Holding Platforms Liable [*]

Xinyu Hua[†] Kathryn E. Spier[‡]
HKUST Harvard University

October 16, 2023

Abstract

Should platforms be held liable for the harms suffered by users? A two-sided platform enables interactions between firms and users. There are two types of firms: harmful and safe. The harmful firms impose larger costs on the users. If firms have deep pockets then platform liability is unwarranted. Holding the firms liable for user harms deters the harmful firms from joining the platform. If firms are judgment proof then platform liability plays an instrumental role in reducing social costs. With platform liability, the platform has an incentive to raise the interaction price to deter harmful firms and invest resources to detect and remove harmful firms from the platform. To prevent overinvestment in detection and removal, the residual liability assigned to the platform may be partial instead of full. The optimal level of platform liability depends on the impact on user participation, the intensity of platform competition, and whether users are involuntary bystanders or voluntary consumers.

[*] We would like to thank Gary Biglaiser, Luís Cabral, Jay Pil Choi, James Dana, Andrei Hagiu, Bård Harstad, Ginger Jin, Yassine Lefouili, Hong Luo, Bentley MacLeod, Sarit Markovich, Haggai Porat, Urs Schweizer, Emil Temnyalov, Marshall Van Alstyne, Rory Van Loo, Julian Wright and seminar audiences at Boston University, Chinese University of Hong Kong, Columbia University, Fudan University, Georgetown University, Harvard University, the Kellogg School at Northwestern University, Asia Pacific Industrial Organization Conference (APIOC 2021), International Industrial Organization Conference (IIOC 2022), Economics of Platforms Seminar (TSE 2022), the 2022 Asia Meeting of the Econometric Society, American Law and Economics Association Conference (ALEA 2022), Society for Institutional & Organizational Economics Conference (SIOE 2022), JRC-TSE Workshop on Liability in the Digital Economy (2022), Society for the Advancement of Economic Theory Conference (SAET 2022). We also thank the support from the Hong Kong Research Grants Council (GRF Grant Number: 16500722).

[†] Hong Kong University of Science and Technology. xyhua@ust.hk.

[‡] Harvard Law School and NBER. kspier@law.harvard.edu.

1 Introduction

Online platforms are ubiquitous in the modern world. We connect with friends on Facebook, shop for products on Amazon, and search online for jobs, information, and entertainment. While the economic and social bene ts created by platforms are undeniable, the costs and hazards for users are very real too. For example, platform users run the risk that their personal data and privacy will be compromised. Users of social networking sites and search engines may be misled by fraudulent advertisements and misinformation. Consumers who shop online run the risk of purchasing counterfeit, defective, or dangerous goods. Should internet platforms like Facebook and Amazon be liable for the harms su ered by users?

In the United States, platforms enjoy relatively broad immunity from lawsuits brought by users, although this immunity is being challenged in legislatures and the courts.[1] Section 230 of the Communications Decency Act, enacted in 1996, shields platforms from liability for the digital content created by their participants.[2] Early proponents argued that the law was necessary to allow the internet to grow and ourish, but its application is controversial and many critics question the law's merits.[3] In 2019, Facebook paid $5 billion to settle charges that they failed to take adequate precautions to protect user data.[4] The FTC has also been investigating how \platforms screen for misleading ads for scams and fraudulent and counterfeit products" and, \in 2022 alone, consumers reported losing more than $1.2 billion to fraud that started on social media, more than any other contact method."[5] Proposed federal legislation would hold platforms liable if they fail to protect users.[6]

Marketplace platforms have largely avoided responsibility for defective products and services sold by third-party vendors. In 2019 the Fourth Circuit held that Amazon.com is not a traditional seller and therefore not subject to strict tort liability.[7] The following

---

[1] See Buiten et al. (2020) for discussion of the European Commission's e-Commerce Directive. Hosting platforms in the EU may avoid liability for illegal content posted by users, assuming they are not aware of it, and are not responsible for monitoring the legality of the posted content.

