2010) The effects of Web assurance seals on consumers' initial trust in an online vendor: A functional perspective. *Decision Support Systems* 48(2):407–418.

Kartik N (2009) Strategic communication with lying costs. *Review of Economic Studies* 76(4):1359–1395.

Kartik N, Ottaviani M, Squintani F (2007) Credulity, lies, and costly talk. *Journal of Economic Theory* 134(1):93–116.

Kim DJ, Steinfield C, Lai YJ (2008) Revisiting the role of web assurance seals in business-to-consumer electronic commerce. *Decision Support Systems* 44(4):1000–1015.

McKnight DH, Kacmar CJ, Choudhury V (2004) Shifting factors and the ineffectiveness of third party assurance seals: A two-stage model of initial trust in a web business. *Electronic Markets* 14(3):252–266.

Milgrom PR (1981) Good news and bad news: Representation theorems and applications. *Bell Journal of Economics* 12(2):380–391.

Milgrom P, Roberts J (1986) Relying on the information of interested parties. *RAND Journal of Economics* 17(1):18–32.

Online Supplement B. Proofs

This part of the Online Supplement provides the proofs of the lemmas, propositions, and corollaries stated in the main text and Online Supplement A, with those in Section A.12 available from the authors upon request.

B.1. Proofs of the Results

Consistency of the pure-strategy assessments. Write $m_i = (r_i, p_i)$ for type i's equilibrium signal and let ν be the probability measure that assigns probability $1/2$ to each claim and is uniform on $[0,1]$ conditional on the claim. For $\varepsilon_n \downarrow 0$, let $k_{i,n}$ be a continuous triangular probability density relative to ν, concentrated on claim r_i and prices within ε_n of p_i, with the interval truncated at 0 and 1 when necessary. Its height at m_i is of order $1/\varepsilon_n$. If $m_H = m_L$, use the same density for both types. Define the strictly positive signal densities $f_{H,n} = (1-\varepsilon_n^2)k_{H,n} + \varepsilon_n^2$ and $f_{L,n} = (1-\varepsilon_n)k_{L,n} + \varepsilon_n$. Each integrates to one, has full support, and converges weakly to the corresponding pure signal. Bayes' rule gives

$$\beta_n(m) = \frac{\alpha f_{H,n}(m)}{\alpha f_{H,n}(m) + (1-\alpha) f_{L,n}(m)}.$$

At any fixed off-path signal, both triangular densities eventually vanish, so $\beta_n(m) = \alpha\varepsilon_n/[\alpha\varepsilon_n + 1-\alpha] \to 0$. At distinct on-path signals, the appropriate triangular density dominates and the posterior converges to one at m_H and zero at m_L. Under pooling, the common triangular density dominates for both types and the posterior converges to α. Thus the Bayesian on-path beliefs and the pessimistic off-path beliefs used to support the paper's pure outcomes are generated by a single sequence of full-support trembles.

The costless-separation refinement proof below sometimes assigns belief one to the additional off-path signal $m_e = (r_L, p_H^m)$ only when $\lambda_s(\gamma) \le 0$. This signal is distinct from both separating signals because $p_H^m < p_L^m$. To obtain that limit, take $0 < \varepsilon_n < 1/4$, let $k_{e,n}$ be a triangular density centered at m_e with price width ε_n^2, and replace the high type's density by

$$f_{H,n} = (1 - \varepsilon_n^2 - \varepsilon_n^{3/2})k_{H,n} + \varepsilon_n^2 + \varepsilon_n^{3/2} k_{e,n}.$$

The extra component has vanishing mass but height of order $\varepsilon_n^{-1/2}$ at m_e, so $\beta_n(m_e) \to 1$; all other stated posterior limits and both weak limits are unchanged. Consumers' purchase rules and hackers' attack rules can additionally receive independent binary-action trembles that vanish with n. These do not alter the signal-based posterior limits, and their limiting rules are the sequentially rational continuation rules verified in the PBE analysis. This establishes consistency of the supporting assessments under the density-based convention stated in Section 3.5.

Proof of Lemma 1. Let p_H denote the price charged by the high-type firm. The expected utility of a consumer with valuation v is given by $U_c = v - p_H - \bar{N}_{h,H}(1-z)$, where $\bar{N}_{h,H}$ is the consumers'

expectation of the number of attacking hackers. The indifferent consumer who is willing to purchase the product has a marginal valuation $v_H = p_H + \bar{N}_{h,H}(1-z)$. Then, the realized market demand is $N_{c,H} = 1 - v_H = 1 - p_H - \bar{N}_{h,H}(1-z)$.

Simultaneously, hackers will launch an attack if their expected payoff exceeds the cost, i.e., $h\bar{N}_{c,H} \geq c$, where $\bar{N}_{c,H}$ is the hackers' expectation of the demand. The realized number of attacking hackers is $N_{h,H} = \frac{h\bar{N}_{c,H}}{\bar{c}}$. In a rational expectations equilibrium, the realized values must match the expected values, yielding $N_{c,H} = \bar{N}_{c,H}$ and $N_{h,H} = \bar{N}_{h,H}$. Solving for $N_{c,H}$ and $N_{h,H}$, we obtain $N_{c,H} = \frac{1-p_H}{1+\gamma(1-z)}$ and $N_{h,H} = \frac{\gamma(1-p_H)}{1+\gamma(1-z)}$, where $\gamma = \frac{h}{\bar{c}}$.

Given the demand and the number of hackers, the high-type firm's profit function is formulated as

$$\pi_H = p_H \cdot \frac{1-p_H}{1+\gamma(1-z)} - \frac{\gamma(1-p_H)}{1+\gamma(1-z)} \cdot (1-z)F.$$

Taking the first-order derivative of π_H with respect to p_H yields $\frac{\partial \pi_H}{\partial p_H} = \frac{1-2p_H+\gamma F(1-z)}{1+\gamma(1-z)}$. Solving the first-order condition $\frac{\partial \pi_H}{\partial p_H} = 0$, we obtain the optimal price $p_H^m = \frac{1+\gamma F(1-z)}{2}$. Substituting p_H^m into the profit function, it follows that $\pi_H^m = \frac{(1-\gamma F(1-z))^2}{4(1+\gamma(1-z))}$.

Next, let p_L denote the price set by the low-type firm. This scenario is mathematically equivalent to the high-type case with $z = 0$. The consumer utility becomes $U_c = v - p_L - \bar{N}_{h,L}$, and hackers launch an attack if $h\bar{N}_{c,L} \geq c$. Following the identical rational expectations equilibrium derivations, we obtain $N_{c,L} = \frac{1-p_L}{1+\gamma}$ and $N_{h,L} = \frac{\gamma(1-p_L)}{1+\gamma}$. The profit function of the low-type firm is therefore

$$\pi_L = p_L \cdot \frac{1-p_L}{1+\gamma} - \frac{\gamma(1-p_L)}{1+\gamma} F.$$

Applying the first-order condition similarly yields the low-type firm's optimal price and profit $p_L^m = \frac{1+\gamma F}{2}$ and $\pi_L^m = \frac{(1-\gamma F)^2}{4(1+\gamma)}$.

Finally, $p_L^m - p_H^m = \frac{z\gamma F}{2} > 0$, so the high-type firm charges a lower price than the low-type firm. Direct subtraction also gives $\pi_H^m - \pi_L^m = \frac{z\gamma[2(1+\gamma)(1-\gamma F)F + (1+\gamma)\gamma F^2 z + (1-\gamma F)^2]}{4(1+\gamma)[1+\gamma(1-z)]} > 0$ under $\gamma F < 1$. This proves the result. □

Proof of Corollaries A1 and A2. Substituting p_i^m into $N_{c,i} = (1-p_i^m)/[1+\gamma(1-z_i)]$ and $N_{h,i} = \gamma N_{c,i}$ gives the symmetric-information expressions of Online Supplement A.4; $N_{c,H}^m > N_{c,L}^m$ because the numerator $1-\gamma F(1-z_i)$ is larger and the denominator $1+\gamma(1-z_i)$ is smaller for $i = H$. In a separating equilibrium the high type's pair is believed to come from the high type ($\beta = 1$) and the low type charges p_L^m with $\beta = 0$, so the low type's outcomes coincide with its symmetric-information outcomes, and $N_{c,H}^{sep} = (1-p_H^{sep})/[1+\gamma(1-z)]$ with p_H^{sep} from Proposition 1. In the costless case $p_H^{sep} = p_H^m$. In the costly case $p_H^{sep} = (1+\gamma F - S(\lambda))/2$, which gives the stated expression, and $N_{c,H}^{sep} > N_{c,H}^m$ because $S(\lambda) > z\gamma F$ for $\lambda < \lambda_s(\gamma)$: S is decreasing in λ and $S(\lambda_s) = z\gamma F$, since $p_H^{sep} = p_H^m$ at $\lambda = \lambda_s$. The rankings between types follow from $N_{c,H}^{sep} \geq N_{c,H}^m > N_{c,L}^m = N_{c,L}^{sep}$,

and the attack intensities are γ times the customer bases. In the LMSE pooling equilibrium both types charge p_H^m and $\beta = 1/2$, so $N_c^p = (1 - p_H^m)/[1 + \gamma(1 - z/2)] = \frac{1-\gamma F(1-z)}{2+\gamma(2-z)}$. Compared with $N_{c,L}^m = \frac{1-\gamma F}{2(1+\gamma)}$, the numerator is larger and the denominator smaller, so $N_c^p > N_{c,L}^m$; compared with $N_{c,H}^m = \frac{1-\gamma F(1-z)}{2+2\gamma(1-z)}$, the numerators coincide and the denominator is larger, so $N_c^p < N_{c,H}^m$; the last inequality is Corollary A1. □

Proof of Lemma 2. By (8), for any pair (r_H, p) and belief β, $\pi_H(r_H, p \mid \beta) = \pi_L(r_H, p \mid \beta) + \lambda + z\gamma F\, N_c(p, \beta)$. If the low type is indifferent between mimicking at (r_H, p) with belief β and truthful disclosure, then $\pi_L(r_H, p \mid \beta) = \pi_L^m$, and the lemma follows with $N_{c,H} = N_c(p, \beta)$. For the comparison with semi-separating outcomes, fix a belief $\beta \in (0, 1]$ and consider any price at which the low type's constraint binds. With $d(\beta) = 1 + \gamma(1 - \beta z)$, binding indifference gives $(p - \gamma F)(1 - p) = (\pi_L^m + \lambda)d(\beta)$. The two price roots therefore imply the customer-base bound

$$N_{c,H} \le N_{c,+}(\beta) \equiv \frac{(1-\gamma F)/2 + \sqrt{(1-\gamma F)^2/4 - (\pi_L^m + \lambda)d(\beta)}}{d(\beta)}.$$

This compares candidate outcomes and does not presume that changing price leaves the equilibrium posterior unchanged. Whenever the roots are real, $N_{c,+}(\beta)$ is strictly increasing in β: its denominator decreases, while its positive numerator increases. When pure separation is costly, $N_{c,+}(1) = N_{c,H}^{sep}$. Hence every binding outcome with $\beta < 1$ has a strictly smaller customer base than the pure costly separating outcome, irrespective of which price root it uses. □

Proof of Proposition 1. Given the posterior belief β, a consumer with expected utility $U_c = v - p - \bar{N}_h(1 - \beta z) \ge 0$ will purchase the product. Meanwhile, hackers with an expected payoff $h \cdot \bar{N}_c \ge c$ will choose to launch an attack. By solving the rational expectations equilibrium conditions, i.e., $\bar{N}_h = N_h$ and $\bar{N}_c = N_c$, the realized number of active consumers and attacking hackers are given by $N_c(p \mid \beta) = \frac{1-p}{1+\gamma(1-\beta z)}$ and $N_h(p \mid \beta) = \frac{\gamma(1-p)}{1+\gamma(1-\beta z)}$, respectively.

Accordingly, the expected profit function of the type-i firm, given belief β, is formulated as

$$\pi_i(r, p \mid \beta) = \begin{cases} \frac{p(1-p)}{1+\gamma(1-\beta z)} - \frac{\gamma(1-p)}{1+\gamma(1-\beta z)} \cdot (1 - z)F, & \text{if } i = H, \\ \frac{p(1-p)}{1+\gamma(1-\beta z)} - \frac{\gamma(1-p)}{1+\gamma(1-\beta z)} \cdot F - \lambda \cdot \mathbf{1}\{r = r_H\}, & \text{if } i = L. \end{cases}$$

The incentive compatibility (IC) constraints for a separating equilibrium ensure that neither type of firm has an incentive to mimic the other. These are given by

$$\pi_H(r_H, p_H^{sep} \mid \beta = 1) \ge \pi_H(r_L, p_L^{sep} \mid \beta = 0) \tag{ICH}$$

$$\pi_L(r_L, p_L^{sep} \mid \beta = 0) \ge \pi_L(r_H, p_H^{sep} \mid \beta = 1) \tag{ICL}$$

We first specify the belief system and the full set of deviations. Beliefs are $\beta(r_H, p_H^{sep}) = 1$ and $\beta(r_L, p_L^{sep}) = 0$ on the equilibrium path, and $\beta(r, p) = 0$ for every other claim–price pair in $\{r_H, r_L\} \times [0, 1]$. Under the admissibility condition $\gamma < \bar{\gamma}_B$, the continuation outcome after every pair is interior,

so $\pi_i(r,p \mid \beta)$ is given by the expression above for every (r,p,β). Since $\gamma F<1$, each type can guarantee a positive payoff by claiming r_L at $p=[1+\gamma(1-z_i)F]/2$. Deviations with $p\le\gamma(1-z_i)F$ or $p=1$ yield nonpositive payoffs and can therefore be excluded. On $\gamma(1-z_i)F<p<1$, profit is strictly increasing in β, so $\beta=0$ is the least favorable belief. Among pairs assigned $\beta=0$, the most profitable deviation of type i is $\max_p \pi_i(r,p \mid \beta=0)$, which is attained at $p=\frac{1+\gamma F(1-z_i)}{2}$ with value $\pi_{i,dev}=\frac{[1-\gamma F(1-z_i)]^2}{4(1+\gamma)}$ (for the low type, with $r=r_L$ to avoid the penalty). For the low-type firm, its equilibrium pair (r_L,p_L^m) attains $\pi_{L,dev}=\pi_L^m=\frac{(1-\gamma F)^2}{4(1+\gamma)}$, so no deviation to a pair with belief zero is profitable; the only remaining deviation is to the high type's pair (r_H,p_H^{sep}), which is governed by (ICL). For the high-type firm, deviations to pairs with belief zero are unprofitable if and only if

$$\pi_H(r_H,p_H^{sep} \mid \beta=1) \;\ge\; \pi_{H,dev}=\frac{[1-\gamma F(1-z)]^2}{4(1+\gamma)}, \tag{19}$$

which implies (ICH) because (r_L,p_L^{sep}) is one such pair. We verify (19) at the end of the proof; it is slack in every separating equilibrium characterized below, so only the low-type firm's constraint (ICL) binds. In any separating equilibrium, consumers correctly infer the firm's true type. For the low-type firm, its optimal strategy coincides with the symmetric information outcome: it chooses r_L and sets $p_L^{sep}=p_L^m$, yielding the optimal reservation profit $\pi_L^{sep}=\pi_L^m=\frac{(1-\gamma F)^2}{4(1+\gamma)}$.

To maximize its profit while maintaining separation, the high-type firm's equilibrium strategy (r_H,p_H^{sep}) must solve the following constrained optimization problem:

$$\begin{aligned}\max_{p_H^{sep}}\quad & \pi_H(r_H,p_H^{sep} \mid \beta=1)=\frac{p_H^{sep}(1-p_H^{sep})}{1+\gamma(1-z)}-\frac{\gamma(1-p_H^{sep})}{1+\gamma(1-z)}(1-z)F\\ \text{s.t.}\quad & \frac{p_H^{sep}(1-p_H^{sep})}{1+\gamma(1-z)}-\frac{\gamma(1-p_H^{sep})}{1+\gamma(1-z)}\cdot F-\lambda\le\frac{(1-\gamma F)^2}{4(1+\gamma)}\end{aligned}$$

For notational convenience, we use p_H to denote p_H^{sep} in the subsequent calculations. The unconstrained optimal price that maximizes $\pi_H(r_H,p_H \mid \beta=1)$ is $\bar p_H=\frac{1+\gamma F(1-z)}{2}$. To characterize the low-type firm's IC constraint, we introduce the quadratic form $w(p_H)=-p_H^2+(1+\gamma F)p_H-C$, where $C=\frac{4(1+\gamma)[\gamma F+\lambda(1+\gamma(1-z))]+(1+\gamma(1-z))(1-\gamma F)^2}{4(1+\gamma)}$. It follows that the required constraint is satisfied if and only if $w(p_H)\le 0$.

Case 1: If the discriminant $(1+\gamma F)^2-4C\le 0$, then $w(p_H)\le 0$ naturally holds for any p_H. In this case, the (ICL) constraint is slack, and the high-type firm's optimal price is its first-best price $p_H^{sep}=\bar p_H$.

Case 2: If $(1+\gamma F)^2-4C>0$, the equation $w(p_H)=0$ has two distinct roots, $0<p_{H1}<p_{H2}<1$, given by $p_{H1,2}=\frac{1+\gamma F\mp S(\lambda)}{2}$ where $S(\lambda)=\sqrt{\frac{z\gamma(1-\gamma F)^2-4(1+\gamma)(1+\gamma(1-z))\lambda}{1+\gamma}}$. The constraint $w(p_H)\le 0$ is satisfied if $p_H\in(0,p_{H1}]\cup[p_{H2},1)$. Both roots bind L's constraint, and $p_{H2}-p_{H1}=S(\lambda)$, so $\pi_H(r_H,p_{H1} \mid 1)-\pi_H(r_H,p_{H2} \mid 1)=\frac{z\gamma F S(\lambda)}{1+\gamma(1-z)}>0$. Since $p_{H2}>\bar p_H$ and profit decreases above $\bar p_H$, the high type's optimal constrained price is $p_H^{sep}=\min\{\bar p_H,p_{H1}\}$.

Substituting the expression for C, we can expand the discriminant as

$$(1+\gamma F)^2 - 4C = \frac{-4(1+\gamma)(1+\gamma(1-z))\lambda + z\gamma(1-\gamma F)^2}{1+\gamma}.$$

It is straightforward to verify that $(1+\gamma F)^2 - 4C \le 0$ when $\lambda \ge \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(1+\gamma(1-z))}$. Conversely, $(1+\gamma F)^2 - 4C > 0$ when $0 < \lambda < \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(1+\gamma(1-z))}$. Under this condition, we evaluate the difference $\bar{p}_H - p_{H1} = \frac{\sqrt{(1+\gamma F)^2 - 4C} - z\gamma F}{2}$. The inequality $\bar{p}_H \le p_{H1}$ is algebraically equivalent to $(1+\gamma F)^2 - 4C \le z^2\gamma^2 F^2$. Substituting the expression for $(1+\gamma F)^2 - 4C$, we obtain $\lambda \ge \lambda_s(\gamma) = \frac{z\gamma[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2]}{4(1+\gamma)(1+\gamma(1-z))}$. It is clear that $\lambda_s(\gamma) < \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(1+\gamma(1-z))}$.

To determine the sign of the threshold $\lambda_s(\gamma)$, we define the quadratic function $f(\gamma) = F^2(1-z)\gamma^2 - F(2+zF)\gamma + 1$. It is easy to verify that $\mathrm{sgn}(\lambda_s(\gamma)) = \mathrm{sgn}(f(\gamma))$. Recall the assumption that $0 < \gamma < \frac{1}{F}$. Within this range, it can be algebraically proven that $f(\gamma) > 0$ if $0 < \gamma < \gamma_s$, where $\gamma_s = \frac{2+zF-\sqrt{z(zF^2+4(1+F))}}{2F(1-z)}$. Otherwise, if $\gamma_s \le \gamma < \frac{1}{F}$, we have $f(\gamma) \le 0$. Consequently, it follows that $\lambda_s(\gamma) > 0$ if $0 < \gamma < \gamma_s$, and $\lambda_s(\gamma) \le 0$ if $\gamma_s \le \gamma < \frac{1}{F}$.

Thus, the high-type firm's equilibrium price is characterized as follows:

(i) If $\gamma_s \le \gamma < \frac{1}{F}$, we unconditionally have $p_H^{sep} = \bar{p}_H = \frac{1+\gamma F(1-z)}{2}$, resulting in a costless separating equilibrium.

(ii) If $0 < \gamma < \gamma_s$, we have $\lambda_s(\gamma) > 0$. The equilibrium outcome then depends on λ. When $\lambda \ge \lambda_s(\gamma)$, the constraint is satisfied at the first-best price, yielding $p_H^{sep} = \bar{p}_H = \frac{1+\gamma F(1-z)}{2}$, i.e., a costless separating equilibrium. When $0 < \lambda < \lambda_s(\gamma)$, the constraint binds, forcing the high-type firm to distort its price downward to $p_H^{sep} = p_{H1} = \frac{1+\gamma F - S(\lambda)}{2}$, i.e., a costly separating equilibrium.

It remains to verify (19). In a costless separating equilibrium, $\pi_H(r_H, \bar{p}_H \mid 1) = \pi_H^m = \frac{[1-\gamma F(1-z)]^2}{4[1+\gamma(1-z)]} > \pi_{H,dev}$ because $1+\gamma(1-z) < 1+\gamma$. In a costly separating equilibrium, the high type's profit $\pi_H^{sep}(\lambda) = \frac{[1-\gamma F+S(\lambda)][1-\gamma F(1-2z)-S(\lambda)]}{4[1+\gamma(1-z)]}$ is increasing in λ on $(0, \lambda_s(\gamma))$ (its derivative is $[S(\lambda) - z\gamma F]/S(\lambda) > 0$; see the proof of Proposition 7), so it suffices to check the limit $\lambda \to 0^+$. Let $k = \sqrt{z\gamma/(1+\gamma)}$ and $a = 1+\gamma(1-z) = (1+\gamma)(1-k^2)$. Since $S(0) = (1-\gamma F)k$, a direct computation gives

$$\pi_H^{sep}(0^+) - \pi_{H,dev} = \frac{z\gamma F}{4(1+\gamma)} \left\{ \frac{2(1-\gamma F)\big[k(1+\gamma) + z\gamma\big]}{a} - z\gamma F \right\}.$$

In the costly region $\lambda_s(\gamma) > 0$, which is equivalent to $(1-\gamma F)^2 > z\gamma(1+\gamma)F^2$, i.e., $1-\gamma F > Fk(1+\gamma)$. Hence $2(1-\gamma F)[k(1+\gamma)+z\gamma] > 2Fk(1+\gamma)[k(1+\gamma)+z\gamma] = 2Fz\gamma(1+\gamma)(1+k)$, using $k^2(1+\gamma) = z\gamma$, and therefore the bracketed term exceeds $2Fz\gamma(1+\gamma)(1+k)/a - z\gamma F > 2Fz\gamma - z\gamma F > 0$, using $a < 1+\gamma$. Thus (19) holds strictly for every $\lambda > 0$, and (ICH) is slack. □

Proof of Proposition 2. In any pooling equilibrium, both the high-type and low-type firms issue the same high-security claim r_H and charge the same equilibrium price p^{pool}. To characterize the feasible pooling prices, let p^p denote a candidate pooling price. Upon observing this equilibrium

strategy, consumers' posterior belief coincides with their prior belief, i.e., $\beta = \alpha$. Consequently, the realized number of active consumers and attacking hackers are $N_c(r_H, p^p) = \frac{1-p^p}{1+\gamma(1-\alpha z)}$ and $N_h(r_H, p^p) = \frac{\gamma(1-p^p)}{1+\gamma(1-\alpha z)}$, respectively. Given the belief α, the expected profit function of the i-type firm is formulated as

$$\pi_i(r_H, p^p \mid \beta = \alpha) = \begin{cases} \frac{p^p(1-p^p)}{1+(1-\alpha z)\gamma} - \frac{\gamma(1-p^p)}{1+(1-\alpha z)\gamma} \cdot (1-z)F, & \text{if } i = H, \\ \frac{p^p(1-p^p)}{1+(1-\alpha z)\gamma} - \frac{\gamma(1-p^p)}{1+(1-\alpha z)\gamma} \cdot F - \lambda, & \text{if } i = L. \end{cases}$$

Off the equilibrium path, we set $\beta = 0$, under which consumers believe the firm to be of the low type. This corresponds to the worst possible belief for a deviating high-type firm. Thus, the highest possible profit a low-type firm can get by deviating is its perfect-information profit $\pi_{L,dev} = \pi_L^m$. If the high-type firm deviates by charging $p_{H,dev}$, its expected profit under the worst-case belief becomes $\pi_{H,dev}(p_{H,dev} \mid \beta = 0) = \frac{(p_{H,dev} - (1-z)\gamma F)(1-p_{H,dev})}{1+\gamma}$. Maximizing this expression with respect to $p_{H,dev}$ yields the optimal deviation price $p_{H,dev} = \frac{1+(1-z)\gamma F}{2}$, which results in the maximum deviation profit $\pi_{H,dev} = \frac{(1-(1-z)\gamma F)^2}{4(1+\gamma)}$.

To ensure that neither type of firm has an incentive to deviate, any valid pooling price p^p must satisfy the following incentive compatibility constraints:

$$\pi_H(r_H, p^p \mid \beta = \alpha) \geq \pi_{H,dev}, \tag{ICH}$$

$$\pi_L(r_H, p^p \mid \beta = \alpha) \geq \pi_{L,dev} \tag{ICL}$$

Let $w_H(p^p) = \pi_H(r_H, p^p \mid \beta = \alpha) - \pi_{H,dev}$ and $w_L(p^p) = \pi_L(r_H, p^p \mid \beta = \alpha) - \pi_{L,dev}$. Expanding these expressions yields

$$w_H(p^p) = \frac{(p^p - (1-z)\gamma F)(1-p^p)}{1+(1-\alpha z)\gamma} - \frac{(1-\gamma F(1-z))^2}{4(1+\gamma)},$$

$$w_L(p^p) = \frac{(p^p - \gamma F)(1-p^p)}{1+(1-\alpha z)\gamma} - \lambda - \frac{(1-\gamma F)^2}{4(1+\gamma)}.$$

By solving the quadratic inequality $w_H(p^p) \geq 0$, we establish that the price must fall within the interval $p^p \in [p_1^p, p_2^p]$, where $p_{1,2}^p = \frac{1}{2}\left(1 + \gamma F(1-z) \mp (1-\gamma F(1-z))\sqrt{\frac{\alpha z \gamma}{1+\gamma}}\right)$.

Regarding the low-type constraint $w_L(p^p) \geq 0$, if the condition $-4(1+\gamma)(1+(1-\alpha z)\gamma)\lambda + \alpha z \gamma (1-\gamma F)^2 < 0$ holds, the discriminant is negative, meaning $w_L(p^p) < 0$ for all p^p, thereby violating constraint (ICL). Hence, a necessary condition for a pooling equilibrium to exist is $0 < \lambda \leq \frac{\alpha z \gamma (1-\gamma F)^2}{4(1+\gamma)(1+(1-\alpha z)\gamma)}$. Under this boundary condition, solving $w_L(p^p) \geq 0$ yields the feasible range $p^p \in [p_3^p, p_4^p]$, where $p_{3,4}^p = \frac{1}{2}\left(1 + \gamma F \mp \sqrt{\frac{-4(1+\gamma)(1+(1-\alpha z)\gamma)\lambda + \alpha z \gamma (1-\gamma F)^2}{1+\gamma}}\right)$.

To satisfy both constraints simultaneously, the intersection $[p_1^p, p_2^p] \cap [p_3^p, p_4^p]$ must be nonempty. Since it is algebraically straightforward that $p_4^p > p_1^p$, the non-emptiness condition reduces to $p_2^p \geq p_3^p$.

For analytical tractability, we henceforth assume the prior belief is uniform, $\alpha = 1/2$. Substituting $\alpha = 1/2$, the penalty threshold simplifies to $0 < \lambda \leq \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(2+(2-z)\gamma)}$. It follows that

$$p_2^p - p_3^p = \frac{1}{2}\left[\sqrt{\frac{z\gamma(1-\gamma F)^2}{2(1+\gamma)} - 2(2+(2-z)\gamma)\lambda} + \sqrt{\frac{z\gamma}{2(1+\gamma)}}(1-\gamma F(1-z)) - z\gamma F\right].$$

Case 1: When $\sqrt{\frac{z\gamma}{2(1+\gamma)}}(1-\gamma F(1-z)) \geq z\gamma F$, it follows that $(1-4z+z^2)F^2\gamma^2 - 2F(1-z+zF)\gamma + 1 \geq 0$. Since $0 < \gamma < 1/F$, this inequality holds if and only if $0 < \gamma \leq \gamma_1$, where $\gamma_1 = \frac{1}{F\left[1-(1-F)z+\sqrt{z[2+F(2-(2-F)z)]}\right]}$. In this case, $p_2^p \geq p_3^p$ holds for $0 < \lambda \leq \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(2+(2-z)\gamma)}$, with equality if and only if $\gamma = \gamma_1$ and λ equals this upper bound.

Case 2: When $\sqrt{\frac{z\gamma}{2(1+\gamma)}}(1-\gamma F(1-z)) < z\gamma F$, which corresponds to $\gamma_1 < \gamma < 1/F$, the condition $p_2^p \geq p_3^p$ is equivalent to $\lambda \leq \tilde{\lambda}$, where $\tilde{\lambda} = \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(2+(2-z)\gamma)} - \frac{2(1+\gamma)\left[z\gamma F-(1-\gamma F(1-z))\sqrt{\frac{z\gamma}{2(1+\gamma)}}\right]^2}{4(1+\gamma)(2+(2-z)\gamma)}$. It is clear that $\tilde{\lambda} < \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(2+(2-z)\gamma)}$. Moreover, it requires $\tilde{\lambda} > 0$, which yields the condition $(z^2-6z+4)F^2\gamma^2 - 2F(4-2z+zF)\gamma + 4 > 0$. We obtain $0 < \gamma < \gamma_p$, where $\gamma_p = \frac{4}{F\left[4-2z+zF+\sqrt{z[8+F(8-4z+zF)]}\right]}$. It is algebraically verifiable that $\gamma_1 < \gamma_p$.

Define $\lambda_1(\gamma) = \begin{cases} \frac{z\gamma(1-\gamma F)^2}{4(1+\gamma)(2+(2-z)\gamma)}, & 0 < \gamma \leq \gamma_1, \\ \tilde{\lambda}, & \gamma_1 < \gamma < \gamma_p. \end{cases}$. The two weak incentive constraints are jointly feasible if and only if $0 < \gamma < \gamma_p$ and $0 < \lambda \leq \lambda_1(\gamma)$, with common feasible interval $[p_1^p, p_2^p] \cap [p_3^p, p_4^p]$. This interval is nondegenerate for $\lambda < \lambda_1(\gamma)$ and a singleton at $\lambda = \lambda_1(\gamma)$. For $0 < \gamma \leq \gamma_1$, this singleton is $p_3^p = p_4^p = p_L^m = (1+\gamma F)/2$. At that price, L earns π_L^m both by claiming r_H under belief α and by claiming r_L under belief zero, so the same-price truth-telling tie-break excludes pooling. For $\gamma_1 < \gamma < \gamma_p$, the singleton is $p_2^p = p_3^p < p_L^m$, so truthful disclosure at that same price earns strictly less than π_L^m and pooling is supported. Thus pooling exists precisely on the region stated in Proposition 2. Every remaining candidate is supported by $\beta(r_H, p^{pool}) = \alpha$ on the path and $\beta(r,p) = 0$ elsewhere: the two weak constraints rule out all profitable deviations, and the preceding qualification enforces the same-price tie-break. □

Proof of Propositions 3 and 4. We first present the detailed regions and the complete mathematical statement, combining both propositions into Proposition B1 below. Throughout, $\lambda(\bar{\gamma}_B)$ abbreviates $\lambda_p(\bar{\gamma}_B)$.

Proposition B1. *The unique LMSE outcome is characterized as follows:*

(i) The LMSE is costless separating if $\lambda \geq \lambda_s(\gamma)$, costly separating if $\lambda_p(\gamma) < \lambda < \lambda_s(\gamma)$, and pooling if $0 < \lambda \leq \lambda_p(\gamma)$.

(ii) Moreover, $\lambda_s(\gamma)$ is strictly unimodal on $(0, \gamma_s)$, and $\lambda_p(\gamma)$ is strictly unimodal on $(0, \bar{\gamma})$. Furthermore, for any given $\lambda > \bar{\lambda}_p$, the LMSE is separating in the whole admissible range; for any given $0 < \lambda \leq \bar{\lambda}_p$, there exist thresholds $\bar{\gamma}_1(\lambda) \leq \bar{\gamma}_2(\lambda)$ such that:

If $F \geq F^o$, the LMSE is separating when $0 < \gamma < \bar{\gamma}_1(\lambda)$, pooling when $\bar{\gamma}_1(\lambda) \leq \gamma \leq \bar{\gamma}_2(\lambda)$, and switches back to separating when $\bar{\gamma}_2(\lambda) < \gamma < \bar{\gamma}_B$.

If $F < F^o$, define γ_0 as the unique solution to $\lambda_p(\gamma) = \bar{\lambda}_p$.

- *When $\bar{\gamma}_B \leq \gamma_0$, if $\lambda(\bar{\gamma}_B) \leq \lambda \leq \bar{\lambda}_p$, the LMSE is separating in the whole admissible range $0 < \gamma < \bar{\gamma}_B$. If $0 < \lambda < \lambda(\bar{\gamma}_B)$, the LMSE is separating for $0 < \gamma < \bar{\gamma}_1(\lambda)$ and pooling for $\bar{\gamma}_1(\lambda) \leq \gamma < \bar{\gamma}_B$.*
- *When $\bar{\gamma}_B > \gamma_0$, if $\lambda(\bar{\gamma}_B) < \lambda \leq \bar{\lambda}_p$, the LMSE is separating for $0 < \gamma < \bar{\gamma}_1(\lambda)$, pooling for $\bar{\gamma}_1(\lambda) \leq \gamma \leq \bar{\gamma}_2(\lambda)$, and switches back to separating for $\bar{\gamma}_2(\lambda) < \gamma < \bar{\gamma}_B$. If $0 < \lambda \leq \lambda(\bar{\gamma}_B)$, the LMSE is separating for $0 < \gamma < \bar{\gamma}_1(\lambda)$ and pooling for $\bar{\gamma}_1(\lambda) \leq \gamma < \bar{\gamma}_B$.*

The proof of Proposition B1 proceeds in four steps as follows:

i. Optimal pooling price. We first identify the pooling equilibrium selected by LMSE within the pooling-existence region of Proposition 2. Without the restriction of $p^p \in [p_1^p, p_2^p] \cap [p_3^p, p_4^p]$, the high-type firm's profit $\pi_H(r_H, p^p \mid \beta = \alpha)$ is maximized at $p_m^p = \frac{1+\gamma F(1-z)}{2}$. It is straightforward that $p_m^p \in [p_1^p, p_2^p]$ and $p_m^p < p_4^p$. Therefore, p_m^p is feasible if and only if $p_m^p \geq p_3^p$. By calculation, $p_m^p \geq p_3^p$ is equivalent to $\lambda \leq \lambda_2(\gamma)$, where $\lambda_2(\gamma) = \frac{z\gamma\left[(1-\gamma F)^2 - 2z\gamma F^2(1+\gamma)\right]}{4(1+\gamma)(2+(2-z)\gamma)}$. For $\lambda_2(\gamma) > 0$, it is necessary and sufficient that $(1-2z)F^2\gamma^2 - 2F(1+zF)\gamma + 1 > 0$. Solving this quadratic inequality yields $\lambda_2(\gamma) > 0$ if and only if $0 < \gamma < \gamma_2$, where $\gamma_2 = \frac{1}{F\left[1+zF+\sqrt{z(2+(2+zF)F)}\right]}$. It is algebraically verifiable that $\gamma_2 < \gamma_1 < \gamma_p$. Moreover, for $0 < \gamma \leq \gamma_1$, we have $\lambda_1(\gamma) - \lambda_2(\gamma) = \frac{z^2\gamma^2F^2}{2(2+(2-z)\gamma)} > 0$.

Consequently, when $0 < \gamma < \gamma_2$ and $0 < \lambda \leq \lambda_2(\gamma)$, it follows that $p_m^p \in [p_3^p, p_4^p]$, and the LMSE pooling price is $p^{pool} = \frac{1+\gamma F(1-z)}{2}$. Within the pooling-existence region, in all remaining cases, we have $p_m^p < p_3^p$. Since the high-type firm's profit is strictly decreasing over the feasible region $[p_3^p, p_4^p]$, the LMSE pooling price is $p^{pool} = p_3^p$.

ii. LMSE selection. Next, we compare π_H^{sep} and π_H^{pool} to characterize the LMSE outcome. When the costless separating equilibrium exists, i.e., $\gamma \geq \gamma_s = \frac{2+zF-\sqrt{z\left(zF^2+4(1+F)\right)}}{2F(1-z)}$, or $0 < \gamma < \gamma_s$ and $\lambda \geq \lambda_s(\gamma) = \frac{z\gamma[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2]}{4(1+\gamma)(1+\gamma(1-z))}$, the high-type firm's profit under any pooling equilibrium is bounded above by its unconstrained maximum $\frac{(1-\gamma F(1-z))^2}{2(2+(2-z)\gamma)}$. Under the costless separating equilibrium, its profit is $\frac{(1-\gamma F(1-z))^2}{4(1+\gamma(1-z))}$. Therefore, we have $\pi_H^{sep} > \pi_H^{pool}$ whenever the costless separating equilibrium exists.

When the costly separating equilibrium exists, i.e., $0 < \gamma < \gamma_s$ and $0 < \lambda < \lambda_s(\gamma)$, we have $\pi_H^{sep} = \frac{[1-\gamma F+S(\lambda)][1-\gamma F(1-2z)-S(\lambda)]}{4[1+(1-z)\gamma]}$, where $S(\lambda) = \sqrt{\frac{z\gamma(1-\gamma F)^2 - 4(1+\gamma)(1+(1-z)\gamma)\lambda}{1+\gamma}}$. We then compare the high-type firm's profit under the costly separating equilibrium with that under the pooling equilibrium.

Case 1: When $p^{pool} = p_3^p$, the high-type firm's profit is $\pi_H^{pool} = \frac{[2(1-\gamma F)+S_1(\lambda)][2(1-\gamma F(1-2z))-S_1(\lambda)]}{8[2+(2-z)\gamma]}$, where $S_1(\lambda) = \sqrt{\frac{2z\gamma(1-\gamma F)^2 - 8(1+\gamma)(2+(2-z)\gamma)\lambda}{1+\gamma}}$. We proceed by contradiction to show that $\pi_H^{sep} > \pi_H^{pool}$. Suppose, to the contrary, that $\pi_H^{sep} \leq \pi_H^{pool}$. Then, it follows that

$$[2+(2-z)\gamma] \cdot S(\lambda) \leq -z\gamma(1-\gamma F) + [1+(1-z)\gamma] \cdot S_1(\lambda). \tag{B1}$$

A necessary condition for inequality (B1) to hold is that $[1+(1-z)\gamma]\cdot S_1(\lambda) > z\gamma(1-\gamma F)$, which is equivalent to $8(1+\gamma)(1+(1-z)\gamma)^2\lambda - z\gamma(1-\gamma F)^2(1-\gamma(2z-1)) < 0$. By calculation, we have $\lambda < \frac{z\gamma(1-\gamma F)^2(1-(2z-1)\gamma)}{8(1+\gamma)(1+(1-z)\gamma)^2}$ with $1-\gamma(2z-1) > 0$. Then, we square both sides of inequality (B1) and obtain

$$S_1(\lambda) \le \frac{2(2+(2-z)\gamma)\lambda}{1-\gamma F} - \frac{(1-\gamma F)(2+(2+z)\gamma)}{2(1+\gamma)}. \tag{B2}$$

Similarly, for the inequality (B2) to hold, it is required that $\frac{2(2+(2-z)\gamma)\lambda}{1-\gamma F} - \frac{(1-\gamma F)(2+(2+z)\gamma)}{2(1+\gamma)} > 0$, which is equivalent to $4(1+\gamma)(2+(2-z)\gamma)\lambda - (1-\gamma F)^2(2+\gamma(z+2)) > 0$. By calculation, we have $\lambda > \frac{(1-\gamma F)^2(2+(2+z)\gamma)}{4(1+\gamma)(2+(2-z)\gamma)}$. If $\pi_H^{sep} \le \pi_H^{pool}$ were to hold, then it would require that $\lambda < \frac{z\gamma(1-\gamma F)^2(1-(2z-1)\gamma)}{8(1+\gamma)(1+(1-z)\gamma)^2}$ with $1-\gamma(2z-1) > 0$, and $\lambda > \frac{(1-\gamma F)^2(2+(2+z)\gamma)}{4(1+\gamma)(2+(2-z)\gamma)}$ simultaneously. However, it is shown that $\frac{z\gamma(1-\gamma F)^2(1-(2z-1)\gamma)}{8(1+\gamma)(1+(1-z)\gamma)^2} - \frac{(1-\gamma F)^2(2+(2+z)\gamma)}{4(1+\gamma)(2+(2-z)\gamma)} = \frac{-(1-\gamma F)^2[(5z^2-8z+4)\gamma^2+8(1-z)\gamma+4]}{8(2+(2-z)\gamma)(1+\gamma(1-z))^2} < 0$. A contradiction arises. Therefore, $\pi_H^{sep} \le \pi_H^{pool}$ cannot hold, and it follows that $\pi_H^{sep} > \pi_H^{pool}$.

Case 2: When $p^{pool} = \frac{1+\gamma F(1-z)}{2}$, the high-type firm's profit is $\pi_H^{pool} = \frac{(1-\gamma F(1-z))^2}{2(2+(2-z)\gamma)}$. By Step i. this pooling price is selected if and only if $0 < \gamma < \gamma_2$ and $0 < \lambda \le \lambda_2(\gamma)$. First, we show that the existence conditions for this pooling equilibrium are a subset of those for the costly separating equilibrium.

(i) Comparison of γ_s and γ_2. If $z \ne 1/2$, it can be shown that γ_s is the smaller positive root of $f_1(\gamma) = (1-z)F^2\gamma^2 - F(2+zF)\gamma + 1 = 0$, and γ_2 is the unique root in $(0, 1/F)$ of $f_2(\gamma) = (1-2z)F^2\gamma^2 - 2F(1+zF)\gamma + 1 = 0$. By calculation, we have $f_1(\gamma) - f_2(\gamma) = z\gamma F^2(1+\gamma) > 0$. Since $f_2(0) = 1 > 0$ and $f_2(\gamma_s) = -z\gamma_s F^2(1+\gamma_s) < 0$, the unique root γ_2 in $(0, 1/F)$ satisfies $\gamma_2 < \gamma_s$.

If $z = 1/2$, it can be shown that γ_2 is the root of $\bar{f}_2(\gamma) = -2F(1+F/2)\gamma + 1$. Substituting γ_s into $\bar{f}_2(\gamma)$ yields $\bar{f}_2(\gamma_s) = (2+F)\sqrt{\frac{F^2}{4}+2F+2} - \left(\frac{F^2}{2}+3F+3\right)$. To determine its sign, consider the inequality $(2+F)\sqrt{\frac{F^2}{4}+2F+2} < \frac{F^2}{2}+3F+3$. Since both sides are positive, squaring both sides gives $(2+F)^2\left(\frac{F^2}{4}+2F+2\right) < \left(\frac{F^2}{2}+3F+3\right)^2$. Expanding and simplifying, this inequality reduces to $-(1+F)^2 < 0$. Hence, $\bar{f}_s(\gamma_s) < 0$ which implies $\gamma_2 < \gamma_s$.

(ii) Comparison of $\lambda_s(\gamma)$ and $\lambda_2(\gamma)$. For $0 < \gamma < \gamma_2$, we have $\lambda_s(\gamma) - \lambda_2(\gamma) = \frac{z\gamma[(1-\gamma F)^2 - z^2\gamma^2F^2]}{4(1+\gamma(1-z))(2+\gamma(2-z))}$. Since $\lambda_2(\gamma) > 0$, it follows that $(1-\gamma F)^2 > 2z\gamma(1+\gamma)F^2$. Moreover, $2z\gamma(1+\gamma)F^2 - z^2\gamma^2F^2 = z\gamma F^2(2+\gamma(2-z)) > 0$, which implies $(1-\gamma F)^2 > 2z\gamma(1+\gamma)F^2 > z^2\gamma^2F^2$. Therefore, $\lambda_2(\gamma) < \lambda_s(\gamma)$ holds. Consequently, the intersection of the existence conditions of the two equilibria coincides with the existence region of pooling equilibrium with $p^{pool} = \frac{1+\gamma F(1-z)}{2}$, that is, $0 < \gamma < \gamma_2$ and $0 < \lambda \le \lambda_2(\gamma)$.

Second, we compare the high-type firm's profits under the separating and pooling equilibria for $0 < \gamma < \gamma_2$ and $0 < \lambda \le \lambda_2(\gamma)$. Let $\Delta_H(\lambda)$ denote the profit difference between the costly separating equilibrium and the pooling equilibrium:

$$\Delta_H(\lambda) = \pi_H^{sep}(\lambda) - \pi_H^{pool}.$$

Note that π_H^{pool} is completely independent of λ when $p^{pool} = \frac{1+\gamma F(1-z)}{2}$. Taking the first-order derivative of $\Delta_H(\lambda)$ with respect to λ, we obtain $\frac{\partial \Delta_H(\lambda)}{\partial \lambda} = \frac{\partial \pi_H^{sep}}{\partial \lambda} = \frac{S(\lambda)-z\gamma F}{S(\lambda)}$. If $S(\lambda) > z\gamma F$, solving this inequality yields $0 < \lambda < \frac{z\gamma[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2]}{4(1+\gamma)(1+\gamma(1-z))}$. Recall that $\lambda_s(\gamma) = \frac{z\gamma[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2]}{4(1+\gamma)(1+\gamma(1-z))} > \lambda_2(\gamma)$. Therefore, it follows that $\frac{\partial \Delta_H(\lambda)}{\partial \lambda} > 0$ for $0 < \lambda \le \lambda_2(\gamma)$. Evaluating at $\lambda = \lambda_2$ gives $\pi_H^{sep}(\lambda_2) = \frac{(1-\gamma F)^2 + 2z\gamma F(1-\gamma F) - S(\lambda_2)^2 + 2z\gamma F S(\lambda_2)}{4(1+(1-z)\gamma)}$. Since $S(\lambda_2) = \sqrt{\frac{2z^2\gamma^2F^2(1+\gamma(1-z)) + z\gamma(1-\gamma F)^2}{2+(2-z)\gamma}}$, it follows that $\pi_H^{sep}(\lambda_2) = \frac{(1-\gamma F)^2 - z^2\gamma^2F^2}{2(2+\gamma(2-z))} + \frac{z\gamma F(1-\gamma F + S(\lambda_2))}{2(1+\gamma(1-z))}$. Then, we can obtain $\Delta_H(\lambda_2) = \frac{(1-\gamma F)^2 - z^2\gamma^2F^2}{2(2+\gamma(2-z))} + \frac{z\gamma F(1-\gamma F + S(\lambda_2))}{2(1+\gamma(1-z))} - \frac{(1-\gamma F + z\gamma F)^2}{2(2+\gamma(2-z))} = z\gamma F \cdot \frac{S(\lambda_2)(2+\gamma(2-z)) - z\gamma[2F(1+\gamma(1-z)) - (1-\gamma F)]}{2(1+\gamma(1-z))(2+\gamma(2-z))}$. Let $L_1 = S(\lambda_2)(2+\gamma(2-z))$ and $L_2 = z\gamma[2F(1+\gamma(1-z)) - (1-\gamma F)]$. Then, $L_1^2 - L_2^2 = 2z\gamma(1+\gamma(1-z))[1-\gamma F(1-z)]^2 > 0$. Since $L_1 > 0$, it follows that $L_1 > |L_2| \ge L_2$. Hence, $\Delta_H(\lambda_2) > 0$.