[2] Section 230(c)(1) says that \No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."Proponents hoped Section 230 would address the \perverse incentives" created by

year, a California court found that Amazon could be held strictly liable for a defective laptop battery that was sold by third-party vendors but \Ful lled by Amazon." [8] Then, in 2021, Amazon was held strictly liable for harms caused by a defective hoverboard that was shipped directly to the consumer by an overseas third-party vendor. Although Amazon did not ful ll the hoverboard order, the court opined that Amazon was \instrumental" in its sale and that \Amazon is well situated to take cost-e ective measures to minimize the social costs of accidents." [9] In short, the law is far from settled.

Such settings include social and professional networking platforms such as Facebook and LinkedIn where the users enjoy same-side network benefits from sharing content with each other and the firms pay the platform to access user data or to engage in influential activities (e.g., advertising). Platform users may be harmed by the firms when their private data is breached or when they are exposed to harmful advertising or misinformation. Absent liability the harmful firms have no incentive to leave the platform, and the platform has an insufficient incentive to detect and remove them. Holding the firms and the platform jointly liable gets them to internalize the negative externalities on the user-bystanders.

If the firms have deep pockets, and must pay in full for the harms they cause, then platform liability is unwarranted. Holding just the firms liable achieves the first-best outcome. Platform liability is socially desirable when the firms are *judgment proof* and immune from liability.[17] First, if the platform is held liable, the platform will raise the interaction price for the firms to reflect the platform's future liability costs. If the harmful firms are "marginal" (i.e., the harmful firms have a lower willingness to pay than the safe firms) then the higher interaction price deters the harmful firms from joining the platform. Second, if the harmful firms are "inframarginal" and undeterrable, the platform will invest resources to detect and remove the harmful firms from the platform.[18] Interestingly, the optimal level of platform liability may be partial instead of full, as full liability could lead to excessive auditing by the platform.[19]

We then consider the more general setting with *heterogeneous* users where some join the platform and others do not. We show that platform liability has the added benefit of stimulating user participation. This happens for two reasons. First, users anticipate that the platform's auditing incentives are improved and that the platform is safer. Second, users view the larger damage award as a "rebate" for joining the platform. Because of the *user-participation effect*, the optimal platform liability is higher than in the baseline model.

Next, we extend the baseline model to settings where *users are customers* of the firms, so interactions require the users' consent. Relevant settings include online marketplaces like eBay and Amazon where participants enjoy cross-side benefits from the sale of goods and services. As in the baseline model there are two types of sellers, harmful and safe. The harmful sellers have lower production costs but cause harms more frequently. The consumers are sophisticated and their willingness-to-pay reflects their rational expectations about product risks. The risk of harmful products depresses the price that consumers are willing to pay and, by extension, depresses the revenues that the platform can generate.

[17] Shavell (1986) provides the first rigorous treatment of the judgment proof problem, where injurers with limited assets tend to engage in risky activities too frequently and take too little care.

[18] If the firms are very judgment proof and can evade liability, then the harmful firms are inframarginal (i.e. the harmful firms have a strictly higher willingness to pay than the safe firms). If the firms are moderately judgment proof, then the harmful firms are "marginal."

[19] If the firms are very judgment proof, then the safe firms are marginal and the harmful firms get information rents. When choosing its audit intensity, the platform does not take into account the lost

If the harmful rms are marginal, then platform liability is unwarranted. Since consumers are willing to pay more for safer products, the platform has a private incentive to raise the interaction price to deter the harmful rms from joining the platform. If the harmful rms are inframarginal, however, then partial platform liability gives the platform an appropriate incentive to audit and remove the harmful rms.[20] Since the platform internalizes the average harm to consumers, the socially-optimal platform liability is *lower* than in the baseline model

one-sided platforms (e.g., for socialplatfors(s.)]TJ/F65 7.9701 Tf 69.4336 4.339 Td [(1r)]TJ/F34 11.95
are btst(anmers)240((and)241[(m)1(nd)241[(partcip(ate)-40(ion)-40(iolyn)-40(ioed)241[ofn)-40((the)-407(plat
(auditing)4127(and)4128(llso)-27((ate)4127(thmaseiv)27(es)4127(qunaly))4127(the)-709(Tf)(1r)]TJ 0 -14.445
raisms the platf(rm's)-264(incen)27(tiv)27(sg)2643(to)-247(deter)-264(the)-248(harmful)-243(rms)-25(bt)28
baseline (mo)-27(dem.)-33(Ino)-91[(t(is)-390cbasl,)2983(platform)-91[(liabo)1llity is so((ale)-91[(fo)-91[(the)2
platforms tds-4063boyes