Accordingly, solving the equation $\pi_H^{sep}(\lambda) - \pi_H^{pool} = 0$ yields the threshold $\lambda = \lambda_p(\gamma)$, such that $\pi_H^{sep}(\lambda) \le \pi_H^{pool}$ for $0 < \lambda \le \max\{\lambda_p(\gamma), 0\}$, and $\pi_H^{sep}(\lambda) > \pi_H^{pool}$ for $\max\{\lambda_p(\gamma), 0\} < \lambda \le \lambda_2(\gamma)$. Through algebraic calculation, the expression for this threshold is given by:

$$\lambda_p(\gamma) = \frac{z\gamma F(1-\gamma F(1-z))}{2(2+\gamma(2-z))(1+\gamma(1-z))} \cdot G(\gamma),$$

where $G(\gamma) = \frac{(1+\gamma)(1-\gamma F)^2 - z\gamma[(1+F)^2 + (1+\gamma)^2F^2]}{2F(1+\gamma)(1-\gamma F + z\gamma F)} - \sqrt{z\gamma(2+\gamma(2-z))}$.

Differentiating $G(\gamma)$ with respect to γ, we obtain $\frac{\partial G(\gamma)}{\partial \gamma} = -\frac{1}{2} - \frac{z(1+F)\left[F(1+\gamma)^2 + (1+F)(1+F(1-z)\gamma^2)\right]}{2F(1+\gamma)^2(1-\gamma F(1-z))^2} - \frac{\sqrt{z}(1+\gamma(2-z))}{\sqrt{\gamma(2+\gamma(2-z))}} < 0$ for $0 < \gamma < \gamma_2$. It follows that $G(0) = \frac{1}{2F} > 0$. Note that γ_2 is the root of $(1-2z)F^2\gamma^2 - 2F(1+zF)\gamma + 1 = 0$. Evaluating at $\gamma = \gamma_2$ yields $z = \frac{(1-\gamma_2 F)^2}{2\gamma_2 F^2(1+\gamma_2)}$. Substituting $z\gamma_2 = \frac{(1-\gamma_2 F)^2}{2F^2(1+\gamma_2)}$, we obtain $\frac{(1+\gamma_2)(1-\gamma_2 F)^2 - z\gamma_2[(1+F)^2 + (1+\gamma_2)^2F^2]}{2F(1+\gamma_2)(1-\gamma_2 F + z\gamma_2 F)} = -\frac{(1-\gamma_2 F)^2}{2F^2(1+\gamma_2)}$. Hence, we have $G(\gamma_2) = -\frac{(1-\gamma_2 F)^2}{2F^2(1+\gamma_2)} - \sqrt{z\gamma_2(2+\gamma_2(2-z))} < 0$. Therefore, there exists a unique $\bar{\gamma} \in (0, \gamma_2)$ such that $G(\bar{\gamma}) = 0$, and hence $\lambda_p(\bar{\gamma}) = 0$. Moreover, since $\frac{z\gamma F(1-\gamma F(1-z))}{2(2+\gamma(2-z))(1+\gamma(1-z))} > 0$, it follows that $\lambda_p(\gamma) \ge 0$ for $0 < \gamma \le \bar{\gamma}$, and $\lambda_p(\gamma) < 0$ for $\bar{\gamma} < \gamma < \gamma_2$.

Consequently, for $0 < \gamma \le \bar{\gamma}$, we have $\pi_H^{sep} \le \pi_H^{pool}$ for $0 < \lambda \le \lambda_p(\gamma)$, and $\pi_H^{sep} > \pi_H^{pool}$ for $\lambda_p(\gamma) < \lambda \le \lambda_2(\gamma)$. For $\bar{\gamma} < \gamma < \gamma_2$, $\pi_H^{sep}(\lambda) > \pi_H^{pool}$ holds for all $0 < \lambda \le \lambda_2(\gamma)$. At $\lambda = \lambda_p(\gamma)$, the high-type firm is indifferent between separating and pooling, whereas the low-type firm earns a strictly higher profit under pooling. Hence, the equilibrium at $\lambda = \lambda_p(\gamma)$ is pooling. Combining the conditions for the costless and costly separating equilibria in Proposition 1 establishes part (i).

iii. Unimodality of λ_s and λ_p. First, we establish that $\lambda_s(\gamma)$ is strictly unimodal over $(0, \gamma_s)$. By rearranging terms, $\lambda_s(\gamma)$ can be written as $\lambda_s(\gamma) = \frac{z\gamma \cdot g(\gamma)}{4(1+\gamma)(1+\gamma(1-z))}$ where $g(\gamma) = F^2(1-z)\gamma^2 - F(2+zF)\gamma + 1$. Solving $g(\gamma) = 0$ yields the roots $\gamma = \frac{2+zF \pm \sqrt{z(zF^2+4(1+F))}}{2F(1-z)}$. It is straightforward to verify that γ_s is the smaller positive root. Hence, $g(\gamma_s) = 0$ and $g(\gamma) > 0$ for all $\gamma \in (0, \gamma_s)$. In addition, we have $\lambda_s(0) = 0$. It follows that $\lambda_s(\gamma) > 0$ for $\gamma \in (0, \gamma_s)$. Furthermore, we have $\lambda_s'(\gamma_s) = \frac{z\gamma_s}{4(1+\gamma_s)[1+\gamma_s(1-z)]}\, g'(\gamma_s) < 0$ since $g'(\gamma_s) < 0$.

The first-order derivative of $\lambda_s(\gamma)$ with respect to γ is given by

$$\frac{\partial \lambda_s(\gamma)}{\partial \gamma} = \frac{zT(\gamma)}{4(1+\gamma)^2(1+\gamma(1-z))^2}$$

where $T(\gamma) = F^2(1-z)^2\gamma^4 + 2F^2(2-z)(1-z)\gamma^3 + [(3F^2-1)(1-z) - (2-z)(2+zF)F]\gamma^2 - 2F(2+zF)\gamma + 1$.

We apply Descartes' Rule of Signs to evaluate the number of positive real roots for $T(\gamma) = 0$. Regardless of whether the coefficient for γ^2 is positive, negative, or zero, the sequence of non-zero coefficients exhibits exactly two sign changes. Thus, $T(\gamma)$ has at most two positive real roots. Moreover, $T(0) = 1 > 0$, and $T(\gamma_s) < 0$. As $\gamma \to +\infty$, the term $F^2(1-z)^2\gamma^4$ dominates, implying $T(\gamma) \to +\infty$. Therefore, the sign of $T(\gamma)$ changes from positive to negative over $(0, \gamma_s)$, and from negative to positive over $(\gamma_s, +\infty)$. It implies that $T(\gamma)$ must possess at least one root in $(0, \gamma_s)$ and at least one root in $(\gamma_s, +\infty)$. Since Descartes' Rule limits the total number of positive roots to two, these roots are unique within their respective intervals. Consequently, $\lambda_s'(\gamma) = 0$ has exactly one root in $(0, \gamma_s)$, proving that $\lambda_s(\gamma)$ first increases and then decreases over $0 < \gamma < \gamma_s$.

Next, we show that $\lambda_p(\gamma)$ is strictly unimodal on $(0, \bar{\gamma})$, with a unique maximizer in this interval. Fix $F > 0$ and $0 < z < 1$, and first consider the mathematical range $0 < \gamma < 1/F$. Define $J(\gamma) = \frac{F\left[z\gamma + \sqrt{z\gamma[2+(2-z)\gamma]}\right]}{1-F\gamma}$ and $X(\gamma) = \frac{1+\gamma}{1+(1-z)\gamma} J(\gamma)\big[J(\gamma)+2\big]$. The expression for λ_p obtained in Step ii can be factored as

$$\lambda_p(\gamma) = \frac{z\gamma(1-F\gamma)^2}{4(1+\gamma)[2+(2-z)\gamma]}\,[1 - X(\gamma)].$$

The function J is strictly increasing, because its numerator is strictly increasing and its denominator is positive and strictly decreasing. Moreover,

$$\frac{d}{d\gamma}\frac{1+\gamma}{1+(1-z)\gamma} = \frac{z}{[1+(1-z)\gamma]^2} > 0.$$

Consequently, $X'(\gamma) > 0$ on $(0, 1/F)$, with $X(\gamma) \to 0$ as $\gamma \to 0^+$ and $X(\gamma) \to +\infty$ as $\gamma \to (1/F)^-$. Thus $X(\gamma) = 1$ has a unique solution, which is precisely the zero $\bar{\gamma}$ identified in Step ii. In particular, $\lambda_p(\gamma) > 0$ and $X(\gamma) < 1$ for $0 < \gamma < \bar{\gamma}$. Let $\gamma_* = \sqrt{2/(2-z)}$. Since

$$\frac{d}{d\gamma}\left\{\frac{z\gamma}{4(1+\gamma)[2+(2-z)\gamma]}\right\} = \frac{z[2-(2-z)\gamma^2]}{4(1+\gamma)^2[2+(2-z)\gamma]^2},$$

the term in braces is nonincreasing for $\gamma \geq \gamma_*$. On the positive-value interval, the other two factors, $(1-F\gamma)^2$ and $1 - X(\gamma)$, are positive and strictly decreasing. Hence, $\lambda_p'(\gamma) < 0$ whenever $\gamma_* \leq \gamma < \bar{\gamma}$. It remains to rule out multiple turning points for $0 < \gamma < \min\{\gamma_*, \bar{\gamma}\}$.

Set $x = \log\gamma$, and use the subscript x to denote differentiation with respect to x, so that $d/dx = \gamma\, d/d\gamma$. Direct differentiation gives

$$\left[\log\frac{z\gamma(1-F\gamma)^2}{4(1+\gamma)[2+(2-z)\gamma]}\right]_{xx} = -\frac{\gamma}{(1+\gamma)^2} - \frac{2(2-z)\gamma}{[2+(2-z)\gamma]^2} - \frac{2F\gamma}{(1-F\gamma)^2} < 0.$$

For $\gamma \le \gamma_*$, we also have

$$\left[\log \frac{1+\gamma}{1+(1-z)\gamma}\right]_{xx} = \frac{z\gamma[1-(1-z)\gamma^2]}{(1+\gamma)^2[1+(1-z)\gamma]^2} > 0,$$

because $(1-z)\gamma_*^2 = 2(1-z)/(2-z) < 1$. Furthermore,

$$\left[\log \sqrt{z\gamma[2+(2-z)\gamma]}\right]_{xx} = \frac{(2-z)\gamma}{[2+(2-z)\gamma]^2} > 0,$$
$$\left[\log \frac{1}{1-F\gamma}\right]_{xx} = \frac{F\gamma}{(1-F\gamma)^2} > 0.$$

The function $z\gamma$ is log-linear in x. Since sums and products of positive log-convex functions are log-convex, $J(\gamma)$, $J(\gamma)+2$, and $X(\gamma)$ are all log-convex as functions of x on the interval under consideration. In particular, $X_{xx} = X\left[(\log X)_{xx} + ((\log X)_x)^2\right] \ge 0$ and $X_x = \gamma X'(\gamma) > 0$. Because $X < 1$,

$$[\log(1-X)]_{xx} = -\frac{X_{xx}}{1-X} - \frac{X_x^2}{(1-X)^2} < 0.$$

Combining these inequalities with the factorization shows that $x \mapsto \log \lambda_p(e^x)$ is strictly concave for $0 < e^x < \min\{\gamma_*, \bar{\gamma}\}$. Its derivative is therefore strictly decreasing on that interval.

To establish the endpoint signs explicitly, the factorization and its derivative yield $\lim_{\gamma\to 0^+} \frac{\lambda_p(\gamma)}{\gamma} = \frac{z}{8} > 0$, and $\lim_{\gamma\to 0^+} \lambda_p'(\gamma) = \frac{z}{8} > 0$. Thus, λ_p is strictly increasing sufficiently close to zero. At the other endpoint, $X(\bar{\gamma}) = 1$ and $X'(\bar{\gamma}) > 0$ imply

$$\lambda_p'(\bar{\gamma}^-) = -\frac{z\bar{\gamma}(1-F\bar{\gamma})^2}{4(1+\bar{\gamma})[2+(2-z)\bar{\gamma}]} X'(\bar{\gamma}) < 0.$$

If $\bar{\gamma} \le \gamma_*$, the strictly decreasing logarithmic derivative therefore crosses zero exactly once before $\bar{\gamma}$. If $\bar{\gamma} > \gamma_*$, the same conclusion holds before γ_*, because $\lambda_p'(\gamma_*) < 0$, and λ_p remains strictly decreasing thereafter. Since $d\log\lambda_p(e^x)/dx = \gamma\lambda_p'(\gamma)/\lambda_p(\gamma)$ has the same sign as $\lambda_p'(\gamma)$ on the positive-value interval, there is a unique $\gamma_0 \in (0, \bar{\gamma})$ such that λ_p is strictly increasing on $(0, \gamma_0)$ and strictly decreasing on $(\gamma_0, \bar{\gamma})$. The maximum is $\bar{\lambda}_p = \lambda_p(\gamma_0)$.

iv. Admissible equilibrium regions. Finally, we restrict the thresholds to the admissible domain $0 < \gamma < \bar{\gamma}_B$, where $\bar{\gamma}_B = \min\{1/F, 1/z\}$ and examine whether $\bar{\gamma}$ and γ_s lie within the admissible range. From the preceding analysis, we have $0 < \bar{\gamma} < \gamma_2 < \gamma_s < 1/F$. If $F \ge z$, then $1/F \le 1/z$, and hence $\bar{\gamma} < \gamma_s < 1/F \le 1/z$. Thus, both $\bar{\gamma}$ and γ_s lie below $1/z$ when $F \ge z$. We next consider the case $0 < F < z$, for which $1/z < 1/F$.

We first compare γ_s with $1/z$. Evaluating $f(\gamma) = F^2(1-z)\gamma^2 - F(2+zF)\gamma + 1$ at $\gamma = 1/z$ yields $F^2(1-z)\frac{1}{z^2} - F(2+zF)\frac{1}{z} + 1 = \frac{\left[z-\left(1+\sqrt{z(1+z)}\right)F\right]\left[z-\left(1-\sqrt{z(1+z)}\right)F\right]}{z^2}$. For $0 < F < z$, it follows that if $\sqrt{z(1+z)} \ge 1$, then $1-\sqrt{z(1+z)} \le 0$, and hence $z - \left(1-\sqrt{z(1+z)}\right)F > 0$. If $\sqrt{z(1+z)} < 1$, then $0 < 1-\sqrt{z(1+z)} < 1$, and $F < z$ implies $\left(1-\sqrt{z(1+z)}\right)F < F < z$, so $z - \left(1-\sqrt{z(1+z)}\right)F >$

0. Consequently, $\operatorname{sgn} f(1/z) = \operatorname{sgn}\left[z - \left(1+\sqrt{z(1+z)}\right)F\right]$. Since $0 < F < z$ implies $1/z < 1/F$, and $1/F$ lies strictly between the two positive roots of $f(\gamma) = 0$, the sign at $\gamma = 1/z$ determines whether $1/z$ lies below or above the smaller root γ_s. Moreover, $0 < \frac{z}{1+\sqrt{z(1+z)}} < z$. Therefore, we have $\gamma_s > 1/z$ if $F < \frac{z}{1+\sqrt{z(1+z)}}$, and $\gamma_s \le 1/z$ if $F \ge \frac{z}{1+\sqrt{z(1+z)}}$. Next, we compare $\bar{\gamma}$ with $1/z$. Recall that $\bar{\gamma}$ is the unique positive root of $\lambda_p(\gamma) = 0$ for $0 < \gamma < \gamma_2$. From the preceding analysis, $\lambda_p(\gamma) = \frac{z\gamma F[1-\gamma F(1-z)]}{2[2+\gamma(2-z)][1+\gamma(1-z)]} \cdot G(\gamma)$. Moreover, $\frac{\partial G(\gamma)}{\partial \gamma} < 0$. Therefore, $G(\gamma)$ is strictly decreasing, and the position of $\bar{\gamma}$ relative to $1/z$ is determined by the sign of $\lambda_p(1/z)$. Substituting $\gamma = 1/z$, we have

$$\lambda_p\left(\frac{1}{z}\right) = \frac{z}{4}\left[\frac{z-F}{\sqrt{z(1+z)}} - F - \frac{z-F(1-z)}{\sqrt{z(2+z)}}\right] \times \left[\frac{z-F}{\sqrt{z(1+z)}} + F + \frac{z-F(1-z)}{\sqrt{z(2+z)}}\right].$$

The second factor is strictly positive for $0 < F < z$. Therefore, it follows that $\operatorname{sgn}\lambda_p(1/z) = \operatorname{sgn}\left[\frac{z-F}{\sqrt{z(1+z)}} - F - \frac{z-F(1-z)}{\sqrt{z(2+z)}}\right]$. Rearranging the first factor, we can obtain $\frac{z-F}{\sqrt{z(1+z)}} - F - \frac{z-F(1-z)}{\sqrt{z(2+z)}} = z\left[\frac{1}{\sqrt{z(1+z)}} - \frac{1}{\sqrt{z(z+2)}}\right] - F\left[1 + \frac{1}{\sqrt{z(1+z)}} - \frac{1-z}{\sqrt{z(z+2)}}\right]$. Since $\frac{1-z}{\sqrt{z(z+2)}} < \frac{1}{\sqrt{z(z+2)}} < \frac{1}{\sqrt{z(1+z)}}$, we have $1 + \frac{1}{\sqrt{z(1+z)}} - \frac{1-z}{\sqrt{z(z+2)}} > 0$. Hence, $\lambda_p(1/z) \ge 0$ if $F \le \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$, and $\lambda_p(1/z) < 0$ if $F > \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$. Because $G(\gamma)$ is strictly decreasing and $G(\bar{\gamma}) = 0$, it follows that $\bar{\gamma} \ge 1/z$ if $F \le \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$ and $\bar{\gamma} < 1/z$ if $F > \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$.

We then compare the two thresholds in F. At $F = \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$, we have $\bar{\gamma} = \frac{1}{z}$. Since the preceding analysis establishes $\bar{\gamma} < \gamma_2 < \gamma_s$, it follows that $\gamma_s > \frac{1}{z}$ at $F = \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$. From the comparison derived above for γ_s, i.e., $\gamma_s > \frac{1}{z}$ if $F < \frac{z}{1+\sqrt{z(1+z)}}$, we therefore obtain $0 < \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}} < \frac{z}{1+\sqrt{z(1+z)}} < z$.

For any fixed $0 < \lambda \le \bar{\lambda}_p$, let $\bar{\gamma}_1(\lambda) \le \bar{\gamma}_2(\lambda)$ denote the two roots of $\lambda_p(\gamma) = \lambda$. Thus, if $0 < F < F^o = \frac{z\left(\sqrt{z+2}-\sqrt{z+1}\right)}{\sqrt{z(z+1)(z+2)}+\sqrt{z+2}-(1-z)\sqrt{z+1}}$, it follows that $\frac{1}{z} < \bar{\gamma} < \gamma_s$. The equilibrium characterization then depends on the relative position of $1/z$ with respect to $\bar{\gamma}_1(\lambda)$ and $\bar{\gamma}_2(\lambda)$. Within the admissible range, the pooling set is $[\bar{\gamma}_1(\lambda), \bar{\gamma}_2(\lambda)] \cap (0, \bar{\gamma}_B)$, and the LMSE is separating elsewhere. This covers empty pooling sets and assigns admissible crossings to pooling. In particular, if $\bar{\gamma}_B > \gamma_0$ and $\lambda = \lambda_p(\bar{\gamma}_B)$, then $\bar{\gamma}_2(\lambda) = \bar{\gamma}_B$, so separation does not reemerge before the excluded upper bound. If $F \ge F^o$, it follows that $\bar{\gamma} \le \bar{\gamma}_B$. Therefore, $\lambda_p(\bar{\gamma}_B) \le 0$. Thus, we have if $0 < \lambda \le \bar{\lambda}_p$, the LMSE is separating when $0 < \gamma < \bar{\gamma}_1(\lambda)$, pooling when $\bar{\gamma}_1(\lambda) \le \gamma \le \bar{\gamma}_2(\lambda)$ and switches back to separating when $\bar{\gamma}_2(\lambda) < \gamma < \bar{\gamma}_B$. This proves part (ii).

Proof of Proposition 5. To simplify notation in the following comparative-static proofs, superscripts s and p denote separating and pooling equilibrium outcomes, respectively.

For $\lambda \le \lambda_p$, it is the pooling equilibrium. Both firm types charge $p^p = \frac{1+\gamma F(1-z)}{2}$, which is independent of λ. Hence both prices remain constant in the pooling region.

For $\lambda_p < \lambda < \lambda_s$, it is the costly separating equilibrium. The low-type firm charges its symmetric-information price $p_L^s = \frac{1+\gamma F}{2}$, which is independent of λ. At the switching threshold $\lambda = \lambda_p$, $\lim_{\lambda \to \lambda_p^+} p_L^s(\lambda) - p^p = \frac{1+\gamma F}{2} - \frac{1+\gamma F(1-z)}{2} = \frac{z\gamma F}{2} > 0$, so the low-type firm's price increases discontinuously at $\lambda = \lambda_p^+$ and remains constant thereafter. The high-type firm's costly-separating price is $p_H^s(\lambda) = \frac{1+\gamma F - S(\lambda)}{2}$. Thus, we have $\frac{\partial p_H^s}{\partial \lambda} = -\frac{1}{2}\frac{\partial S(\lambda)}{\partial \lambda} = \frac{1+\gamma(1-z)}{S(\lambda)} > 0$. Moreover, at $\lambda = \lambda_p$, we have $p^p - \lim_{\lambda \to \lambda_p^+} p_H^s(\lambda) = \frac{1+\gamma F(1-z)}{2} - \frac{1+\gamma F - S(\lambda_p)}{2} = \frac{S(\lambda_p) - z\gamma F}{2} > 0$, where the last inequality follows from $\lambda_p(\gamma) < \lambda_s(\gamma)$, which implies $S(\lambda_p) > z\gamma F$. Thus, the high-type firm's price decreases discontinuously at $\lambda = \lambda_p^+$ and then increases throughout the costly-separating region.

For $\lambda \geq \lambda_s$, it is the costless separating equilibrium and both firm types charge their symmetric-information prices, which remain constant. □

Proof of Proposition 6. In the pooling equilibrium, both firm types face the same attack intensity, $N_{h,H}^p = N_{h,L}^p = \frac{\gamma[1-\gamma F(1-z)]}{2+\gamma(2-z)}$, which is independent of λ. In the costly separating equilibrium, the high-type and low-type firms face, respectively, $N_{h,H}^s(\lambda) = \frac{\gamma[1-\gamma F + S(\lambda)]}{2[1+\gamma(1-z)]}$ and $N_{h,L}^s = \frac{\gamma(1-\gamma F)}{2(1+\gamma)}$. Since $\frac{\partial S(\lambda)}{\partial \lambda} = -\frac{2[1+\gamma(1-z)]}{S(\lambda)}$, it follows that $\frac{\partial N_{h,H}^s(\lambda)}{\partial \lambda} = -\frac{\gamma}{S(\lambda)} < 0$, and $\frac{\partial N_{h,L}^s}{\partial \lambda} = 0$. Thus, throughout the costly separating region, the attack intensity faced by the high-type firm decreases with λ, whereas that faced by the low-type firm remains constant. In the costless separating region, $N_{h,H}^s = \frac{\gamma[1-\gamma F(1-z)]}{2[1+\gamma(1-z)]}$ and $N_{h,L}^s = \frac{\gamma(1-\gamma F)}{2(1+\gamma)}$. Neither expression depends on λ.

At $\lambda = \lambda_p$, we have $S(\lambda_p) = z\gamma F + [1-\gamma F(1-z)]\sqrt{\frac{z\gamma}{2+\gamma(2-z)}}$. Substituting this into the high-type firm's attack intensity yields $\lim_{\lambda \to \lambda_p^+} N_{h,H}^s(\lambda) = \frac{\gamma[1-\gamma F(1-z)]\left[1+\sqrt{\frac{z\gamma}{2+\gamma(2-z)}}\right]}{2[1+\gamma(1-z)]}$. It follows that $\lim_{\lambda \to \lambda_p^+} N_{h,H}^s(\lambda) - N_{h,H}^p = \frac{\gamma[1-\gamma F(1-z)]\left[\gamma z + [2+\gamma(2-z)]\sqrt{\frac{z\gamma}{2+\gamma(2-z)}}\right]}{2[1+\gamma(1-z)][2+\gamma(2-z)]} > 0$. Hence, the high-type firm's attack intensity increases discontinuously at $\lambda = \lambda_p^+$. For the low-type firm, it follows that $N_{h,L}^p - \lim_{\lambda \to \lambda_p^+} N_{h,L}^s(\lambda) = \frac{\gamma^2 z[1-\gamma F + 2F(1+\gamma)]}{2(1+\gamma)[2+\gamma(2-z)]} > 0$. Thus, the low-type firm's attack intensity decreases discontinuously at $\lambda = \lambda_p^+$. □

Proof of Proposition 7. For $\lambda \leq \lambda_p$, the high-type firm's pooling profit is independent of λ, whereas the low-type firm's pooling profit includes the false-claim penalty and is therefore decreasing in λ. At $\lambda = \lambda_p$, the high-type firm is indifferent between pooling and costly separation, so its profit is continuous at the threshold. Comparing the low-type firm's pooling and separating profits at the threshold gives $\pi_L^p - \lim_{\lambda \to \lambda_p^+} \pi_L^s(\lambda) = \frac{1-\gamma F\left(2-\gamma F\left(1-z^2\right)\right)}{2\gamma(2-z)+4} - \frac{(1-\gamma F)^2}{4(1+\gamma)} - \lambda_p = \frac{z\gamma F(1-\gamma F(1-z))\left(z\gamma + \sqrt{z\gamma(\gamma(2-z)+2)}\right)}{2(2+\gamma(2-z))(1+\gamma(1-z))} > 0$. Thus the low-type firm's profit decreases discontinuously at $\lambda = \lambda_p^+$ and remains constant for $\lambda > \lambda_p$, because it earns its symmetric-information profits in both separating regimes.

It remains to show the high-type firm's profit increases in the costly-separating region. Its costly-separating profit is $\pi_H^s(\lambda) = \frac{(1-\gamma F + S(\lambda))(1-\gamma F(1-2z) - S(\lambda))}{4(1+\gamma(1-z))}$. Differentiating with respect to λ yields

$\frac{\partial \pi_H^s}{\partial \lambda} = \frac{S(\lambda) - z\gamma F}{S(\lambda)}$. Since $\lambda < \lambda_s(\gamma)$ is equivalent to $S(\lambda) > z\gamma F$ in the costly-separating region, we have $\partial \pi_H^s / \partial \lambda > 0$. For $\lambda \geq \lambda_s$, the high-type firm earns its symmetric-information profit, which is independent of λ. □

Proof of Proposition 8. Let $\rho = \sqrt{\frac{z\gamma}{2+(2-z)\gamma}}$. Under pooling, demand is $N_c^p = \frac{1-p^p}{1+\gamma(1-z/2)} = \frac{1-\gamma F(1-z)}{2+(2-z)\gamma}$. Hence, consumer surplus is $CS^p = \frac{1}{2}(CS_H^p + CS_L^p) = \frac{(N_c^p)^2}{2} = \frac{[1-\gamma F(1-z)]^2}{2[2+(2-z)\gamma]^2}$. Under costly separation, $N_{c,H}^s(\lambda) = \frac{1-p_H^s(\lambda)}{1+\gamma(1-z)} = \frac{1-\gamma F + S(\lambda)}{2(1+\gamma(1-z))}$ and $N_{c,L}^s = \frac{1-p_L^s}{1+\gamma} = \frac{1-\gamma F}{2(1+\gamma)}$. Thus, $CS^s(\lambda) = \frac{1}{2}(CS_H^s + CS_L^s) = \frac{1}{4}(N_{c,H}^s(\lambda))^2 + \frac{1}{4}(N_{c,L}^s)^2$.

i. We compare CS^p and CS^s at the threshold $\lambda = \lambda_p(\gamma)$. At this threshold, the high-type firm is indifferent between costly separating and pooling equilibrium. Thus, $\frac{(1-\gamma F + S(\lambda_p))(1-\gamma F + 2z\gamma F - S(\lambda_p))}{4(1+\gamma(1-z))} = \frac{[1-\gamma F(1-z)]^2}{2[2+(2-z)\gamma]}$. Rearranging yields $S(\lambda_p)^2 - 2z\gamma F S(\lambda_p) + (z\gamma F)^2 - \frac{z\gamma(1-\gamma F(1-z))^2}{2+(2-z)\gamma} = 0$, where $z\gamma F < (1-\gamma F(1-z))\rho$. Taking the positive root, we obtain $S(\lambda_p) = z\gamma F + (1-\gamma F(1-z))\rho$. Hence, $1 - \gamma F + S(\lambda_p) = [1-\gamma F(1-z)](1+\rho)$. It follows that $\lim_{\lambda \to \lambda_p^+} N_{c,H}^s(\lambda) = \frac{[1-\gamma F(1-z)](1+\rho)}{2[1+\gamma(1-z)]}$ and $\lim_{\lambda \to \lambda_p^+} CS^s(\lambda) = \frac{[1-\gamma F(1-z)]^2(1+\rho)^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2}$. Subtracting CS^p, we obtain

$$\Delta CS = \lim_{\lambda \to \lambda_p^+} CS^s(\lambda) - CS^p = \frac{[1-\gamma F(1-z)]^2(1+\rho)^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2} - \frac{[1-\gamma F(1-z)]^2}{2[2+(2-z)\gamma]^2}.$$

Using $\rho^2 = \frac{z\gamma}{2+(2-z)\gamma}$, and $2 + (2-z)\gamma = 2[1+\gamma(1-z)]/(1-\rho^2)$, we obtain $\frac{[1-\gamma F(1-z)]^2}{2[2+(2-z)\gamma]^2} = \frac{[1-\gamma F(1-z)]^2(1-\rho^2)^2}{8[1+\gamma(1-z)]^2}$. Substituting this into ΔCS and using $(1+\rho)^2 - 2(1-\rho^2)^2 = -(1+\rho)^2(1-4\rho+2\rho^2)$, we obtain

$$\Delta CS = \frac{1}{16}\left[\frac{(1-\gamma F)^2}{(1+\gamma)^2} - \frac{(1-\gamma F(1-z))^2}{(1+\gamma(1-z))^2}(1+\rho)^2(1-4\rho+2\rho^2)\right].$$

Since $\lambda_p > 0$, we have $S(\lambda_p) < (1-\gamma F)\sqrt{\frac{z\gamma}{1+\gamma}}$. Substituting $S(\lambda_p) = z\gamma F + (1-\gamma F(1-z))\rho$ yields $z\gamma F + [1-\gamma F(1-z)]\rho < (1-\gamma F)\sqrt{\frac{z\gamma}{1+\gamma}}$. Let $k = \sqrt{\frac{z\gamma}{1+\gamma}}$, so that $0 < k < 1$. Using $1 - \gamma F(1-z) - (1-\gamma F) = z\gamma F$, this inequality implies $1 - \gamma F > \frac{[1-\gamma F(1-z)](1+\rho)}{1+k}$.

Case 1: If $1 - 4\rho + 2\rho^2 \leq 0$, it is clear that $\Delta CS > 0$.

Case 2: If $1 - 4\rho + 2\rho^2 > 0$, then showing $\Delta CS > 0$ is equivalent to proving $\frac{1-\gamma F}{1+\gamma} > \frac{1-\gamma F(1-z)}{1+\gamma(1-z)}(1+\rho)\sqrt{1-4\rho+2\rho^2}$. Using $1 - \gamma F > \frac{(1-\gamma F(1-z))(1+\rho)}{1+k}$, it suffices to show $\frac{1}{(1+\gamma)(1+k)} \geq \frac{\sqrt{1-4\rho+2\rho^2}}{1+\gamma(1-z)}$. Note that $1+\gamma(1-z) = 1+\gamma-z\gamma = (1+\gamma)(1-\frac{z\gamma}{1+\gamma}) = (1+\gamma)(1-k^2)$. Substituting this into the inequality gives $1 - k \geq \sqrt{1-4\rho+2\rho^2}$. Since $0 < k < 1$, squaring both sides yields $1 - 2k + k^2 \geq 1 - 4\rho + 2\rho^2$, which is equivalent to $k^2 - 2k \geq 2\rho^2 - 4\rho$. Note that $\rho = \sqrt{\frac{z\gamma}{2+(2-z)\gamma}} = \sqrt{\frac{\frac{z\gamma}{1+\gamma}}{2-\frac{z\gamma}{1+\gamma}}} = \sqrt{\frac{k^2}{2-k^2}} = \frac{k}{\sqrt{2-k^2}}$. Substituting $\rho = \frac{k}{\sqrt{2-k^2}}$ into the inequality yields $k^2 - 2k \geq \frac{2k^2}{2-k^2} - \frac{4k}{\sqrt{2-k^2}}$, which is equivalent to $\frac{4}{\sqrt{2-k^2}} - 2 \geq \frac{k^3}{2-k^2}$. Multiplying both sides by $2 - k^2$, we obtain $4(\sqrt{2-k^2} - 1) + k^2(2-k) \geq 0$. Since $0 < k < 1$, the left-hand side is nonnegative, and the inequality clearly holds. Therefore, when $1 - 4\rho + 2\rho^2 > 0$, it follows that $\Delta CS > 0$. In summary, consumer surplus increases discontinuously at $\lambda = \lambda_p^+$.

ii. At $\lambda = \lambda_p$, the difference in social welfare between costly separation and pooling can be decomposed as $\Delta SW = \lim_{\lambda \to \lambda_p^+} SW^s(\lambda) - SW^p = \Delta CS + \Delta\pi + \Delta R$, where ΔR denotes the difference in the regulator's expected penalty revenue. Because no false-claim penalty is collected under separation, whereas the low-type firm pays λ_p under pooling with probability $1/2$, we have $\Delta R = -\frac{\lambda_p}{2}$. Moreover, since $\lim_{\lambda \to \lambda_p^+} \pi_H^s(\lambda) = \pi_H^p$, it follows that $\Delta\pi = \frac{1}{2}\left(\lim_{\lambda \to \lambda_p^+} \pi_L^s(\lambda) - \pi_L^p\right) = \frac{1}{2}\left[\frac{(1-\gamma F)^2}{4(1+\gamma)} - \frac{[1-\gamma F(1-z)][1-\gamma F(1+z)]}{2[2+(2-z)\gamma]} + \lambda_p\right]$. Substituting ΔCS, $\Delta\pi$ and ΔR, we obtain

$$\Delta SW = \frac{(3+2\gamma)(1-\gamma F)^2}{16(1+\gamma)^2} - \frac{[1-\gamma F(1-z)](1-\gamma F)}{2[2+(2-z)\gamma]} + \left[\frac{1}{4(2+(2-z)\gamma)} - \frac{(1+\rho)^2(1-4\rho+2\rho^2)}{16(1+\gamma(1-z))^2}\right][1-\gamma F(1-z)]^2.$$

Let $x(F) = \frac{1-\gamma F}{1-\gamma F(1-z)}$ where $0 < x < 1$. We can obtain $\Delta SW = [1-\gamma F(1-z)]^2 \xi(x)$, in which

$$\xi(x) = \frac{3+2\gamma}{16(1+\gamma)^2}x^2 - \frac{x}{2[2+(2-z)\gamma]} + \frac{1}{4[2+(2-z)\gamma]} - \frac{(1+\rho)^2(1-4\rho+2\rho^2)}{16[1+\gamma(1-z)]^2}.$$

Because $[1-\gamma F(1-z)]^2 > 0$, the sign of ΔSW is determined by the sign of $\xi(x)$.

We first characterize the feasible range of x. Recall that $S(\lambda_p) = z\gamma F + [1-\gamma F(1-z)]\rho$. Using $z\gamma F = [1-\gamma F(1-z)] - (1-\gamma F)$, we can rewrite this expression as $S(\lambda_p) = [1-\gamma F(1-z)](1+\rho) - (1-\gamma F)$. Moreover, $S(0) = (1-\gamma F)\sqrt{\frac{z\gamma}{1+\gamma}}$. Since $S(\lambda)$ is strictly decreasing in λ and $\lambda_p > 0$, we have $S(\lambda_p) < S(0)$. Therefore, $[1-\gamma F(1-z)](1+\rho) < (1-\gamma F)\left(1+\sqrt{\frac{z\gamma}{1+\gamma}}\right)$, which implies $x > x_p = \frac{1+\rho}{1+\sqrt{\frac{z\gamma}{1+\gamma}}}$. Note that $\rho^2 = \frac{z\gamma}{2+(2-z)\gamma} < \frac{z\gamma}{1+\gamma}$, because $2+(2-z)\gamma - (1+\gamma) = 1+\gamma(1-z) > 0$. Hence, $0 < \rho < \sqrt{\frac{z\gamma}{1+\gamma}}$, and therefore $0 < x_p(z,\gamma) < 1$. Thus, under $\lambda_p > 0$, the feasible range of x is $x_p < x < 1$.

We next characterize the sign of $\xi(x)$. Since $\frac{\partial^2 \xi(x)}{\partial x^2} = \frac{3+2\gamma}{8(1+\gamma)^2} > 0$, $\xi(x)$ is strictly convex in x. Its discriminant is $\Delta_\xi = \frac{1}{4[2+(2-z)\gamma]^2} - \frac{3+2\gamma}{4(1+\gamma)^2}\left[\frac{1}{4[2+(2-z)\gamma]} - \frac{(1+\rho)^2(1-4\rho+2\rho^2)}{16[1+\gamma(1-z)]^2}\right]$. We now characterize the sign of Δ_ξ. From $\rho^2 = \frac{z\gamma}{2+(2-z)\gamma}$, we have $z\gamma = \rho^2[2+(2-z)\gamma]$. Since $2+(2-z)\gamma = 2(1+\gamma) - z\gamma$, it follows that $2+(2-z)\gamma = \frac{2(1+\gamma)}{1+\rho^2}$. Similarly, $1+\gamma(1-z) = 1+\gamma-z\gamma = 1+\gamma - \frac{2(1+\gamma)\rho^2}{1+\rho^2} = \frac{(1+\gamma)(1-\rho^2)}{1+\rho^2}$. Substituting these two identities into Δ_ξ and simplifying yields $\Delta_\xi = \frac{(1+\rho^2)P}{64(1+\gamma)^4(1-\rho)^2}$, where $P = (1-\rho)^2[1+\rho^2((2\gamma+3)^2+1)] - 2\rho(2\gamma+3)$. Because $\frac{1+\rho^2}{64(1+\gamma)^4(1-\rho)^2} > 0$, we have $\operatorname{sgn}(\Delta_\xi) = \operatorname{sgn}(P)$. It follows that $P > 0$ if and only if $\{[(2\gamma+3)^2+1]\rho^4 + [(2\gamma+3)^2+2]\rho^2 + 1\}^2 - 4\rho^2\{[(2\gamma+3)^2+1]\rho^2 + (2\gamma+4)\}^2 > 0$. Substituting $\rho^2 = \frac{z\gamma}{2+(2-z)\gamma}$ and simplifying, the left-hand side becomes $\frac{4(1+\gamma)^2}{[2+(2-z)\gamma]^4}T(z,\gamma)$, where

$$T(z,\gamma) = 16z^2(1-z)^2\gamma^6 + 16z^2(3z^2-6z+2)\gamma^5 + 4z(9z^3-14z^2-14z-4)\gamma^4 + 8z(3z^2-17z-12)\gamma^3 + (4-172z-47z^2)\gamma^2 + (8-92z)\gamma + 4.$$

Since $\frac{4(1+\gamma)^2}{[2+(2-z)\gamma]^4} > 0$, we obtain $\Delta_\xi(z,\gamma) > 0 \iff T(z,\gamma) > 0$.

We next characterize the sign of $T(z,\gamma)$ with respect to γ. First, $T(z,0)=4>0$. Moreover, because the coefficient of γ^6 is $16z^2(1-z)^2>0$, we have $\lim_{\gamma\to\infty}T(z,\gamma)=+\infty$. We now show that $T(z,\gamma)$ is negative at an intermediate value of γ. Substituting $\gamma=1/\sqrt{2z}$ into $T(z,\gamma)$ gives $T\left(z,\frac{1}{\sqrt{2z}}\right)=9z^2+12\sqrt{2}\,z^{3/2}-\frac{71}{2}z-92\sqrt{2z}-100-\frac{16\sqrt{2}}{\sqrt{z}}$. Because $0<z<1$, we have $z^2<z$ and $z^{3/2}<\sqrt{z}$. Therefore, $T\left(z,\frac{1}{\sqrt{2z}}\right)<9z+12\sqrt{2}\,\sqrt{z}-\frac{71}{2}z-92\sqrt{2z}-100-\frac{16\sqrt{2}}{\sqrt{z}}=-\frac{53}{2}z-80\sqrt{2z}-100-\frac{16\sqrt{2}}{\sqrt{z}}<0$. Hence, $T(z,\gamma)=0$ has at least one positive root in $(0,1/\sqrt{2z})$ and at least one positive root in $(1/\sqrt{2z},\infty)$.

It remains to show that these are the only positive roots. We have $T(z,\gamma)=\tau_6\gamma^6+\tau_5\gamma^5+\tau_4\gamma^4+\tau_3\gamma^3+\tau_2\gamma^2+\tau_1\gamma+\tau_0$, where $\tau_6=16z^2(1-z)^2>0$, $\tau_5=16z^2(3z^2-6z+2)$, $\tau_4=4z(9z^3-14z^2-14z-4)<0$, $\tau_3=8z(3z^2-17z-12)<0$, $\tau_2=-47z^2-172z+4$, $\tau_1=8-92z$, and $\tau_0=4>0$. The coefficient τ_5 changes sign at $z=1-1/\sqrt{3}$, τ_2 changes sign at its unique positive root $z_a=\frac{-86+4\sqrt{474}}{47}$, and τ_1 changes sign at $z_b=2/23$. These thresholds satisfy $0<z_a<z_b<1-\frac{1}{\sqrt{3}}<1$. Therefore, the coefficient sign sequences are

z interval	$(\tau_6,\tau_5,\tau_4,\tau_3,\tau_2,\tau_1,\tau_0)$
$(0,z_a)$	$+,+,-,-,+,+,+$
(z_a,z_b)	$+,+,-,-,-,+,+$
$(z_b,1-1/\sqrt{3})$	$+,+,-,-,-,-,+$
$(1-1/\sqrt{3},1)$	$+,-,-,-,-,-,+.$

Each sequence has exactly two sign changes. By Descartes' Rule of Signs, $T(z,\gamma)=0$ has at most two positive roots. Since we have already established that it has at least two positive roots, it has exactly two. Denote them by $0<\gamma_2^T(z)<\frac{1}{\sqrt{2z}}<\gamma_1^T(z)$. Because $T(z,0)>0$, $T(z,1/\sqrt{2z})<0$, and $T(z,\gamma)\to+\infty$ as $\gamma\to\infty$, these two roots imply $T(z,\gamma)>0$ if $0<\gamma<\gamma_2^T(z)$, or $\gamma>\gamma_1^T(z)$, and $T(z,\gamma)\le 0$ if $\gamma_2^T(z)\le\gamma\le\gamma_1^T(z)$. Combining this result with $\operatorname{sgn}(\Delta_\xi)=\operatorname{sgn}(T(z,\gamma))$, we have $\Delta_\xi(z,\gamma)>0$ if $0<\gamma<\gamma_2^T(z)$ or $\gamma>\gamma_1^T(z)$, whereas $\Delta_\xi(z,\gamma)\le 0$ if $\gamma_2^T(z)\le\gamma\le\gamma_1^T(z)$.

Case 1: If $\gamma_2^T(z)\le\gamma\le\gamma_1^T(z)$, then $\Delta_\xi(z,\gamma)\le 0$. Since $\xi(x)$ is strictly convex, we have $\xi(x)\ge 0$ for all x. Hence, $\Delta SW\ge 0$.

Case 2: If $0<\gamma<\gamma_2^T(z)$, $\Delta_\xi(z,\gamma)>0$ and $\xi(x)=0$ has two distinct positive roots $y^\pm=\frac{8(1+\gamma)^2}{3+2\gamma}\left[\frac{1}{2[2+(2-z)\gamma]}\pm\sqrt{\Delta_\xi(z,\gamma)}\right]$, where $0<y^-<y^+$. We next show that the interval (y^-,y^+) lies entirely below the feasible range $x_p<x<1$. Evaluating $\xi(x)$ at $x=1-\rho$ gives $\xi(1-\rho)=\frac{\rho^2\left[2+2\gamma\rho(1-\rho)^2-(1-\rho)^4\right]}{8(1+\gamma)^2(1-\rho)^2}>0$. Moreover, $\xi'(1-\rho)=\frac{1+\rho^2-(2\gamma+3)\rho(1+\rho)}{8(1+\gamma)^2}$. We now show that $\xi'(1-\rho)>0$ for $0<\gamma<\gamma_2^T(z)$. Suppose, to the contrary, that $1+\rho^2-(2\gamma+3)\rho(1+\rho)=0$ at some γ in this region. Then, $2\gamma+3=\frac{1+\rho^2}{\rho(1+\rho)}$. Substituting this expression into P yields $P=\frac{2\rho(1+\rho^2)(\rho^3-\rho^2-2)}{(1+\rho)^2}<0$, because $0<\rho<1$. This contradicts $P>0$, which holds for $0<\gamma<\gamma_2^T(z)$. Since the numerator of $\xi'(1-\rho)$ converges to 1 as $\gamma\to 0$, it follows that $\xi'(1-\rho)>0$ throughout this region. Because $\xi(x)$ is strictly convex and has two distinct roots, $\xi(1-\rho)>0$ and $\xi'(1-\rho)>0$ imply $y^+<1-\rho$. Next,

recall that $x_p = \frac{1+\rho}{1+\sqrt{\frac{z\gamma}{1+\gamma}}}$. Since $\rho^2 = \frac{z\gamma}{2+(2-z)\gamma} > \frac{1}{2}\frac{z\gamma}{1+\gamma}$, we have $2\rho > \sqrt{\frac{z\gamma}{1+\gamma}}$. Therefore, $x_p - (1-\rho) = \frac{2\rho-(1-\rho)\sqrt{\frac{z\gamma}{1+\gamma}}}{1+\sqrt{\frac{z\gamma}{1+\gamma}}} > 0$. Hence, $y^+ < 1-\rho < x_p$. Since the feasible range is $x_p < x < 1$, it follows that every feasible x lies to the right of y^+. Therefore, $\xi(x) > 0$ for $x_p < x < 1$, and hence $\Delta SW > 0$ for $0 < \gamma < \gamma_2^T(z)$.