Shares consumers rms.-9414Pproductl
liability imty()(so)-27(cialle)-52(desirialre)-52(ife)-52((com)1ms)1[umiispe produce(Sp) 1977;l
anderons,
e produce afeity the tmhe urcv Simons,yReninran
)975(.,)30 5583(Huag)4847(and)4847Spiers(*525(d)4847emphmslizd)4847(thd)483[(partculare)]TJ-01s030 -14.446 T

liability to the consumer
n irec(ly) for the tm'is harrs.-570(Hpa)28
shoulds-313(())-27sy a consumer the produc.s
(y)- (t(rnd)-324(part)27(y)-)4(bt)27(st(anm
ccouenable then the liability to the iman chan
internal, the harm[40]

extending liability to an injurer's lenders[25] and Dari Mattiacci and Parisi (2003) consider vicarious liability where liability is extended to the injurer's employer.[26] Arlen and MacLeod (2005a) show that holding managed care organizations liable for medical malpractice by their physicians can raise the physicians' incentives to take care. Our model, which has not been previously studied, investigates the design of platform liability when the platform can audit and remove harmful participants.[27]

There is a vast literature on multi-sided platforms. The early studies (e.g., Caillaud and Jullien, 2003; Rochet and Tirole, 2003, 2006; Amstrong, 2006; and Weyl, 2010) have identi ed how cross-side externalities a ect platform pricing schemes and users' participation incentives. The literature also examines the impact of seller competition[28] or the impact of platform competition on pricing.[29] Some recent studies pay attention to non-pricing strategies, including seller exclusion (Hagiu, 2009), information management (Julien and Pavan, 2019; Choi and Mukherjee, 2020), control right allocation (Hagiu and Wright, 2015, 2018), and platform governance (Teh, 2022).

There is a small but growing literature on platform liability. The policy papers by Buiten et al. (2020) and Lefouili and Madio (2022) discuss informally whether platforms should bear liability for harms caused by participants. A few working papers study copyright infringement and retail settings. De Chiara et al. (2021) examine hosting platforms' incentives to lter copyright-infringing materials. They focus on harms to copyright owners and do not consider platforms' pricing strategies. Jeon et al. (2022) examine how negligence-based liability changes platforms' incentives to remove IP-infringing products, which in turn a ects brand owners' innovation incentives. Zennyo (2023) considers the impact of platform liability on sellers' e orts to improve product safety, without discussing platforms' screening or auditing actions. Yasui (2022) discusses sellers' incentives to maintain reputation and platforms' ex-post e orts to discover and announce potential safety risks after consumers purchase products from sellers. Our paper considers a broad array of platform types and investigates the e ects of liability on platform pricing, incentives to block bad actors, and social welfare.

Our paper is organized as follows. Section 2 presents the baseline model where users are homogeneous bystanders of the rms. Section 3 generalizes the baseline model by considering heterogeneous users with endogenous participation. Section 4 examines sev-

[25]See also Boyer and La ont (1997) and Che and Spier (2008). Bebchuk and Fried (1996) argue informally for raising the priority of tort victims in bankruptcy above debt claims gives the debtholders an incentive to better monitor the borrower.

[26]There are related legal studies. See Kraakman (1986) for a general taxonomy of gatekeeper enforcement strategies, Hamdani (2002) for liability on internet service providers, Hamdani (2003) on accountants and lawyers, and Van Loo (2020a) on big technology.

[27]Our paper is also related to the studies comparing joint and several liability (JSL) to several liability (SL) for harms caused by multiple defendants (e.g., see Landes and Posner, 1980; Carvell et al., 2012). With JSL, the victim may recover full damages from a single deep-pocketed defendant. With SL, the victim's recovery from each defendant is limited by the defendant's share of responsibility.

[28]See Nocke et al. (2007), Galeotti and Moraga-Gonzalez (2009), Hagiu (2009), Gomes (2014), Belleamme and Peitz (2019).