Case 3: If $\gamma > \gamma_1^T(z)$, we also have $\Delta_\xi(z,\gamma) > 0$, and $\xi(x) = 0$ has the two distinct roots y^- and y^+. To incorporate the feasible range $x_p < x < 1$, define $\underline{y} = \min\{1, \max\{y^-, x_p\}\}$, and $\overline{y} = \max\{x_p, \min\{y^+, 1\}\}$. Then, $\Delta SW < 0$ if and only if $\underline{y} < x < \overline{y}$. If $x_p < x \le \underline{y}$ or $\overline{y} \le x < 1$, it follows that $\Delta SW \ge 0$. Finally, solving $x = \frac{1-\gamma F}{1-\gamma F(1-z)}$ for F gives $F = \frac{1-x}{\gamma[1-(1-z)x]}$. Moreover, $\frac{\partial}{\partial x}\left[\frac{1-x}{\gamma[1-(1-z)x]}\right] = -\frac{z}{\gamma[1-(1-z)x]^2} < 0$, so the mapping from x to F is strictly decreasing. Define $\underline{F}_s = \frac{1-\overline{y}(z,\gamma)}{\gamma[1-(1-z)\overline{y}(z,\gamma)]}$, and $\overline{F}_s = \frac{1-\underline{y}(z,\gamma)}{\gamma[1-(1-z)\underline{y}(z,\gamma)]}$. It follows that $\Delta SW < 0$ if and only if $\underline{F}_s < F < \overline{F}_s$. Otherwise, $\Delta SW \ge 0$ if $0 < F \le \underline{F}_s$ or $\overline{F}_s \le F < \frac{1-x_p}{\gamma[1-(1-z)x_p]}$.

Since $0 < \gamma < 1/z$, we define $\hat{\gamma} = \min\{\gamma_1^T(z), 1/z\}$. Combining the three cases, social welfare decreases at $\lambda = \lambda_p^+$ if and only if $\hat{\gamma} < \gamma < 1/z$ and $\underline{F}_s < F < \overline{F}_s$. Otherwise, social welfare weakly increases at $\lambda = \lambda_p^+$ if $0 < \gamma \le \hat{\gamma}$, or if $\hat{\gamma} < \gamma < 1/z$ and either $0 < F \le \underline{F}_s$ or $\overline{F}_s \le F < \frac{1-x_p}{\gamma[1-(1-z)x_p]}$.

iii. Next, for $\lambda_p < \lambda < \lambda_s$, the equilibrium is a costly separating equilibrium. In this region, consumer surplus is given by $CS^s(\lambda) = \frac{[1-\gamma F+S(\lambda)]^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2}$. Taking the first derivative with respect to λ, we obtain $\frac{\partial CS^s(\lambda)}{\partial \lambda} = -\frac{1-\gamma F+S(\lambda)}{4[1+\gamma(1-z)]S(\lambda)} < 0$. Together with the preceding comparison at $\lambda = \lambda_p$, this proves that consumer surplus jumps up at λ_p^+ and strictly decreases over $\lambda_p < \lambda < \lambda_s$. Hence $\sup_{\lambda>0} CS(\lambda) = \lim_{\lambda \downarrow \lambda_p} CS^s(\lambda)$. The supremum is not attained: at $\lambda = \lambda_p$ the LMSE selects pooling (Proposition 3), so $CS(\lambda_p) = CS^p < \lim_{\lambda \downarrow \lambda_p} CS^s(\lambda)$, and for every $\lambda > \lambda_p$, choose $\lambda' \in (\lambda_p, \min\{\lambda, \lambda_s\})$. Then $CS(\lambda') > CS(\lambda)$ by strict decrease on the costly separating interval and constancy on the costless separating interval. Similarly, social welfare can be expressed as $SW^s(\lambda) = CS^s(\lambda) + \frac{1}{2}\left(\pi_H^s + \pi_L^s\right)$. Differentiating with respect to λ yields $\frac{\partial SW^s(\lambda)}{\partial \lambda} = \frac{\{2[1+\gamma(1-z)]-1\}S(\lambda)-\gamma F\{2z[1+\gamma(1-z)]-1\}-1}{4[1+\gamma(1-z)]S(\lambda)}$. The second derivative with respect to λ is given by $\frac{\partial^2 SW^s(\lambda)}{\partial \lambda^2} = -\frac{1+\gamma F\{2z[1+\gamma(1-z)]-1\}}{2S(\lambda)^3} < 0$. Hence, $\frac{\partial SW^s(\lambda)}{\partial \lambda}$ is decreasing in λ. Moreover, we have $\left.\frac{\partial SW^s(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_s} = -\frac{1-\gamma F(1-z)}{4z\gamma F(1+\gamma(1-z))} < 0$. Therefore, if $\left.\frac{\partial SW^s(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_p} \le 0$, it follows that $\frac{\partial SW^s(\lambda)}{\partial \lambda} \le 0$ for all $\lambda_p < \lambda < \lambda_s$. Otherwise, there exists a unique $\lambda^o \in (\lambda_p, \lambda_s)$ such that $\frac{\partial SW^s(\lambda)}{\partial \lambda} = 0$, with $\frac{\partial SW^s(\lambda)}{\partial \lambda} > 0$ for $\lambda_p < \lambda < \lambda^o$ and $\frac{\partial SW^s(\lambda)}{\partial \lambda} \le 0$ for $\lambda^o \le \lambda < \lambda_s$. Evaluating at $\lambda = \lambda_p$, we obtain

$$\left.\frac{\partial SW^s(\lambda)}{\partial \lambda}\right|_{\lambda=\lambda_p} = \frac{\{2[1+\gamma(1-z)]-1\}S(\lambda_p) - \gamma F\{2z[1+\gamma(1-z)]-1\} - 1}{4[1+\gamma(1-z)]S(\lambda_p)}.$$

Since $S(\lambda_p) = z\gamma F + (1-\gamma F(1-z))\rho$, the numerator reduces to $[1-\gamma F(1-z)]\{\rho[2(1+\gamma(1-z))-1]-1\}$. If $\rho[2(1+\gamma(1-z))-1]-1 > 0$, then $\rho^2[2(1+\gamma(1-z))-1]^2 > 1$. Substituting $\rho^2 = \frac{z\gamma}{2[1+\gamma(1-z)]+z\gamma}$, we obtain $\frac{z\gamma}{2[1+\gamma(1-z)]+z\gamma}\{2[1+\gamma(1-z)]-1\}^2 > 1 \Longrightarrow 2z(1-z)\gamma^2 > 1$, which

yields $\gamma > \frac{1}{\sqrt{2z(1-z)}}$. Consequently, if $\min\left\{\frac{1}{\sqrt{2z(1-z)}}, \bar{\gamma}_B\right\} < \gamma < \bar{\gamma}_B$, then $\frac{\partial SW^s(\lambda)}{\partial \lambda} \geq 0$ for $\lambda_p < \lambda \leq \lambda^o$ and $\frac{\partial SW^s(\lambda)}{\partial \lambda} < 0$ for $\lambda^o < \lambda < \lambda_s$. Otherwise, if $0 < \gamma \leq \min\left\{\frac{1}{\sqrt{2z(1-z)}}, \bar{\gamma}_B\right\}$, then $\frac{\partial SW^s(\lambda)}{\partial \lambda} < 0$ for $\lambda_p < \lambda < \lambda_s$, so that, as for consumer surplus, the welfare supremum over the separating regimes is $\lim_{\lambda \downarrow \lambda_p} SW^s(\lambda)$ and is not attained.

Finally, we obtain λ^o in closed form and interpret it. Write $a = 1 + \gamma(1-z)$. The first-order condition $\frac{\partial SW^s(\lambda)}{\partial \lambda} = 0$ reads $(2a-1)S(\lambda^o) = 1 + \gamma F(2za-1)$, so $S(\lambda^o) = \frac{1+\gamma F(2za-1)}{2a-1}$, and since $S(\lambda)^2 = \frac{z\gamma(1-\gamma F)^2}{1+\gamma} - 4a\lambda$,

$$\lambda^o = \frac{1}{4a}\left[\frac{z\gamma(1-\gamma F)^2}{1+\gamma} - \left(\frac{1+\gamma F(2za-1)}{2a-1}\right)^2\right].$$

At this penalty the high type's customer base is $N^s_{c,H}(\lambda^o) = \frac{1-\gamma F+S(\lambda^o)}{2a} = \frac{(1-\gamma F)(2a-1)+1+\gamma F(2za-1)}{2a(2a-1)} = \frac{2a[1-\gamma F(1-z)]}{2a(2a-1)} = \frac{1-\gamma F(1-z)}{1+2\gamma(1-z)} = q^W_H$, the maximizer of $W_H(q)$ in (17). Since the low type's outcome does not depend on λ in the separating regimes and $SW^s(\lambda) = \frac{1}{2}[W_H(N^s_{c,H}(\lambda)) + W_L(N^m_{c,L})]$ with $N^s_{c,H}(\lambda)$ strictly decreasing in λ, welfare increases in λ exactly while $N^s_{c,H}(\lambda) > q^W_H$; the condition $\gamma > 1/\sqrt{2z(1-z)}$ is equivalent to $N^s_{c,H}(\lambda^+_p) = \frac{[1-\gamma F(1-z)](1+\rho)}{2a} > q^W_H$, i.e., to $\rho(2a-1) > 1$. This also shows directly that SW^s is strictly concave in λ on (λ_p, λ_s).

iv. We now establish the global welfare optimum, retaining the assumption $\lambda_p > 0$. Define $\gamma^W = 1/\sqrt{2z(1-z)}$. The polynomial analysis in part ii places γ^W strictly between $\gamma^T_2(z)$ and $\gamma^T_1(z)$. To verify this, substituting $2z(1-z)(\gamma^W)^2 = 1$ gives

$$2z(1-z)^2 T(z, \gamma^W) = z\left[-200 + 81z + 73z^2 + 2\gamma^W(-104z^3 + 284z^2 - 152z - 32)\right] < 0.$$

Indeed, $-200 + 81z + 73z^2 < -46$, and

$$-104z^3 + 284z^2 - 152z - 32 = -4\left[8(1-z)^3 + 62z(1-z)^2 + 29z^2(1-z) + z^3\right] < 0.$$

It follows that the welfare jump is strictly positive whenever $0 < \gamma \leq \gamma^W$. For $\gamma < \gamma^T_2$, this is Case 2 of part ii. At $\gamma = \gamma^T_2$, continuity of that argument places the double root of ξ no higher than $1 - \rho < x_p$, so $\xi(x) > 0$ throughout the feasible interval $x_p < x < 1$. For $\gamma^T_2 < \gamma \leq \gamma^W < \gamma^T_1$, the discriminant of ξ is strictly negative and its leading coefficient is positive, again giving $\xi(x) > 0$. Combining this strict upward jump with part iii, which shows strict decrease throughout costly separation for $\gamma \leq \gamma^W$, proves that the global welfare supremum in this case is approached as $\lambda \downarrow \lambda_p$ and is not attained.

Suppose now that $\gamma > \gamma^W$. Part iii establishes that $\lambda^o \in (\lambda_p, \lambda_s)$ uniquely maximizes welfare over the separating regimes. It remains to compare this value with pooling welfare. Write $a = 1 + \gamma(1-z)$,

$b = 1+\gamma$, $d = a+b$, $t = b-a = z\gamma > 0$, $A = 1-\gamma F(1-z) > 0$ and $x = \frac{1-\gamma F}{A}$. Pooling demand is A/d, so

$$SW^p = \frac{A^2}{2d^2} + \frac{A(1-\gamma F)}{2d}.$$

At λ^o, the high type serves its welfare-optimal quantity $A/(2a-1)$, while the low type serves its monopoly quantity $(1-\gamma F)/(2b)$. Thus, by (17),

$$SW^s(\lambda^o) = \frac{A^2}{4(2a-1)} + \frac{(2b+1)(1-\gamma F)^2}{16b^2}.$$

Completing the square yields

$$\frac{SW^s(\lambda^o) - SW^p}{A^2} = \frac{2b+1}{16b^2}\left(x - \frac{4b^2}{(2b+1)d}\right)^2 + \frac{2+4(b+1)t+(2b+1)t^2}{4d^2(2a-1)(2b+1)} > 0.$$

The strict inequality holds because $a > 1$, $b > 1$, and $t > 0$. Pooling welfare is therefore strictly below the separating maximum, independently of the sign of the welfare jump at λ_p^+. Since welfare is constant in the pooling and costless separating regimes and strictly concave in the costly separating regime, λ^o is the unique global welfare-maximizing penalty. This also proves that pooling is never welfare-optimal when $\lambda_p > 0$. □

Proof of Corollary 1. Under the pooling equilibrium, consumer surplus is $CS^p = \frac{[1-\gamma F(1-z)]^2}{2[2+\gamma(2-z)]^2}$, and social welfare is $SW^p = \frac{[1-\gamma F(1-z)]^2}{2[2+\gamma(2-z)]^2} + \frac{(1-\gamma F)[1-\gamma F(1-z)]}{2[2+\gamma(2-z)]}$. Under the costless separating equilibrium, consumer surplus is $CS^s = \frac{[1-\gamma F(1-z)]^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2}$, and social welfare is $SW^s = \frac{[1-\gamma F(1-z)]^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2} + \frac{[1-\gamma F(1-z)]^2}{8[1+\gamma(1-z)]} + \frac{(1-\gamma F)^2}{8(1+\gamma)}$.

i. Subtracting pooling consumer surplus from costless-separating consumer surplus gives

$$CS^s - CS^p = \frac{[1-\gamma F(1-z)]^2}{16[1+\gamma(1-z)]^2} + \frac{(1-\gamma F)^2}{16(1+\gamma)^2} - \frac{[1-\gamma F(1-z)]^2}{2[2+\gamma(2-z)]^2}.$$

Consequently, $CS^s \geq CS^p$ is equivalent to $\frac{(1-\gamma F)^2}{(1+\gamma)^2} \geq [1-\gamma F(1-z)]^2\left\{\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right\}$. Because both expressions are strictly positive, it follows that $8[1+\gamma(1-z)]^2 \leq [2+\gamma(2-z)]^2 \Longleftrightarrow 2\sqrt{2}\left(1-\frac{z\gamma}{1+\gamma}\right) \leq 2 - \frac{z\gamma}{1+\gamma} \Longleftrightarrow \left[(2\sqrt{2}-1)z - 2(\sqrt{2}-1)\right]\gamma \geq 2(\sqrt{2}-1)$. Therefore, define

$$\bar{\gamma}_c = \begin{cases} \dfrac{2(\sqrt{2}-1)}{(2\sqrt{2}-1)z - 2(\sqrt{2}-1)}, & z > \dfrac{2(\sqrt{2}-1)}{2\sqrt{2}-1}, \\ +\infty, & 0 < z \leq \dfrac{2(\sqrt{2}-1)}{2\sqrt{2}-1}. \end{cases}$$

Then, it follows that $8[1+\gamma(1-z)]^2 \leq [2+\gamma(2-z)]^2 \Longleftrightarrow \gamma \geq \bar{\gamma}_c$. In this region, the inequality $\frac{(1-\gamma F)^2}{(1+\gamma)^2} > [1-\gamma F(1-z)]^2\left\{\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right\}$ always holds. Hence, it follows that $CS^s > CS^p$ if $\gamma \geq \bar{\gamma}_c$.

In all remaining cases, $\gamma < \bar{\gamma}_c$. Expanding the consumer-surplus condition as a quadratic function of F gives

$$\begin{aligned} K(F) = \gamma^2 &\left[\frac{1}{(1+\gamma)^2} - (1-z)^2\left(\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right)\right] F^2 \\ &+ 2\gamma\left[(1-z)\left(\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right) - \frac{1}{(1+\gamma)^2}\right] F \\ &+ \frac{1}{(1+\gamma)^2} - \frac{8}{[2+\gamma(2-z)]^2} + \frac{1}{[1+\gamma(1-z)]^2}. \end{aligned}$$

Thus, $CS^s \geq CS^p$ if and only if $K(F) \geq 0$.

We next establish the properties of this quadratic. Notice that $\frac{z\gamma}{1+\gamma} \in (0,1)$, and hence we have $1-(1+\gamma)^2\left[\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right] = \frac{\left(\frac{z\gamma}{1+\gamma}\right)^2\left[\left(\frac{z\gamma}{1+\gamma}\right)^2 - 6\left(\frac{z\gamma}{1+\gamma}\right)+6\right]}{\left(2-\frac{z\gamma}{1+\gamma}\right)^2\left(1-\frac{z\gamma}{1+\gamma}\right)^2} > 0$. It follows that the coefficient on F^2 in $K(F)$ is strictly positive because $\frac{1}{(1+\gamma)^2} - (1-z)^2\left[\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right] > \frac{1}{(1+\gamma)^2} - \left[\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right] > 0$. Hence, $K(F)$ is strictly convex in F. Evaluating it at $F=0$ and $F=\frac{1}{\gamma}$ gives $K(0) = \frac{1}{(1+\gamma)^2} - \frac{8}{[2+\gamma(2-z)]^2} + \frac{1}{[1+\gamma(1-z)]^2} > 0$ and $K(\frac{1}{\gamma}) = -z^2\left[\frac{8}{[2+\gamma(2-z)]^2} - \frac{1}{[1+\gamma(1-z)]^2}\right] < 0$. Therefore, the quadratic has two distinct real roots. Its smaller root lies uniquely in $(0, 1/\gamma)$, whereas its larger root is strictly greater than $1/\gamma$. Define F_c as this smaller root, namely, the unique solution to $K(F_c) = 0$ in $(0, 1/\gamma)$. Since the model imposes $0 < F < 1/\gamma$, it follows that $CS^s > CS^p$ if $0 < F < F_c$, and otherwise $CS^s \leq CS^p$ if $F_c \leq F < 1/\gamma$. Thus, within the feasible range $0 < \gamma < 1/z$, pooling yields higher consumer surplus than costless separation if and only if $0 < \gamma < \gamma_c$, where $\gamma_c = \min\{\bar{\gamma}_c, 1/z\}$, and $F_c < F < 1/\gamma$. This proves part (i).

ii. Subtracting SW^p from SW^s and collecting terms with respect to F, we can obtain $W(F) = A_W F^2 + B_W F + C_W$, where $A_W = \frac{\gamma^2}{16}\left[\frac{3+2\gamma}{(1+\gamma)^2} + \frac{(1-z)^2[3+2\gamma(1-z)]}{[1+\gamma(1-z)]^2} - \frac{8(1-z)[3-z+\gamma(2-z)]}{[2+\gamma(2-z)]^2}\right]$, $B_W = \frac{\gamma}{8}\left[\frac{4\left(6-4z+\gamma(2-z)^2\right)}{(2+\gamma(2-z))^2} - \frac{3+2\gamma}{(1+\gamma)^2} - \frac{(1-z)[3+2\gamma(1-z)]}{[1+\gamma(1-z)]^2}\right]$ and $C_W = \frac{1}{16}\left[\frac{3+2\gamma}{(1+\gamma)^2} + \frac{3+2\gamma(1-z)}{[1+\gamma(1-z)]^2} - \frac{8[3+\gamma(2-z)]}{[2+\gamma(2-z)]^2}\right]$.
To characterize the sign of $W(F)$, first note that

$$W(F) - (CS^s - CS^p) = \frac{\gamma^2 z^2\left\{1 + 2F + [2+\gamma(2-z)+\gamma^2(1-z)]\,F^2\right\}}{8(1+\gamma)[1+\gamma(1-z)][2+\gamma(2-z)]} > 0$$

Therefore, $W(F) > CS^s - CS^p$. As established above, $CS^s \geq CS^p$ if $\gamma \geq \gamma_c$, or if $0 < \gamma < \gamma_c$ and $0 < F \leq F_c$. Therefore, $W(F) > 0$ if $\gamma \geq \gamma_c$, or if $0 < \gamma < \gamma_c$ and $0 < F \leq F_c$. Hence, costless separation yields strictly higher social welfare than pooling throughout these regions.

We next consider the region $0 < \gamma < \gamma_c$ and $F_c < F < 1/\gamma$, where $CS^s < CS^p$. Part i also establishes that the coefficient on F^2 in $CS^s - CS^p$ is strictly positive. The coefficient on F^2 in $W(F) - (CS^s - CS^p)$ is $\frac{\gamma^2 z^2[2+\gamma(2-z)+\gamma^2(1-z)]}{8(1+\gamma)[1+\gamma(1-z)][2+\gamma(2-z)]} > 0$. Consequently, $A_W > 0$, so $W(F)$ is strictly convex in F. It follows that $W\left(\frac{1}{\gamma}\right) = z^2[\frac{1}{16[1+\gamma(1-z)]^2} + \frac{1}{8[1+\gamma(1-z)]} - \frac{1}{2[2+\gamma(2-z)]^2}]$. Since $2+\gamma(2-z) = 2[1+\gamma(1-z)] + z\gamma > 2[1+\gamma(1-z)]$, we have $\frac{1}{2[2+\gamma(2-z)]^2} < \frac{1}{8[1+\gamma(1-z)]^2}$. Therefore, $W\left(\frac{1}{\gamma}\right) > z^2\left[-\frac{1}{16[1+\gamma(1-z)]^2} + \frac{1}{8[1+\gamma(1-z)]}\right] = \frac{z^2[1+2\gamma(1-z)]}{16[1+\gamma(1-z)]^2} > 0$. Moreover, $W(0) = C_W > 0$. The unconstrained

global minimum of $W(F)$ is attained at $F = F_w^o = -\frac{B_W}{2A_W}$ and equals $W(F_w^o) = -\frac{B_W^2 - 4A_W C_W}{4A_W}$. Directly substituting the expressions for A_W, B_W, and C_W and collecting terms, we obtain $B_W^2 - 4A_W C_W = \frac{\gamma^2 z^2}{64(1+\gamma)^2[1+\gamma(1-z)]^2[2+\gamma(2-z)]^2} Q_W(z,\gamma)$, where $Q_W(z,\gamma) = 4z^2(z-1)\gamma^4 + 2z(3z^2-2z-4)\gamma^3 + (7z^2-28z+4)\gamma^2 + 4(2-5z)\gamma + 4$. Let $\Delta_W = B_W^2 - 4A_W C_W$. Thus, we have $\mathrm{sgn}(\Delta_W) = \mathrm{sgn}(Q_W(z,\gamma))$.

We then characterize the sign of $Q_W(z,\gamma)$. First, $Q_W(z,0) = 4 > 0$, and $Q_W\left(z, \frac{1}{z}\right) = 4z^2(z-1)\frac{1}{z^4} + 2z(3z^2-2z-4)\frac{1}{z^3} + (7z^2-28z+4)\frac{1}{z^2} + 4(2-5z)\frac{1}{z} + 4 = -\frac{3z^2+20z+8}{z^2} < 0$. Therefore, $Q_W(z,\gamma) = 0$ has at least one root in $(0, 1/z)$. To establish uniqueness, arrange the coefficients of Q_W in descending powers of γ: $\omega_4 = 4z^2(z-1)$, $\omega_3 = 2z(3z^2-2z-4)$, $\omega_2 = 7z^2-28z+4$, $\omega_1 = 4(2-5z)$ and $\omega_0 = 4$. For $0 < z < 1$, ω_4 and ω_3 are strictly negative, whereas ω_0 is strictly positive. The coefficient ω_2 changes sign at $z_c = 2 - \frac{2\sqrt{42}}{7}$, whereas ω_1 changes sign at $z = 2/5$. Since $0 < z_c < 2/5$, the signs of the nonzero coefficients are

z interval	$(\omega_4, \omega_3, \omega_2, \omega_1, \omega_0)$
$(0, z_c)$	$-,-,+,+,+$
$(z_c, 2/5)$	$-,-,-,+,+$
$(2/5, 1)$	$-,-,-,-,+$

with zero coefficients omitted at the two boundary values. Each sequence has exactly one sign change. By Descartes' rule of signs, $Q_W(z,\gamma) = 0$ has at most one positive root. Combining this result with the existence argument above, it has exactly one positive root in $(0, 1/z)$. Define this root as $\gamma_w(z)$, so that $Q_W(z, \gamma_w(z)) = 0$, and $0 < \gamma_w(z) < \frac{1}{z}$. It follows that $Q_W(z,\gamma) \geq 0$ if $0 < \gamma \leq \gamma_w(z)$ and $Q_W(z,\gamma) < 0$ if $\gamma_w(z) < \gamma < 1/z$.