[29]See Dukes and Gal-Or (2003), Hagiu (2006), Armstrong and Wright (2007), White and Weyl (2010), Karle et al. (2020), Tan and Zhou (2021).

eral extensions including a retail setting where the firms are sellers and the users are consumers and a setting with two competing platforms. Section 5 provides concluding thoughts. The proofs are in the appendix.

2 The Baseline Model

Consider a two-sided platform (P) with two kinds of participants, firms (S) and users (B). The platform is a monopolist and necessary for interactions between firms and users. Firms and users are small, have outside options of zero, and the mass of each is normalized to unity.

The platform provides two goods. First, the platform provides a quasi-public good that gives each user a private benefit $v > 0$. For simplicity, we first consider the special case where users are homogeneous and have the same v. Section 3 generalizes the analysis to include heterogeneous users with different valuations. Second, the platform provides opportunities for the firms and the users to interact.

We assume that interactions between firms and users do not require the users' consent and so the users are effectively "bystanders."[30] The benefits and costs of these interactions depend on the firms' type, $i \in \{H, L\}$, where is the mass of type H and 1 is the mass of type L in the firm population.[31] The H-type firms have higher interaction benefits, $_H > _L$, but impose higher interaction losses on users, $_H d > _L d$ where $_i \in [0, 1]$ is the probability of harm and $d > 0$ is the level of harm per firm-user interaction.[32] The firms privately observe their types.

We assume that the platform charges the firms a price p per interaction and allows users to join the platform for free. This is broadly aligned with what we often observe in practice. Platforms such as Google and Facebook monetize quasi-pubic goods by selling online advertising to businesses and/or sharing user data and do not charge users for access. In theory, this pricing strategy can be very profitable for the platform in strategic environments with strong network effects.[33] Our assumption is also aligned with other papers in the platform literature.[34]

The platform has the capability to detect and block the H-type firms. We will refer to the platform's efforts to detect the H-types as auditing. By virtue of their scale, data,

---

[30] Section 4.1 extends the analysis to retail platforms where interactions require the users' consent.

[31] For simplicity, is taken as exogenous. One may endogenize by allowing firms to invest resources to increase the likelihood being safe. Section 4.3 discusses an extension with firm moral hazard problems.

[32] If $_H < _L$ then the H-types are marginal for all liability rules and auditing is unnecessary. The threshold $\hat{v}$ defined in (5) below is identically equal to zero, and all of our results apply.

[33] Suppose that each user receives v only if a large number of users join the platform. A user's decision to join depends on the price and their expectations about the number of other users. Following Harsanyi and Selten (1988), to avoid coordination failure, the platform should set a sufficiently low price (or even zero price) for the users. The appendix provides an illustrative example of the coordination game.

[34] Armstrong (2006) shows that, with strong network effects, platforms have incentives to set negative prices. However, negative prices may be infeasible. Armstrong and Wright (2007) and Choi and Jeon (2021) justify non-negative prices on adverse selection and moral hazard grounds. Gans (2022) justifies this based on free disposal. See also Belleflamme and Peitz (2021).

and technological sophistication, platforms like Google may be in a good position to root out harmful platform participants.[35] Specifically, by spending effort $e \in [0, 1)$ per firm, the platform can detect H-type firms with probability e and block them from interacting with users.[36] We assume that the cost of effort $c(e)$ satisfies $c(0) = 0$, $c'(e) > 0$, $c''(e) > 0$, $c'(0) = 0$, and $c'(e)$

to join the platform.[43] A2 guarantees that the platform always gets non-negative pro ts

2.1 Motivating Examples

In our baseline model, bad actors on one side of a two-sided platform may harm users on the other side of the platform. In the following, we motivate the baseline model with three broad examples: fraudulent advertising, data misuse by technology partners, and the sale of harmful products. For each of these three settings, we will document the platform's nancial incentives, the presence of bad actors, and the potential for user harm.