Next, we show that $\gamma_w(z) < \bar{\gamma}_c$. If $\bar{\gamma}_c = +\infty$, the result is immediate. Otherwise, $z > \frac{2(\sqrt{2}-1)}{2\sqrt{2}-1}$. Substituting $\bar{\gamma}_c$ into $Q_W(z,\gamma)$ gives $Q_W(z, \bar{\gamma}_c) = -\frac{32z^3\left[(5\sqrt{2}-6)z - (\sqrt{2}-1)\right]}{\left[(2\sqrt{2}-1)z - 2(\sqrt{2}-1)\right]^4}$. Moreover, $(5\sqrt{2}-6)z - (\sqrt{2}-1) > (5\sqrt{2}-6)\frac{2(\sqrt{2}-1)}{2\sqrt{2}-1} - (\sqrt{2}-1) = 5\sqrt{2} - 7 > 0$. Hence, $Q_W(z, \bar{\gamma}_c) < 0$. Since $\gamma_w(z)$ is the unique positive root of $Q_W(z,\gamma) = 0$, it follows that $\gamma_w(z) < \min\{\bar{\gamma}_c, 1/z\} = \gamma_c$. Therefore, for $\gamma_w(z) \leq \gamma < \gamma_c$, we have $Q_W(z,\gamma) \leq 0$ and hence $\Delta_W \leq 0$. Since $A_W > 0$, $W(F) \geq 0$ for all F.

It remains to consider $0 < \gamma < \gamma_w(z)$. In this region, $\Delta_W > 0$, and hence the global minimum of $W(F)$ is strictly negative. Moreover, its minimizer $F_w^o = -\frac{B_W}{2A_W}$ lies in $(0, 1/\gamma)$. To see this, note that $B_W < 0$ throughout $0 < \gamma < \gamma_w(z)$: as $\gamma \to 0$, $B_W/\gamma \to -z/8 < 0$, and B_W cannot cross zero because $B_W = 0$ would imply $\Delta_W = -4A_W C_W < 0$, contradicting $\Delta_W > 0$. Thus, $B_W < 0$ and $F_w^o > 0$. We next show that $F_w^o < 1/\gamma$. As $\gamma \to 0$, $\gamma F_w^o = -\frac{\gamma B_W}{2A_W} \to \frac{1}{2+z} < 1$. Hence, $F_w^o < 1/\gamma$ for sufficiently small $\gamma > 0$. By continuity, F_w^o cannot cross the upper boundary $1/\gamma$ for $0 < \gamma < \gamma_w(z)$. Otherwise, there would exist some $\gamma \in (0, \gamma_w(z))$ such that $F_w^o = 1/\gamma$. At this point, $W(F_w^o) = W(1/\gamma) > 0$, contradicting $W(F_w^o) = -\frac{\Delta_W}{4A_W} < 0$. Therefore, $0 < F_w^o < \frac{1}{\gamma}$ for all $0 < \gamma < \gamma_w(z)$.

Moreover, since $CS^s \geq CS^p$ for $0 < F \leq F_c$ and $W(F) - (CS^s - CS^p) > 0$, we have $W(F) > 0$ for $0 < F \leq F_c$. Thus, $F_c < F_w^o$. Together with $W(F_w^o) < 0$ and $W(1/\gamma) > 0$, the strict convexity of

$W(F)$ implies that $W(F)=0$ has exactly two roots in the feasible interval: one in (F_c, F_w^o) and the other in $(F_w^o, 1/\gamma)$. Denote these two roots by $\underline{F}_w$ and $\overline{F}_w$. They are given by $\underline{F}_w = \frac{-B_W - \sqrt{\Delta_W}}{2A_W}$ and $\overline{F}_w = \frac{-B_W + \sqrt{\Delta_W}}{2A_W}$ with $F_c < \underline{F}_w < F_w^o < \overline{F}_w < \frac{1}{\gamma}$. Because $A_W > 0$, $W(F) < 0$ if and only if $\underline{F}_w < F < \overline{F}_w$. Thus, pooling yields higher social welfare if and only if $0 < \gamma < \gamma_w(z)$ and $\underline{F}_w < F < \overline{F}_w$. Otherwise, social welfare is weakly higher under costless separation, with equality at feasible roots of $W(F)=0$. This proves part (ii). □

Proof of Proposition 9. From the proof of Proposition 8, we have established that at $\lambda = \lambda_p$, the right-hand limit of consumer surplus under costly separation is strictly greater than consumer surplus under pooling, i.e., $\Delta CS = \lim_{\lambda \to \lambda_p^+} CS^s(\lambda) - CS^p(\lambda_p) > 0$. Because the condition $\lambda_p(\bar{\gamma}_2) = \lambda$ holds exactly at the threshold $\bar{\gamma}_2$, we can evaluate the discontinuous change in consumer surplus as γ crosses $\bar{\gamma}_2$ from the pooling equilibrium to the costly separating equilibrium

$$\lim_{\gamma \to \bar{\gamma}_2^+} CS^s(\lambda, \gamma) - CS^p(\bar{\gamma}_2) = \lim_{\lambda \to \lambda_p(\bar{\gamma}_2)^+} CS^s(\lambda) - CS^p(\lambda_p(\bar{\gamma}_2)) > 0$$

Hence, the consumer surplus increases discontinuously at $\gamma = \bar{\gamma}_2^+$.

Similarly, the proof of Proposition 8 establishes that, at the boundary $\lambda = \lambda_p$, the welfare difference $\Delta SW = \lim_{\lambda \to \lambda_p^+} SW^s(\lambda) - SW^p(\lambda_p) \geq 0$ if $0 < \gamma \leq \hat{\gamma}$, or if $\gamma > \hat{\gamma}$ and either $0 < F \leq \underline{F}_s$ or $F \geq \overline{F}_s$. Otherwise, $\Delta SW < 0$ if $\gamma > \hat{\gamma}$ and $\underline{F}_s < F < \overline{F}_s$. At $\gamma = \bar{\gamma}_2$, we have $\lambda = \lambda_p(\bar{\gamma}_2)$. Hence, as γ crosses $\bar{\gamma}_2$ from the pooling equilibrium to the costly separating equilibrium, it follows that

$$\lim_{\gamma \to \bar{\gamma}_2^+} SW^s(\lambda, \gamma) - SW^p(\bar{\gamma}_2) = \lim_{\lambda \to \lambda_p(\bar{\gamma}_2)^+} SW^s(\lambda) - SW^p(\lambda_p(\bar{\gamma}_2)).$$

Therefore, social welfare weakly increases at $\bar{\gamma}_2^+$ if $\bar{\gamma}_2 \leq \hat{\gamma}$, or if $\bar{\gamma}_2 > \hat{\gamma}$ and either $0 < F \leq \underline{F}_s(\bar{\gamma}_2)$ or $F \geq \overline{F}_s(\bar{\gamma}_2)$. It decreases if $\bar{\gamma}_2 > \hat{\gamma}$ and $\underline{F}_s(\bar{\gamma}_2) < F < \overline{F}_s(\bar{\gamma}_2)$. This proves part (i).

Next, we prove that when the equilibrium transitions from the costless separating equilibrium to the costly separating equilibrium, both consumer surplus and social welfare are strictly increasing in γ at $\bar{\gamma}_0^+$, where $\bar{\gamma}_0$ is the smaller root of $\lambda_s(\gamma) = \lambda$.

The total consumer surplus is given by $CS = \frac{1}{2}(CS_H + CS_L)$ where $CS_H(p(\gamma), \gamma) = \frac{1}{2}\left(\frac{1-p(\gamma)}{1+\gamma(1-z)}\right)^2$ and $CS_L(p(\gamma), \gamma) = \frac{1}{2}\left(\frac{1-p(\gamma)}{1+\gamma}\right)^2$. Under the costly separating equilibrium, the equilibrium prices are $p_H^s(\gamma) = \frac{1+\gamma F - S(\gamma)}{2}$ and $p_L^s(\gamma) = \frac{1+\gamma F}{2}$, with $S(\gamma) = \sqrt{\frac{z\gamma(1-\gamma F)^2}{1+\gamma} - 4(1+\gamma(1-z))\lambda}$. Therefore, by applying the chain rule for the total derivative, we obtain

$$\left.\frac{dCS_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \left.\frac{\partial CS_H}{\partial p}\right|_{\bar{\gamma}_0} \cdot \left.\frac{dp_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} + \left.\frac{\partial CS_H}{\partial \gamma}\right|_{\bar{\gamma}_0}$$

where $\left.\frac{\partial CS_H}{\partial p}\right|_{\bar{\gamma}_0} = -\frac{1-\bar{\gamma}_0 F(1-z)}{2(1+\bar{\gamma}_0(1-z))^2}$, $\left.\frac{dp_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{F - S'(\bar{\gamma}_0^+)}{2}$, and $\left.\frac{\partial CS_H}{\partial \gamma}\right|_{\bar{\gamma}_0} = -\frac{(1-z)(1-\bar{\gamma}_0 F(1-z))^2}{4(1+\bar{\gamma}_0(1-z))^3}$

To determine $S'(\bar{\gamma}_0^+)$, let $D(\gamma) = 1+\gamma(1-z)$. At the threshold connecting the costless and costly separating regimes, the boundary penalty condition satisfies

$$\lambda_s(\gamma) = \frac{1}{4(1+\gamma(1-z))}\left[\frac{z\gamma(1-\gamma F)^2}{1+\gamma} - z^2\gamma^2F^2\right].$$

Rearranging this equation yields $\frac{z\gamma(1-\gamma F)^2}{1+\gamma} = 4D(\gamma)\lambda_s(\gamma) + z^2\gamma^2F^2$. Substituting this back into the expression for $S(\gamma)^2$, we get $S(\gamma)^2 = 4D(\gamma)[\lambda_s(\gamma)-\lambda] + z^2\gamma^2F^2$. Taking the derivative with respect to γ on both sides yields

$$2S(\gamma)S'(\gamma) = 4D'(\gamma)[\lambda_s(\gamma)-\lambda] + 4D(\gamma)\lambda_s'(\gamma) + 2z^2\gamma F^2.$$

Evaluating the above expression as $\gamma \to \bar{\gamma}_0^+$, we utilize the binding constraints $\lambda = \lambda_s(\bar{\gamma}_0)$ and $S(\bar{\gamma}_0) = z\bar{\gamma}_0F$, which simplifies the equation to

$$S'(\bar{\gamma}_0^+) = \frac{2D(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{z\bar{\gamma}_0F} + zF.$$

It follows that $\left.\frac{dp_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{F(1-z)}{2} - \frac{D(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{z\bar{\gamma}_0F}$. Consequently, substituting this into the total derivative for CS_H^s, we obtain

$$\left.\frac{dCS_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{(1-\bar{\gamma}_0F(1-z))\lambda_s'(\bar{\gamma}_0)}{2D(\bar{\gamma}_0)z\bar{\gamma}_0F} - \frac{(1-\bar{\gamma}_0F(1-z))F(1-z)}{4D(\bar{\gamma}_0)^2} - \frac{(1-\bar{\gamma}_0F(1-z))^2(1-z)}{4D(\bar{\gamma}_0)^3}.$$

Substituting $p_L^s(\gamma) = \frac{1+\gamma F}{2}$, we obtain $\left.\frac{dCS_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = -\frac{(1-\bar{\gamma}_0F)(1+F)}{4(1+\bar{\gamma}_0)^3}$. Define $\Omega_{CS} = \left.\frac{dCS_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} + \left.\frac{dCS_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+}$, and we have

$$\Omega_{CS} = \frac{E(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{2z\bar{\gamma}_0FD(\bar{\gamma}_0)} - \left[\frac{E(\bar{\gamma}_0)F(1-z)}{4D(\bar{\gamma}_0)^2} + \frac{E(\bar{\gamma}_0)^2(1-z)}{4D(\bar{\gamma}_0)^3} + \frac{(1-\bar{\gamma}_0F)(1+F)}{4(1+\bar{\gamma}_0)^3}\right]$$

where $E(\bar{\gamma}_0) = 1-\bar{\gamma}_0F(1-z)$.

To establish that $\left.\frac{dCS^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{\Omega_{CS}}{2}$ is strictly positive, consider a sufficiently small penalty parameter $\lambda > 0$. Recall that $\lambda_s(\gamma)$ is unimodal and initially strictly increases from $\lambda_s(0) = 0$. This ensures that $\lambda_s'(\bar{\gamma}_0) > 0$, and as $\lambda \to 0^+$, it mathematically implies that the threshold $\bar{\gamma}_0 \to 0^+$. Define $\Omega_1 = \frac{E(\bar{\gamma}_0)F(1-z)}{4D(\bar{\gamma}_0)^2} + \frac{E(\bar{\gamma}_0)^2(1-z)}{4D(\bar{\gamma}_0)^3} + \frac{(1-\bar{\gamma}_0F)(1+F)}{4(1+\bar{\gamma}_0)^3}$. Taking the limit as $\bar{\gamma}_0 \to 0^+$, we obtain $\lim_{\bar{\gamma}_0\to 0^+}\frac{E(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{2z\bar{\gamma}_0FD(\bar{\gamma}_0)} = +\infty$, and $\lim_{\bar{\gamma}_0\to 0^+}\Omega_1 = \frac{(1+F)(2-z)}{4} < +\infty$.

Therefore, provided that the penalty λ is sufficiently small (and correspondingly, $\bar{\gamma}_0$ is sufficiently close to zero), we can obtain $\left.\frac{dCS^s}{d\gamma}\right|_{\bar{\gamma}_0^+} > 0$. Hence there exists $\lambda_{CS}^+ > 0$ such that, for each fixed $0 < \lambda < \lambda_{CS}^+$, continuity gives $\epsilon_{CS}(\lambda) > 0$ for which CS^s is increasing on $[\bar{\gamma}_0(\lambda), \bar{\gamma}_0(\lambda)+\epsilon_{CS}(\lambda)]$.

The total social welfare is given by $SW = \frac{1}{2}(CS_H + \pi_H + CS_L + \pi_L)$ where $\pi_H(p(\gamma),\gamma) = \frac{[p(\gamma)-\gamma F(1-z)][1-p(\gamma)]}{1+\gamma(1-z)}$ and $\pi_L(p(\gamma),\gamma) = \frac{[p(\gamma)-\gamma F][1-p(\gamma)]}{1+\gamma}$. Similarly, applying the chain rule for the total derivative, we have

$$\left.\frac{d\pi_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \left.\frac{\partial\pi_L}{\partial p}\right|_{\bar{\gamma}_0} \cdot \left.\frac{dp_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+} + \left.\frac{\partial\pi_L}{\partial\gamma}\right|_{\bar{\gamma}_0}$$

$$\left.\frac{d\pi_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \left.\frac{\partial \pi_H}{\partial p}\right|_{\bar{\gamma}_0} \cdot \left.\frac{dp_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} + \left.\frac{\partial \pi_H}{\partial \gamma}\right|_{\bar{\gamma}_0}$$

Notice that at the threshold $\gamma = \bar{\gamma}_0$, the equilibrium prices exactly coincide with the unconstrained monopoly optimal prices $p_H^s(\bar{\gamma}_0) = \frac{1+\bar{\gamma}_0 F(1-z)}{2}$ and $p_L^s(\bar{\gamma}_0) = \frac{1+\bar{\gamma}_0 F}{2}$. By the Envelope Theorem, since the firm optimally sets prices to maximize per-type profits at this continuous limit, the first-order partial derivatives of profit with respect to price evaluated at this point are zero, i.e., $\left.\frac{\partial \pi_H}{\partial p}\right|_{\bar{\gamma}_0} = 0$ and $\left.\frac{\partial \pi_L}{\partial p}\right|_{\bar{\gamma}_0} = 0$. Therefore, it follows that

$$\left.\frac{d\pi_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \left.\frac{\partial \pi_L}{\partial \gamma}\right|_{\bar{\gamma}_0} = -\frac{(1-p_L^s)(F+p_L^s)}{(1+\bar{\gamma}_0)^2}$$

$$\left.\frac{d\pi_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \left.\frac{\partial \pi_H}{\partial \gamma}\right|_{\bar{\gamma}_0} = -\frac{(1-p_H^s)(1-z)(F+p_H^s)}{D(\bar{\gamma}_0)^2}$$

Substituting the prices $p_H^s(\bar{\gamma}_0)$ and $p_L^s(\bar{\gamma}_0)$, we obtain

$$\left.\frac{d\pi_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = -\frac{(1-\bar{\gamma}_0 F)[1+(2+\bar{\gamma}_0)F]}{4(1+\bar{\gamma}_0)^2}$$

$$\left.\frac{d\pi_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = -\frac{(1-z)(1-\bar{\gamma}_0 F(1-z))}{2D(\bar{\gamma}_0)} \cdot \left[\frac{(1-\bar{\gamma}_0 F(1-z))}{2D(\bar{\gamma}_0)} + F\right]$$

Define the sum of these marginal firm profits as $\Omega_2 = \left.\frac{d\pi_H^s}{d\gamma}\right|_{\bar{\gamma}_0^+} + \left.\frac{d\pi_L^s}{d\gamma}\right|_{\bar{\gamma}_0^+}$. We obtain $\left.\frac{dSW^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{1}{2}(\Omega_{CS} + \Omega_2)$. As established above, when λ is sufficiently low, we have $\bar{\gamma}_0 \to 0^+$. Taking this limit for Ω_2, we obtain $\lim_{\bar{\gamma}_0 \to 0^+} \Omega_2 = -\frac{(2-z)(1+2F)}{4}$. Moreover, $\lim_{\bar{\gamma}_0 \to 0^+} \frac{E(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{2z\bar{\gamma}_0 F D(\bar{\gamma}_0)} = +\infty$, while $\lim_{\bar{\gamma}_0 \to 0^+}(\Omega_1 - \Omega_2) = \frac{(2+3F)(2-z)}{4}$. It follows that $\lim_{\bar{\gamma}_0 \to 0^+} \left.\frac{dSW^s}{d\gamma}\right|_{\bar{\gamma}_0^+} = \frac{1}{2}\lim_{\bar{\gamma}_0 \to 0^+}\left[\frac{E(\bar{\gamma}_0)\lambda_s'(\bar{\gamma}_0)}{2z\bar{\gamma}_0 F D(\bar{\gamma}_0)} - (\Omega_1 - \Omega_2)\right] = +\infty$. Hence, there exists $\lambda_{SW}^+ > 0$ such that, for each fixed $0 < \lambda < \lambda_{SW}^+$, continuity gives $\epsilon_{SW}(\lambda) > 0$ for which SW^s is increasing on $[\bar{\gamma}_0(\lambda), \bar{\gamma}_0(\lambda) + \epsilon_{SW}(\lambda)]$. Take $\lambda^+ = \min\{\lambda_{CS}^+, \lambda_{SW}^+\}$. For each fixed $0 < \lambda < \lambda^+$, choose $\epsilon(\lambda) = \min\{\epsilon_{CS}(\lambda), \epsilon_{SW}(\lambda), [\bar{\gamma}_1(\lambda) - \bar{\gamma}_0(\lambda)]/2\} > 0$. The resulting interval remains below the next regime transition, proving part (ii). □

Proof of Proposition 10. The price optimization and equilibrium comparison follow the arguments for Propositions 1, 2, and 3. We state the modified payoffs and candidates, and prove the additional feasibility and local comparisons required here.

i. Equilibrium candidates. The modified continuation is $N_c(p,\beta) = \frac{1-p}{1+\gamma(1-\beta z)^2}$ and $N_h(p,\beta) = \frac{\gamma(1-\beta z)(1-p)}{1+\gamma(1-\beta z)^2}$, so that

$$\pi_i(r,p \mid \beta) = \frac{[p - \gamma(1-\beta z)(1-z_i)F](1-p)}{1+\gamma(1-\beta z)^2} - \lambda \cdot \mathbf{1}\{i = L,\ r = r_H\}.$$

As in Proposition 1, use Bayesian beliefs on the equilibrium path and $\beta = 0$ elsewhere. Since a higher belief raises positive operating profit, the relevant truthful deviation bounds are $\pi_L^s = \frac{(1-\gamma F)^2}{4(1+\gamma)}$ and $\pi_{H,dev} = \frac{[1-\gamma(1-z)F]^2}{4(1+\gamma)}$.

Under separation, the low type charges $(1+\gamma F)/2$ and earns π_L^s. The high type's unconstrained price is $p_{H,u}^s = [1+\gamma(1-z)^2 F]/2$, and its low-type deterrence threshold is

$$\lambda_s^d(\gamma) = \frac{1}{4}\left\{\frac{[1-\gamma F(1-z^2)][1-\gamma F(1-z)^2]}{1+\gamma(1-z)^2} - \frac{(1-\gamma F)^2}{1+\gamma}\right\} > 0.$$

For the sign, set $t = \gamma F \in (0,1)$. The common numerator is strictly concave in t, with positive endpoint values $\gamma z(2-z)$ and $(1+\gamma)z^3(2-z)$. The high-type-optimal candidate charges $p_{H,u}^s$ for $\lambda \geq \lambda_s^d$; otherwise, its price $p_H^s(\lambda)$ is the lower solution of $\pi_L(r_H, p \mid 1) = \pi_L^s$. The lower root is preferred because

$$\pi_H(r_H, p \mid 1) - \pi_L(r_H, p \mid 1) = \lambda + \frac{\gamma z(1-z)F(1-p)}{1+\gamma(1-z)^2}.$$

Write its payoff as $\pi_H^s(\lambda, \gamma)$, whether or not it is feasible. As λ rises on the costly branch, the lower root increases toward $p_{H,u}^s$, so strict concavity makes π_H^s continuous and strictly increasing; thereafter it is constant and strictly above $\pi_{H,dev}$. Thus separation is feasible exactly when $\lambda \geq \lambda_e^d$, where $\lambda_e^d = \inf\{\lambda \geq 0 : \pi_H^s(\lambda, \gamma) \geq \pi_{H,dev}\}$ and $0 \leq \lambda_e^d < \lambda_s^d$.

For pooling, let $p_{H,u}^p = \frac{1+\gamma(1-z/2)(1-z)F}{2}$, and $p_{L,m}^p = \frac{1+\gamma(1-z/2)F}{2}$, and define the necessary low-type feasibility bound $\lambda_L^d = \frac{[1-\gamma(1-z/2)F]^2}{4[1+\gamma(1-z/2)^2]} - \pi_L^s > 0$. The strict sign follows by comparison with π_L^s. For part (i), the additional condition $F \leq 4(1-z)$ implies $\pi_H(r_H, p_{L,m}^p \mid 1/2) > \pi_{H,dev}$ and $\pi_L^s + \lambda_L^d + \frac{\gamma z(1-z)F[1-\gamma(1-z)^2 F]}{2[1+\gamma(1-z)^2]} > \pi_{H,dev}$. These two inequalities are verified below. The first makes $p_{L,m}^p$ feasible throughout $0 \leq \lambda \leq \lambda_L^d$, which is therefore exactly the pooling-existence interval. As in Proposition 3, the best pooling price is $p^p(\lambda) = \max\{p_{H,u}^p, p_{L,1}^p(\lambda)\}$, where $p_{L,1}^p$ is the lower solution of $\pi_L(r_H, p \mid 1/2) = \pi_L^s$. This price lies between $p_{H,u}^p$ and $p_{L,m}^p$, so the first inequality ensures that the high type's constraint is slack. Its payoff $\pi_H^p(\lambda, \gamma)$ is continuous and nonincreasing in λ.

The second inequality gives $\lambda_e^d < \lambda_L^d$. If $\lambda_L^d \geq \lambda_s^d$, this follows from costless separation. Otherwise $p_H^s(\lambda_L^d) < p_{H,u}^s$, and the binding low-type constraint and the displayed payoff difference imply

$$\pi_H^s(\lambda_L^d, \gamma) > \pi_L^s + \lambda_L^d + \frac{\gamma z(1-z)F[1-\gamma(1-z)^2 F]}{2[1+\gamma(1-z)^2]} > \pi_{H,dev}.$$

Continuity then gives strict overlap. Hence at least one PBE in the maintained class exists for every positive penalty.

ii. Verification of the sufficient condition. For this algebraic verification only, set $\zeta = 1-z \in (0,1)$ and $t = \gamma F \in (0,1)$; the condition becomes $0 < F \leq 4\zeta$. Direct simplification gives

$$\pi_H(r_H, p_{L,m}^p \mid 1/2) - \pi_{H,dev}$$

$$= \frac{tF(1-\zeta)}{4(F+t)[4F+t(1+\zeta)^2]}\Big\{3+\zeta+4F\zeta \\ -[F(1+\zeta+2\zeta^2)+2\zeta+2\zeta^2]t-(1-\zeta)(1+\zeta)^2t^2\Big\}.$$

The polynomial in braces is strictly decreasing in $t\in[0,1]$. At $t=1$, it equals $(1-\zeta)[2-\zeta^2+F(2\zeta-1)]$. This is positive if $\zeta\geq 1/2$. If $\zeta<1/2$, then $F\leq 4\zeta$ implies $2-\zeta^2+F(2\zeta-1)\geq 2-4\zeta+7\zeta^2>0$. This proves the first strict inequality. For the second, the difference between its left-hand side and $\pi_{H,dev}$ simplifies to $\frac{tF(1-\zeta)}{4(F+t)(F+t\zeta^2)[4F+t(1+\zeta)^2]}\,I(F)$ and $I(F)=c_0+c_1F+c_2F^2$, where the coefficients, used only in this verification, are

$$\begin{aligned} c_0 &= \zeta^2(3+\zeta)t(1-t)^2+2\zeta(1+3\zeta-\zeta^2-\zeta^3)t^2(1-t) \\ &\quad +\zeta(1-\zeta)(2+6\zeta+4\zeta^2+\zeta^3)t^3, \\ c_1 &= (3+\zeta)(1-t)^2+2(1+3\zeta-\zeta^2+\zeta^3)t(1-t) \\ &\quad +2\zeta(1-\zeta)(4+5\zeta+3\zeta^2+\zeta^3)t^2, \\ c_2 &= -4+8\zeta+(1+3\zeta-8\zeta^3)t. \end{aligned}$$

Here $c_0,c_1>0$. If $\zeta\geq 1/2$, the endpoint values of the linear function $c_2(t)$ are $4(2\zeta-1)$ and $(1-\zeta)(-3+8\zeta+8\zeta^2)$, both nonnegative, so $I(F)>0$. The same conclusion holds if $\zeta<1/2$ and $c_2\geq 0$.

It remains to consider $\zeta<1/2$ and $c_2<0$. Then I is concave in F, so it is enough to check $I(0)>0$ and $I(4\zeta)>0$. The first is $c_0>0$. For the second, write

$$\frac{I(4\zeta)}{\zeta}=d_0(1-t)^3+3d_1t(1-t)^2+3d_2t^2(1-t)+d_3t^3,$$

where $d_0=4(3-15\zeta+32\zeta^2)$, $d_1=\frac{20-145\zeta+425\zeta^2+8\zeta^3-128\zeta^4}{3}$, $d_2=\frac{2(5-49\zeta+239\zeta^2-5\zeta^3-136\zeta^4-4\zeta^5)}{3}$, and $d_3=(1-\zeta)(2-10\zeta+172\zeta^2+153\zeta^3+8\zeta^4)$. For $0<\zeta<1/2$, $d_1\geq\frac{20-145\zeta+393\zeta^2}{3}>0$ and $d_2\geq\frac{2(5-49\zeta+202\zeta^2)}{3}>0$. Also $d_0>0$ and $d_3\geq(1-\zeta)(2-10\zeta+172\zeta^2)>0$. Each quadratic used in these bounds has a positive leading coefficient and a negative discriminant. Thus all four Bernstein coefficients are positive, so $I(4\zeta)>0$. Concavity yields $I(F)>0$ for $0<F\leq 4\zeta$, completing the verification of the second inequality.

iii. LMSE selection. The comparison in Proposition 3 applies once feasibility is established: costless separation strictly dominates every pooling outcome, while on the costly branch π_H^s strictly increases and π_H^p is nonincreasing. Define $\lambda_p^d(\gamma)=\sup\left(\{0\}\cup\{\lambda\in[0,\lambda_L^d]:\pi_H^p(\lambda,\gamma)\geq\pi_H^s(\lambda,\gamma)\}\right)$. Then $0\leq\lambda_p^d<\lambda_s^d$. Below a positive cutoff, pooling is strictly preferred. Above it, separation is feasible: if pooling still exists, its payoff is above $\pi_{H,dev}$ and below the separating candidate's payoff; otherwise $\lambda>\lambda_L^d>\lambda_e^d$. Strict concavity gives the unique selected outcome away from the cutoff.

At a positive cutoff, continuity and the low type's pooling constraint make pooling an LMSE outcome. Separation is also selected exactly when both types receive the same payoffs. The cutoff may coincide with λ_L^d, so a transition need not be an equality of high-type payoffs. This proves part (i).

iv. Local comparative statics without the additional condition. For part (ii), retain only the original domain $0<\gamma<\min\{1/F,4\}$. First, $\lambda_s^d(0)=0$ and $(\lambda_s^d)'(0)=\frac{z(2+2F-z)}{4}>0$. Thus, for each fixed positive penalty, sufficiently low γ makes costless separation feasible and strictly preferred to pooling.

Next, set $\lambda=0$ and consider small positive γ. At the unconstrained pooling price,

$$\pi_L(r_H,p_{H,u}^p\mid 1/2)-\pi_L^s=\frac{z(4-z+4F)}{16}\gamma+O(\gamma^2)>0.$$

The high type's unconstrained pooling maximum also strictly exceeds $\pi_{H,dev}$, so this pooling price is feasible. Moreover,

$$\pi_H^p(0,\gamma)-\pi_H^s(0,\gamma)=\frac{z[4F(1+z)+4-z]}{16}\gamma+O(\gamma^{3/2})>0.$$

Here the separating candidate payoff is an upper bound on all separating PBE payoffs, whether or not that candidate is itself feasible. Choose a compact interval $[\gamma_a,\gamma_b]$ within this small-γ region, with $0<\gamma_a<\gamma_b$. Continuity preserves $\lambda<\lambda_s^d(\gamma)$, the strict pooling constraints, and the payoff advantage uniformly on this interval for all sufficiently small positive λ. Since $(\lambda_s^d)'(0)>0$, the local boundary $\underline{\gamma}(\lambda)$ solving $\lambda_s^d(\underline{\gamma})=\lambda$ converges to zero as $\lambda\to 0^+$. Hence this fixed pooling interval lies strictly to the right of the low-threat costless separating region for sufficiently small λ.

Finally, let $F\geq 1/4$ and take $\gamma\to(1/F)^-$ at $\lambda=0$. The low type's binding pooling price converges to $1-z/2$, exceeding $p_{H,u}^p$ by $z(1-z)/4$ in the limit. The candidate separating payoff, selected pooling payoff, and deviation bound satisfy $\pi_H^s(0,\gamma)\to\frac{Fz^2(1-z)}{F+(1-z)^2}$, and $\pi_H^p(0,\gamma)\to\frac{Fz^2(2-z)}{4F+(2-z)^2}>\frac{Fz^2}{4(1+F)}=\lim_{\gamma\to(1/F)^-}\pi_{H,dev}$. The strict inequality follows because its numerator after taking a common denominator contains the positive factor $4F(1-z)+4-z^2$. Also, $\lambda_L^d(\gamma)\to\frac{Fz^2}{4[4F+(2-z)^2]}>0$, and $\lambda_s^d(\gamma)\to\frac{Fz^3(2-z)}{4[F+(1-z)^2]}>0$. Thus the low type's pooling discriminant remains positive near $\lambda=0$, and the high type's constraint is strictly slack at the selected binding price. By continuity, pooling remains feasible at that price for sufficiently small positive penalties and γ sufficiently close to $1/F$.

The limiting candidate payoff difference is

$$\lim_{\gamma\to(1/F)^-}[\pi_H^s(0,\gamma)-\pi_H^p(0,\gamma)]=\frac{Fz^2[(1-z)(2-z)+F(2-3z)]}{[F+(1-z)^2][4F+(2-z)^2]}.$$

The numerator factor $(1-z)(2-z)+F(2-3z)$ is strictly decreasing in $z\in(0,1)$ and has its unique zero at $z_d(F)$. For $z<z_d(F)$, the separating candidate strictly outperforms the feasible

pooling outcome, which itself strictly exceeds the high type's deviation bound. The separating candidate therefore satisfies the high type's constraint and is selected. For $z > z_d(F)$, pooling strictly outperforms the best separating candidate and is selected regardless of that candidate's feasibility. For each fixed F and $z \neq z_d(F)$, these strict inequalities persist for all sufficiently small positive λ and all γ sufficiently close to $1/F$ from below. The argument includes $F = 1/4$, where the two admissible upper bounds coincide at $1/F = 4$; the limits are always taken from within the admissible domain. This proves part (ii). □

Proof of Proposition 11. The continuation is unchanged from Proposition 1. It is convenient to write operating profit in terms of demand $n = N_c$, with the associated price $p = 1 - [1 + \gamma(1 - \beta z)]n$:

$$\Pi_i(n \mid \beta) = [1 - \gamma(1 - z_i)F]n - [1 + \gamma(1 - \beta z) + \gamma\theta(1 - z_i)]n^2.$$

The low type additionally pays λ when claiming r_H. Use Bayesian beliefs on the path and $\beta = 0$ at every other claim–price pair, as in Propositions 1 and 2. The best truthful payoffs under these off-path beliefs are $\pi_L^s = \frac{(1-\gamma F)^2}{4[1+\gamma(1+\theta)]}$ and $\pi_{H,dev} = \frac{[1-\gamma(1-z)F]^2}{4[1+\gamma+\gamma\theta(1-z)]}$. All displayed quantities below correspond to prices in $[0, 1]$; positive operating profit implies $0 < n < 1/[1 + \gamma(1 - \beta z)]$. Throughout this proof, write $j = \sqrt{\frac{\gamma z}{1+\gamma(1+\theta)}}$.

i. Separation. The low type earns π_L^s at its monopoly price. The high type's unconstrained demand is $n_{H,u}^s = \frac{1-\gamma(1-z)F}{2[1+\gamma(1-z)(1+\theta)]}$, and its deterrence threshold is

$$\begin{aligned}\lambda_s^b(\gamma) &= \Pi_L(n_{H,u}^s \mid 1) - \pi_L^s \\ &= \frac{j^2 G^b(\gamma)}{4[1 + \gamma(1 - z)(1 + \theta)]^2},\end{aligned}$$

where $G^b(\gamma) = 1 + [(1 + \theta) - z(1 + \theta)^2 - 2F(1 + \theta z) - F^2 z]\gamma + F(1 - z)[F(1 - z) - 2(1 + \theta)]\gamma^2 + F^2(1 + \theta)(1 - z)^2\gamma^3$. Here, $G^b(0) = 1$ and $G^b(1/F) = -\frac{z(F+1+\theta)(F+1+\theta-z)}{F} < 0$. The leading coefficient is positive. Descartes' rule allows at most two positive roots; the sign at $1/F$ and the positive limit at infinity give one on either side of $1/F$. Hence exactly one root γ_s^b lies in $(0, 1/F)$, and λ_s^b is positive on $(0, \gamma_s^b)$ and negative on $(\gamma_s^b, 1/F)$.

When $\lambda \geq \lambda_s^b$, choose $n_{H,u}^s$. Otherwise, choose the larger solution $n_{H,r}^s$ of $\Pi_L(n \mid 1) = \pi_L^s + \lambda$. This is the same quadratic-constraint argument as in Proposition 1: the larger-demand root is preferred because

$$\Pi_H(n \mid 1) - \Pi_L(n \mid 1) = \gamma z(Fn + \theta n^2)$$

is strictly increasing in n. The unconstrained high-type demand exceeds the low type's maximizing demand under $\beta = 1$, so the binding choice satisfies $n_{H,r}^s > n_{H,u}^s$ and requires a lower price.

It remains to check the high type's deviation constraint. Costless separation strictly exceeds $\pi_{H,dev}$ because its quadratic coefficient is smaller than under $\beta = 0$. On the costly branch, relaxing

the low-type constraint moves demand toward $n^s_{H,u}$ and strictly raises the high type's payoff, so it suffices to check $\lambda = 0$. The binding demand at zero penalty is $n^s_{H,r}(0) = \frac{1-\gamma F}{2[1+\gamma(1+\theta)](1-j)}$. The costly condition $n^s_{H,r}(0) > n^s_{H,u}$ gives $\gamma z F(1-j) < (1-\gamma F)j[1-(1+\theta)j]$, and hence $(1+\theta)j < 1$. Direct expansion and collection yield $4[1+\gamma(1+\theta)]\big[\Pi_H(n^s_{H,r}(0) \mid 1) - \pi_{H,dev}\big] = \frac{\theta(1-\gamma F)^2 j^3[2-(1+\theta)j]}{(1-j)^2(1-\theta j^2)} + \frac{\gamma z F}{(1-j)(1-\theta j^2)}\big\{2(1-\gamma F)j(1-\theta j) - \gamma z F(1-j)\big\}$. All denominators are positive and the first term is nonnegative. By the costly condition, the expression in braces strictly exceeds $(1-\gamma F)j[1+(1-\theta)j] > 0$. The last bracket is positive because $\theta j < 1-j$. Thus $\Pi_H(n^s_{H,r}(0) \mid 1) > \pi_{H,dev}$, proving the required constraint. The stated separating regimes follow, with prices recovered from $p = 1-[1+\gamma(1-z)]n$.

ii. Pooling. As in Proposition 2, it suffices to intersect the two types' incentive-compatible intervals. The high type's constraint $\Pi_H(n \mid 1/2) \geq \pi_{H,dev}$ gives the nonempty interval $[n_{H,-}, n_{H,+}]$, where

$$n_{H,\pm} = \frac{1-\gamma(1-z)F}{2[1+\gamma(1-z/2)+\gamma\theta(1-z)]}\left[1 \pm \frac{j}{\sqrt{2(1-\theta j^2)}}\right].$$

The low type's gross pooling profit is maximized at $n_{L,m} = \dfrac{1-\gamma F}{2[1+\gamma(1-z/2)+\gamma\theta]}$, which is below the high type's maximizing demand and hence below $n_{H,+}$. Therefore its best demand within the high-type interval is $n^p_* = \max\{n_{L,m}, n_{H,-}\}$ and $\lambda^b_1(\gamma) = \Pi_L(n^p_* \mid 1/2) - \pi^s_L$. The two weak incentive constraints are feasible exactly when $0 < \lambda \leq \lambda^b_1(\gamma)$. For $0 < \lambda < \lambda^b_1(\gamma)$, the candidate n^p_* gives L strictly more than π^s_L, so it also satisfies the same-price tie-break. At $\lambda = \lambda^b_1(\gamma) > 0$, strict concavity leaves the unique candidate demand n^p_*, at which L earns π^s_L. Under the specified beliefs, the tie-break excludes this candidate if and only if its price equals the low type's symmetric-information price, that is, if

$$1-[1+\gamma(1-z/2)]n^p_* = 1 - \frac{(1+\gamma)(1-\gamma F)}{2[1+\gamma(1+\theta)]}.$$

At any other price, L's payoff from truthful disclosure under belief zero is strictly below π^s_L, so the boundary candidate is supported. For $\theta = 0$, this qualification reduces to the boundary restriction in Proposition 2.

The threshold is continuous in γ. We now show that it has exactly one zero on $(0, 1/F)$. At zero penalty, the low type's constraint gives the interval $[n_{L,-}, n_{L,+}]$, where

$$n_{L,\pm} = \frac{1-\gamma F}{2[1+\gamma(1-z/2)+\gamma\theta]}\left(1 \pm \frac{j}{\sqrt{2}}\right).$$

Since $n_{L,-} < n_{L,m} < n_{H,+}$, the sign of $\lambda^b_1(\gamma)$ is the sign of $n_{L,+} - n_{H,-}$: strict overlap gives a strictly positive penalty bound, tangency gives zero, and disjoint intervals give a negative bound.

The variable j increases strictly with γ and satisfies $0 < j < \sqrt{\frac{z}{1+\theta+F}}$ and $\gamma = \frac{j^2}{z-(1+\theta)j^2}$. Substituting the demand bounds into $n_{H,-} \leq n_{L,+}$ and simplifying gives the equivalent condition

$$\frac{zFj(\sqrt{2}-j)}{z-(1+\theta+F)j^2} \leq 1 - \sqrt{2}\theta j + \sqrt{1-\theta j^2}.$$

All denominators and square roots are positive on the stated interval. The derivative of the left-hand side with respect to j is

$$\frac{\sqrt{2}zF\,[z-\sqrt{2}zj+(1+\theta+F)j^2]}{[z-(1+\theta+F)j^2]^2}>0.$$

The quadratic in brackets has positive leading coefficient and discriminant $2z[z-2(1+\theta+F)]<0$. The right-hand side is nonincreasing, with derivative $-\sqrt{2}\theta-\theta j/\sqrt{1-\theta j^2}\le 0$. As $\gamma\to 0^+$, $j\to 0$ and the two sides tend to zero and two, respectively. As $\gamma\to(1/F)^-$, $j\to\sqrt{z/(1+\theta+F)}$ and the left-hand side tends to $+\infty$ while the right-hand side remains finite. They therefore cross exactly once, defining a unique threshold $\gamma_p^b\in(0,1/F)$ such that $\lambda_1^b(\gamma)>0$ for $0<\gamma<\gamma_p^b$, $\lambda_1^b(\gamma_p^b)=0$, and $\lambda_1^b(\gamma)<0$ for $\gamma_p^b<\gamma<1/F$. Thus, under the specified off-path beliefs, pooling exists for $0<\gamma<\gamma_p^b$ and $0<\lambda<\lambda_1^b(\gamma)$; the equality case is included exactly when it satisfies the preceding same-price qualification. Since $F\ge z$, this range of γ is admissible, proving part (ii). □

Proof of Proposition A1. With exogenous conditional attack intensity $\bar{N}_h$, demand under belief β is $q^e(p,\beta)=[1-p-(1-\beta z)\bar{N}_h]^+$, which is zero at prices above $1-(1-\beta z)\bar{N}_h$, so such deviations yield gross revenue zero and are never profitable; type i's profit is $\pi_i^e(r,p\mid\beta)=p\,q^e(p,\beta)-(1-z_i)\bar{N}_hF-\lambda\mathbf{1}\{i=L,r=r_H\}$. Maximizing $p\,q^e(p,\beta)$ gives the price $[1-(1-\beta z)\bar{N}_h]/2$ and the maximized gross revenue $V(\beta)=[1-(1-\beta z)\bar{N}_h]^2/4$.

i. For any (r,p,β), $\pi_H^e-\pi_L^e=z\bar{N}_hF+\lambda\mathbf{1}\{r=r_H\}$, which does not depend on p. Under symmetric information $\beta=1$ for the high type and $\beta=0$ for the low type, so $p_H^{e,m}=[1-(1-z)\bar{N}_h]/2>[1-\bar{N}_h]/2=p_L^{e,m}$. In a separating equilibrium supported by the belief $\beta=0$ off the path, the low type earns $V(0)-\bar{N}_hF$ truthfully, so its incentive constraint at the high type's pair (r_H,p_H^s) is $p_H^s\,q^e(p_H^s,1)-\lambda\le V(0)$, while the high type's payoff is $p_H^s\,q^e(p_H^s,1)-(1-z)\bar{N}_hF$ and its best deviation under $\beta=0$ yields $V(0)-(1-z)\bar{N}_hF$. If $\lambda\ge V(1)-V(0)$ the constraint holds at $p_H^{e,m}$ and separation is costless. If $\lambda<V(1)-V(0)$, the high type maximizes $p\,q^e(p,1)$ subject to $p\,q^e(p,1)\le V(0)+\lambda$; since $p\,q^e(p,1)=V(1)-(p-p_H^{e,m})^2$ is symmetric around $p_H^{e,m}$, the constraint binds at the two prices $p_H^{e,m}\pm\sqrt{V(1)-V(0)-\lambda}$, which yield the same gross revenue $V(0)+\lambda$ and hence the same profit, and the high type's rent relative to the low type's reservation payoff $V(0)-\bar{N}_hF$ equals $\lambda+z\bar{N}_hF$, independently of its customer base.

ii. In a pooling equilibrium at a common price p with belief α, the high type's best pooling payoff is $V(\alpha)-(1-z)\bar{N}_hF$ at $p=[1-(1-\alpha z)\bar{N}_h]/2$, and the low type's incentive constraint at that price is $V(\alpha)-\lambda\ge V(0)$, i.e., $\lambda\le V(\alpha)-V(0)$. The high type prefers this pooling outcome to its best separating outcome, whose gross revenue is $\min\{V(1),V(0)+\lambda\}$, if and only if $V(\alpha)\ge V(0)+\lambda$, the same condition. Hence the LMSE is pooling if $\lambda<V(\alpha)-V(0)$, costly separating if $V(\alpha)-V(0)<\lambda<V(1)-V(0)$, and costless separating if $\lambda\ge V(1)-V(0)$. At $\lambda=V(\alpha)-V(0)$

the high type's pooling and separating payoffs coincide and so do the low type's, which equal $V(0) - \bar{N}_h F$ in both outcomes; the lexicographic ordering therefore does not separate the two outcomes, and the proposition assigns the boundary to pooling by convention (unlike in the main model, where the low type strictly prefers pooling at $\lambda = \lambda_p$). Neither $V(\alpha) - V(0)$ nor $V(1) - V(0)$ involves F. Direct computation gives $V(\alpha) - V(0) = \alpha z \bar{N}_h[2 - (2 - \alpha z)\bar{N}_h]/4$, a concave quadratic in $\bar{N}_h$ with maximum at $\bar{N}_h = 1/(2 - \alpha z) \in (1/2, 1)$; this is the inverted-U shape noted in the text.

iii. For a known type serving q, price is $1 - q - (1 - z_i)\bar{N}_h$, consumer surplus is $q^2/2$, and profit is $q[1 - q - (1 - z_i)\bar{N}_h] - (1 - z_i)\bar{N}_h F$, so net surplus is $[1 - (1 - z_i)\bar{N}_h]q - q^2/2 - (1 - z_i)\bar{N}_h F$, which is increasing in q for $q < 1 - (1 - z_i)\bar{N}_h = 2q_i^{e,m}$. In the costly separating regime the low type serves $q_L^{e,m}$, and the high type serves $q^e(p_H^{e,m} \mp \sqrt{V(1) - V(0) - \lambda}, 1) < q^e(0, 1) = 2q_H^{e,m}$ because both separating prices are positive ($\sqrt{V(1) - V(0)} < \sqrt{V(1)} = p_H^{e,m}$). Every customer base that arises in that regime therefore lies below the level at which net surplus peaks, so a change in λ, which moves only the high type's customer base, moves consumer surplus and welfare in the same direction: when the downward price is played both decrease in λ and both suprema over the regime are approached as λ decreases to $V(\alpha) - V(0)$; when the upward price is played both increase in λ. Thus, along each price branch, consumer surplus and social welfare approach their highest values at the same penalty boundary. Finally, the sign of the change in consumer surplus at the transition from pooling to costly separation depends on which price is played: for λ close to $V(\alpha) - V(0)$, the high type's customer base exceeds its pooling customer base at the downward price and falls short of it at the upward price, while the low type's customer base is smaller than under pooling in both cases. Since the LMSE ranks outcomes by the firm's profits, it is silent on the choice between the two prices, which is the indeterminacy noted in the text. □

Proof of Proposition A2. i. Throughout this part, consider the baseline model and the maintained class in which the high type claims r_H. Write $c_H = \gamma(1 - z)F$, $c_L = \gamma F$, $a = 1 + \gamma(1 - z)$ and $b_h = 1 + \gamma(1 - z/2)$. Both equilibrium payoffs are positive: a type can claim r_L at its optimal price under belief zero and obtain a positive payoff under any belief. A deviation with $p \le c_i$ is therefore unprofitable for type i. For the remaining prices, type i's payoff is increasing in the belief, so belief one gives its largest deviation payoff.

Every pooling price is excluded. Let p^0 be any pooling PBE price, π_i^0 its payoffs, and $N_c^0 = (1 - p^0)/b_h$. The high type's payoff under belief one is $(p - c_H)(1 - p)/a$. At $p = p^0$ it strictly exceeds π_H^0, whereas it is zero at $p = c_H$. Its lower indifference root p_- thus satisfies $c_H < p_- < p^0$ and $p_- < p_H^m$, with $\pi_H(r_H, p_- \mid 1) = \pi_H^0$ and $N_c(p_-, 1) > N_c^0$. By (8), $\pi_L(r_H, p_- \mid 1) = \pi_H^0 - \lambda - z\gamma F N_c(p_-, 1) < \pi_H^0 - \lambda - z\gamma F N_c^0 = \pi_L^0$. Choose $p' = p_- + \epsilon$ with $\epsilon > 0$ sufficiently small. The high type strictly gains under belief one, while the low type remains equilibrium-dominated: if $p' > c_L$, its largest deviation payoff occurs under belief one and is below π_L^0; if $p' \le c_L$, every belief gives a nonpositive deviation

payoff. The Intuitive Criterion therefore assigns probability one to H and breaks the pooling PBE. D1 does so as well. This argument covers every pooling price, not only p_H^m or the LMSE pooling region.

Other separating prices are excluded. In any separating PBE in the maintained class, L claims r_L and charges p_L^m: if it instead claimed r_H, changing only its claim would save the penalty without worsening its belief; its optimal truthful payoff is π_L^m. Let p^* denote the unique high-type-optimal separating price of Proposition 1. It satisfies $p^* \le p_H^m < p_L^m$. Any other separating price $p^0 \neq p^*$ gives H strictly less than $\pi_H(r_H, p^* \mid 1)$. Choose $p' < p^*$ sufficiently close to p^*. Since $(p - c_L)(1 - p)$ is strictly increasing below p_L^m, L's deterrence constraint is strict at p', while H's payoff there still exceeds its payoff at p^0. The same positive-margin argument makes L equilibrium-dominated under every belief. The Intuitive Criterion assigns the deviation to H, eliminating every such separating outcome.

The optimal costly separating outcome survives. Let $K_H = \pi_H^{sep}$ and suppose $0 < \lambda < \lambda_s(\gamma)$. The low type's constraint binds at p^*, so $K_H \frac{p^* - c_L}{p^* - c_H} = \pi_L^m + \lambda$ and $p^* > c_L$. If H can weakly gain at an off-path pair (r, p) under some belief β, then $p \ge p^*$: at prices below p^* even belief one gives H less than K_H. Moreover, $\pi_L(r, p \mid \beta) \ge K_H \frac{p - c_L}{p - c_H} - \lambda \mathbf{1}\{r = r_H\}$. The ratio $(p - c_L)/(p - c_H)$ is strictly increasing in p. Thus for $p > p^*$ the right-hand side strictly exceeds π_L^m for either claim. At $p = p^*$ the only off-path claim is r_L, and L gains by $\lambda > 0$. Consequently, whenever H weakly gains, L strictly gains. Continuity makes L's gain set strictly larger, since H's constraint against every belief-zero deviation is strictly slack by Proposition 1. D1 therefore places weight only on L at these pairs. At pairs where H cannot weakly gain, L is not eliminated in favor of H, so belief zero is admissible as well. The pessimistic off-path beliefs hence deter all deviations and satisfy D1 and the Intuitive Criterion. This argument covers prices on both sides of p_H^m and both claims.

The costless outcome survives, including boundary cases. If $\lambda \ge \lambda_s(\gamma)$, H obtains its global maximum under belief one, so it has no strictly profitable deviation under any belief. The only off-path pair at which H can even be indifferent is (r_L, p_H^m) under belief one. At that pair L's payoff minus π_L^m equals $\lambda_s(\gamma)$. If $\lambda_s(\gamma) > 0$, L strictly gains whenever H is indifferent, so belief zero satisfies both criteria. If $\lambda_s(\gamma) \le 0$, assign belief one to this single off-path pair: H is indifferent and L does not gain, and neither criterion eliminates H there. At every other off-path pair, H cannot weakly gain and belief zero is admissible. These beliefs support the same costless separating outcome, including $\lambda = \lambda_s(\gamma) > 0$. Combined with the preceding exclusions, this proves uniqueness of the surviving pure-strategy outcome within the maintained class for every $\lambda > 0$.

ii. Following Mailath et al. (1993), a PBE (σ', μ') *defeats* a PBE (σ, μ) if there is a claim–price pair m that is sent with zero probability under σ and with positive probability under σ' by a nonempty set K of types such that every type in K weakly prefers (σ', μ') to (σ, μ), at least one

type in K strictly prefers it, and the belief $\mu(\cdot \mid m)$ is not the restriction of the prior to K (with the standard qualification for types in K that are indifferent); a PBE is *undefeated* if no PBE defeats it. Let (σ, μ) be the LMSE equilibrium and (σ', μ') any other pure-strategy PBE in which the high type claims r_H. Suppose first that $\lambda \le \lambda_p(\gamma)$, so that (σ, μ) is the pooling equilibrium at $\bar{p}_H$. If (σ', μ') is separating, the pairs sent under σ' and not under σ are (r_H, p'_H), sent by the high type alone, and (r_L, p_L^m), sent by the low type alone. For the first, $K = \{H\}$ and the defeat condition requires the high type to strictly prefer (σ', μ'); but its separating profit is at most $\pi_H^{sep}(\lambda) \le \pi_H^{pool}$ by the definition of λ_p. For the second, $K = \{L\}$ and the low type would have to weakly prefer its separating profit π_L^m to π_L^{pool}; but $\pi_L^{pool} > \pi_L^m$ because $\lambda \le \lambda_p < \lambda_2$ (Step i of the proof of Proposition B1). If (σ', μ') is pooling at $p'' \neq \bar{p}_H$, then $K = \{H, L\}$ and the high type strictly prefers $\bar{p}_H$, its unconstrained optimum under belief α. Hence no PBE defeats (σ, μ). Suppose next that $\lambda > \lambda_p(\gamma)$, so that (σ, μ) is the high-type-optimal separating equilibrium. If (σ', μ') is pooling at p'', the pair (r_H, p'') is off the path of σ (it coincides with (r_H, p_H^{sep}) only if $p'' = p_H^{sep}$, in which case it is on the path and is not a candidate), $K = \{H, L\}$, and the high type's pooling profit is at most $\pi_H^{pool} < \pi_H^{sep}(\lambda)$; if (σ', μ') is another separating equilibrium, $K = \{H\}$ at the new high-type pair and the high type's profit there is at most $\pi_H^{sep}(\lambda)$, so it cannot strictly prefer (σ', μ'). Hence (σ, μ) is undefeated.

iii. In the specified semi-separating PBE, L mixes between (r_L, p_L^m) and H's pair, so Lemma 2 gives $\pi_H = \pi_L^m + \lambda + z\gamma F N_{c,H}$ with $\beta \in (\alpha, 1)$. When pure separation is costly, the bound established with that lemma gives $N_{c,H} \le N_{c,+}(\beta) < N_{c,+}(1) = N_{c,H}^{sep}$, so H's semi-separating payoff is strictly below π_H^{sep}. When pure separation is costless, H's positive operating margin and $\beta < 1$ imply $\pi_H(r_H, p \mid \beta) < \pi_H(r_H, p \mid 1) \le \pi_H^m$. In either case the semi-separating payoff is strictly below the best pure separating payoff, which is at most H's LMSE payoff. Thus no such semi-separating outcome enters $\mathcal{E}_H$ when this class is added, and the selected outcome is unchanged. □

Proof of Proposition A3. The equilibrium characterization follows the same logic as the proofs of Propositions 1–4, so we omit its full derivation. The LMSE is costless separating when $\lambda \ge \lambda_s$, costly separating when $\lambda_p < \lambda < \lambda_s$, and pooling when $0 < \lambda \le \lambda_p$, where $\lambda_s = \frac{\gamma z[(1-\gamma F)^2 - \gamma z(1+\gamma)F^2]}{4(1+\gamma)[1+\gamma(1-z)]}$ and $\lambda_p(\alpha) = \frac{\gamma z}{4[1+\gamma(1-z)]}\left\{\frac{(1-\gamma F)^2}{1+\gamma} - \left[F\sqrt{\gamma z} + [1-\gamma F(1-z)]\sqrt{\frac{1-\alpha}{1+\gamma(1-\alpha z)}}\right]^2\right\}$. The expression for λ_s is independent of the prior belief α. Differentiating $\lambda_p(\alpha)$ gives

$$\frac{\partial \lambda_p}{\partial \alpha} = \frac{\gamma z[1-\gamma F(1-z)]\left[\gamma z F + [1-\gamma F(1-z)]\sqrt{\frac{\gamma z(1-\alpha)}{1+\gamma(1-\alpha z)}}\right]}{4[1+\gamma(1-\alpha z)]^2\sqrt{\frac{\gamma z(1-\alpha)}{1+\gamma(1-\alpha z)}}} > 0,$$

because $0 < \alpha < 1$ and $\gamma F < 1$. This proves part (i).

For $0 < \lambda < \lambda_s$, the high-type firm's profit under costly separation is $\pi_H^s = \frac{[1-\gamma F(1-z)]^2 - [S(\lambda) - \gamma z F]^2}{4[1+\gamma(1-z)]}$ where $S(\lambda) = \sqrt{\frac{\gamma z(1-\gamma F)^2}{1+\gamma} - 4[1+\gamma(1-z)]\lambda}$. Its pooling profit is $\pi_H^p(\alpha) = [1-\gamma F(1-z)]^2/[4(1+\gamma(1-\alpha z))]$. Consequently, $\partial \pi_H^p / \partial \alpha = \gamma z[1-\gamma F(1-z)]^2/[4(1+\gamma(1-\alpha z))^2] > 0$.

Since $0 < \lambda < \lambda_s$, we have $0 < S(\lambda) - \gamma z F < (1-\gamma F)\sqrt{\gamma z/(1+\gamma)} - \gamma z F < [1-\gamma F(1-z)]\sqrt{\gamma z/(1+\gamma)}$. Substituting this inequality into π_H^s gives $\pi_H^p(0) < \pi_H^s < \lim_{\alpha\to 1^-}\pi_H^p(\alpha)$. By continuity and strict monotonicity of $\pi_H^p(\alpha)$, there is a unique $\alpha_p(\lambda) \in (0,1)$ at which $\pi_H^p(\alpha_p) = \pi_H^s$, given by $\alpha_p(\lambda) = \frac{\gamma z[1-\gamma F(1-z)]^2 - (1+\gamma)[S(\lambda)-\gamma z F]^2}{\gamma z\left\{[1-\gamma F(1-z)]^2 - [S(\lambda)-\gamma z F]^2\right\}}$. At this point, the high-type pooling incentive constraint is satisfied because $\pi_H^p(\alpha_p) = \pi_H^s(\lambda) > [1-\gamma F(1-z)]^2/[4(1+\gamma)]$, where the right-hand side is its deviation payoff. The low-type incentive constraint also holds: its binding constraint under costly separation and high-type indifference imply

$$\pi_L^p(\alpha_p) - \pi_L^s = \frac{\gamma z F[S(\lambda)-\gamma z F][1-\gamma F + S(\lambda)]}{2[1+\gamma(1-z)][1-\gamma F(1-z)]} > 0.$$

Thus, the LMSE selects pooling at α_p. Moreover, $\pi_H^p(\alpha_p) = \pi_H^s(\lambda)$ is equivalent to $\lambda_p(\alpha_p) = \lambda$. Since $\lambda_p(\alpha)$ is strictly increasing in α, we have $\lambda_p(\alpha) < \lambda$ for $\alpha < \alpha_p$ and $\lambda_p(\alpha) \geq \lambda$ for $\alpha \geq \alpha_p$. Therefore, the LMSE is costly separating for $\alpha < \alpha_p$ and pooling for $\alpha \geq \alpha_p$. This proves part (ii). □

Proof of Proposition A4. Under costly separation, the type-specific prices and demands do not depend on α, so both types' profits are independent of α . Under pooling, $p^p = [1+\gamma F(1-z)]/2$ and $N_c^p = [1-\gamma F(1-z)]/[2(1+\gamma(1-\alpha z))]$. Consequently, $\pi_H^p = [1-\gamma F(1-z)]^2/[4(1+\gamma(1-\alpha z))]$ and $\pi_L^p = (p^p - \gamma F)N_c^p - \lambda$. Differentiation gives $\partial\pi_H^p/\partial\alpha = \gamma z[1-\gamma F(1-z)]^2/[4(1+\gamma(1-\alpha z))^2] > 0$ and $\partial\pi_L^p/\partial\alpha = \gamma z[1-\gamma F(1-z)](p^p-\gamma F)/[2(1+\gamma(1-\alpha z))^2] > 0$; the latter inequality holds because $\pi_L^p(\alpha_p) = (p^p-\gamma F)N_c^p(\alpha_p) - \lambda > \pi_L^s > 0$, so $p^p - \gamma F > 0$. At α_p, $\pi_H^p = \pi_H^s$, while $\pi_L^p(\alpha_p) - \pi_L^s = \frac{\gamma z F[S(\lambda)-\gamma z F][1-\gamma F+S(\lambda)]}{2[1+\gamma(1-z)][1-\gamma F(1-z)]} > 0$. This proves part (i).

Under costly separation, $N_{c,H}^s = [1-\gamma F + S(\lambda)]/[2(1+\gamma(1-z))]$ and $N_{c,L}^s = (1-\gamma F)/[2(1+\gamma)]$, with $N_{c,H}^s > N_{c,L}^s$. Hence, $CS^s(\alpha) = [\alpha(N_{c,H}^s)^2 + (1-\alpha)(N_{c,L}^s)^2]/2$ increases strictly in α. Under pooling, $CS^p(\alpha) = [1-\gamma F(1-z)]^2/[8(1+\gamma(1-\alpha z))^2]$, so $\partial CS^p/\partial\alpha = \gamma z[1-\gamma F(1-z)]^2/[4(1+\gamma(1-\alpha z))^3] > 0$.

At the switching point, high-type indifference gives $N_c^p(\alpha_p)/N_{c,H}^s = 1 - [S(\lambda)-\gamma z F]/[1-\gamma F(1-z)]$. Also, $1-\gamma F + S(\lambda) < (1-\gamma F)(1+\sqrt{\gamma z/(1+\gamma)})$ implies $N_{c,L}^s/N_{c,H}^s > 1 - \sqrt{\gamma z/(1+\gamma)}$. Substituting the expression for α_p from the preceding proof and simplifying yields $\frac{S(\lambda)-\gamma z F}{1-\gamma F(1-z)} - (1-\alpha_p)\sqrt{\frac{\gamma z}{1+\gamma}} > 0$, because $0 < [S(\lambda)-\gamma z F]/[1-\gamma F(1-z)] < \sqrt{\gamma z/(1+\gamma)}$. Therefore $\alpha_p N_{c,H}^s + (1-\alpha_p)N_{c,L}^s > N_c^p(\alpha_p)$. By convexity of the square function, $CS^s(\alpha_p) \geq [\alpha_p N_{c,H}^s + (1-\alpha_p)N_{c,L}^s]^2/2 > [N_c^p(\alpha_p)]^2/2 = CS^p(\alpha_p)$. Thus consumer surplus falls strictly when the LMSE switches to pooling.

It remains to compare consumer surplus as $\alpha \to \alpha_p^-$ and as $\alpha \to 1^-$. Direct substitution of α_p yields

$$\lim_{\alpha\to\alpha_p^-} CS^s(\alpha) - \lim_{\alpha\to 1^-} CS^p(\alpha) = \frac{v Q_\alpha(v)}{8\gamma z[1+\gamma(1-z)]^2\left\{[1-\gamma F(1-z)]^2 - [S(\lambda)-\gamma z F]^2\right\}},$$

where $Q_\alpha(v) = 2\gamma z[1-\gamma F(1-z)]^3 + \left[(2\gamma z - 1 - \gamma)[1-\gamma F(1-z)]^2 + \frac{[1+\gamma(1-z)]^3(1-\gamma F)^2}{(1+\gamma)^2}\right]v - 2(1+\gamma)[1-\gamma F(1-z)]v^2 - (1+\gamma)v^3$ and $v = S(\lambda) - \gamma z F$. We have $Q_\alpha(0) = 2\gamma z[1-\gamma F(1-z)]^3 >$

0. Its coefficients in descending powers of v have signs $-,-,-,+$ or $-,-,+,+$, so Descartes' rule of signs implies that Q_α has exactly one positive root. At the algebraic endpoint $v = [1 - \gamma F(1-z)]\sqrt{\gamma z/(1+\gamma)}$, the corresponding expression for α_p equals zero, and the surplus difference becomes $(N^s_{c,L})^2/2 - [1-\gamma F(1-z)]^2/[8(1+\gamma(1-z))^2] < 0$. Hence this unique root lies strictly between zero and that endpoint. Define $\bar{\lambda}_\alpha$ as the unique solution to $Q_\alpha(S(\bar{\lambda}_\alpha) - \gamma z F) = 0$ with $0 < S(\bar{\lambda}_\alpha) - \gamma z F < [1-\gamma F(1-z)]\sqrt{\gamma z/(1+\gamma)}$, and set $\lambda_\alpha = \max\{\bar{\lambda}_\alpha, 0\}$. As λ increases through $(0, \lambda_s)$, $S(\lambda) - \gamma z F$ decreases strictly to zero. Thus, $\lim_{\alpha \to \alpha_p^-} CS^s(\alpha) > \lim_{\alpha \to 1^-} CS^p(\alpha)$ if $\lambda_\alpha < \lambda < \lambda_s$ and $\lim_{\alpha \to \alpha_p^-} CS^s(\alpha) \leq \lim_{\alpha \to 1^-} CS^p(\alpha)$ if $\lambda \leq \lambda_\alpha$. This proves part (ii). □

Online Supplement C. Summary of Notation and Thresholds

This part of the Online Supplement provides a summary of the key notation in the paper (Table A5) and a table collecting the threshold functions used in the main results (Table A6).

C.1. Summary of Notation

Table A5 Summary of Main Notation.

Notation	Description
z_i	Firm's cybersecurity level, $i \in \{L, H\}$
α	Prior probability that the firm is of the high type
β	Posterior probability that the firm is of the high type
r	Firm's cybersecurity claim, $r \in \{r_H, r_L\}$
v	Consumers' valuation for the firm's digital product
p	Firm's price
$\bar{z}$	Expected cybersecurity level of the firm from consumers' perspective
l	Consumer loss from a hacker attack
$\bar{N}_h(r,p)$	Intensity of hacker attacks anticipated by consumers
$N_h(r,p)$	Actual intensity of hacker attacks
c	Hacker's attack cost, $c \sim U[0, \bar{c}]$
h	Hacker's payoff per active consumer of the attacked firm
$\bar{N}_c(r,p)$	Number of active consumers anticipated by hackers
$N_c(r,p)$	Actual number of active consumers of the firm
γ	Hacker payoff-to-cost ratio, $\gamma = h/\bar{c}$
F	Loss incurred by the firm per attack
θ	Degree to which the firm's loss per attack increases with its customer base
λ	Expected penalty for a false high-security claim
U_c	Expected utility of a consumer
U_h	Expected utility of a hacker
π_i	Profit of the type-i firm
CS	Consumer surplus
SW	Social welfare

C.2. Threshold Functions

Table A6 Threshold functions of the baseline model

($\alpha = 1/2$; $a = 1+\gamma(1-z)$, $A = 1-\gamma F(1-z)$, $\rho = \sqrt{z\gamma/(2+(2-z)\gamma)}$).

Symbol	Definition	Role
$\bar{\gamma}_B$	$\min\{1/F, 1/z\}$	Admissible upper bound on γ (positive demand, interior hacker participation)
γ_s	smaller root of $F^2(1-z)\gamma^2 - F(2+zF)\gamma + 1 = 0$	$\lambda_s(\gamma) > 0$ iff $\gamma < \gamma_s$
$\lambda_s(\gamma)$	$\dfrac{z\gamma[(1-\gamma F)^2 - z\gamma(1+\gamma)F^2]}{4(1+\gamma)a}$	Costly-to-costless separating boundary (Propositions 1, 3)
$S(\lambda)$	$\sqrt{[z\gamma(1-\gamma F)^2 - 4(1+\gamma)a\lambda]/(1+\gamma)}$	Costly separating price $p_H^{sep} = (1+\gamma F - S)/2$
$\gamma_1,\ \gamma_p,\ \lambda_1(\gamma)$	given in the proof of Proposition 2	Existence of pooling PBE
$\lambda_2(\gamma)$	$\dfrac{z\gamma[(1-\gamma F)^2 - 2z\gamma(1+\gamma)F^2]}{4(1+\gamma)(2+(2-z)\gamma)}$	Pooling price equals p_H^m iff $\lambda \le \lambda_2$
$\lambda_p(\gamma)$	$\dfrac{1}{4a}\left[\dfrac{z\gamma(1-\gamma F)^2}{1+\gamma} - (z\gamma F + A\rho)^2\right]$	Pooling-to-separation boundary of the LMSE; $0 < \lambda_p < \lambda_2 < \lambda_s$ on $(0, \bar{\gamma})$
$\bar{\gamma}$	unique zero of λ_p on $(0, \gamma_2)$	$\lambda_p > 0$ on $(0, \bar{\gamma})$
$\gamma_0,\ \bar{\lambda}_p$	$\gamma_0 = \arg\max \lambda_p$, $\bar{\lambda}_p = \lambda_p(\gamma_0)$	Peak of the unimodal boundary
$\bar{\gamma}_1(\lambda), \bar{\gamma}_2(\lambda)$	roots of $\lambda_p(\gamma) = \lambda$ on $(0, \bar{\gamma})$	Pooling interval in γ (Proposition 4)
$F^o(z)$	$\dfrac{z(\sqrt{z+2} - \sqrt{z+1})}{\sqrt{z(z+1)(z+2)} + \sqrt{z+2} - (1-z)\sqrt{z+1}}$	$\bar{\gamma} \le \bar{\gamma}_B$ iff $F \ge F^o$ (second separation inside the admissible range)
$\bar{\gamma}_0(\lambda)$	smaller root of $\lambda_s(\gamma) = \lambda$	Costless-to-costly transition in γ (Proposition 9)
$\hat{\gamma},\ \underline{F}_s, \overline{F}_s$	$\hat{\gamma} = \min\{\gamma_1^T(z), 1/z\}$; $\underline{F}_s, \overline{F}_s$ from the roots of ξ	Sign of the welfare jump at λ_p^+ (Proposition 8)
λ^o	$\dfrac{1}{4a}\left[\dfrac{z\gamma(1-\gamma F)^2}{1+\gamma} - \left(\dfrac{1+\gamma F(2za-1)}{2a-1}\right)^2\right]$	Interior welfare maximum; $N_{c,H}^s(\lambda^o) = q_H^W$
q_i^W	$[1-\gamma(1-z_i)F]/[1+2\gamma(1-z_i)]$	Welfare-optimal customer base of a known type
$\gamma_c, F_c, \gamma_w, \underline{F}_w, \overline{F}_w$	given in the proof of Corollary 1	Pooling versus costless separation