Advertisers. Many platforms rely on paid advertising as their main source of revenue.[44] Although most online advertising is benign, fraudulent and misleading ads abisle82(partners,)-29bbo ing app developers, who can use the data to improve their o erings and the experiences of platform users. When deciding whether to grant developers access to user data, platforms may consider the nancial bene t (among other things). For example, in 2013, Facebook allegedly granted or denied access based on the developer spending at least, $250 in mobile advertising.[51] In the Spring of 2023, Twitter, Reddit, and other platforms announced hefty charges for developers to access the platforms' API, leading some partners to reduce their data usage and others to terminate their contracts.[52] There is substantial evidence that technology partners violate their platform developer agreements[53] and sell data to others.[54] In the wrong hands, platform data can \be used for identity theft, phishing, fraud, and other harmful purposes.[55]

Third-Party Sellers. Retail and gaming platforms make money by sharing sales revenue with third-party sellers. While many consumers are sophisticated and can understand the risks that they face online,[56] others may be unaware of the risks or cannot meaningfully consent to transactions (e.g., children). Consumers who are naïve and unaware of the harms are, for all intents and purposes, bystanders. For example, some online games use \dark patterns" to exploit cognitive biases and to manipulate users into making online purchases. Trickery was central in the FTC cases against Apple, Google, and Amazon for in-app charges associated with \free" games for children.[57] The FTC complaint against Apple described third-party game Dragon Story \as `sucker[ing] young children into spending huge amounts of money' without their parents' knowledge."[58] In-app sales are nancially lucrative for third-party sellers and for the platform. Apple, for example, retained thirty percent of all revenue, including in-app sales.[59]

---

[51] \Facebook's enforcement of its policies, terms, and conditions, however, was inadequate and was in uenced by the nancial bene t that violator third-party app developers provided to Facebook." See *United States of America v. Facebook Inc.*, Case 1:19-cv-02184, Complaint for Civil Penalties, Injunction, and Other Relief (Filed 07/24/19). https://www.justice.gov/opa/press-release/ le/1186506/download

[52] See \Reddit Wants to Get Paid for Helping to Teach Big AI Systems," *New York Times*, Apr 18, 2023.

[53] The \misuse of data shared with third-party apps on Facebook [includes] ransomware, spam, and targeted advertising." See Farooqi et al. (2020)

[54] App developer Aleksandr Kogan shared the personal information of 87 million Facebook users with Cambridge Analytica. Kogan allegedly received over $800,000 for the cothesom0yn8(elop)-e14T44cvv.infrserya2a84ox-33

2.2 Equilibrium Analysis

In this subsection, we characterize the platform's pricing and auditing strategies, p and e, given the assignment of liability, w_s and w_p. A type-i rm will seek to join the platform when their expected pro t per interaction is non-negative,

$$_i \quad _i w_s \quad p \quad 0; \tag{4}$$

where $_i$ is the rm's interaction bene t, $_i w_s$ is the rm's expected liability, and p is the price paid to the platform. Note that depending on the level of rm liability, w_s, the H-type may have higher or lower rents than the L-type. The rents of the two types are equal when

$$w_s = \hat{w} = \frac{_H \qquad _L}{_H \qquad _L} < d: \tag{5}$$

The threshold $\hat{w}$ de ned in (5) is critical for understanding the impact of platform liability on the interaction price and audit intensity. If the rms are su ciently judgment-proof, $w_s <$ b

We now explore the platform's incentive to audit and remove the H-type firms. The platform's aggregate profits are:

$$(e) = (1$$

(a) *If* $_H(d \quad w) > r_H(w_s)$ *then* $0 < e < e$.

(b) *If* $_H(d \quad w) = r_H(w_s)$ *then* $0 < e = e$.

(c) *If* $_H(d \quad w) < r_H(w_s)$ *then* $0 < e < e$.

To summarize, when rm liability is below the threshold, $w_s \quad \hat{w}$, the H-type rms cannot be deterred from joining the platform by the interaction price p. The platform invests in auditing if and only if the joint platform- rm surplus is larger than the rms' information rent. Note that the platform's incentives to audit are stronger when w_p and w_s are larger. The platform's incentive to audit and remove the H-types is socially insu cient when the joint liability for the platform and rms is small (as in case 2(a)) but socially excessive if the joint liability is large (as in case 2(c)).

Case 2: $w_s > \hat{w}$. Now suppose that rm liability is above the threshold, so the H-type rms are marginal. The platform's pro t-maximizing strategy is to either charge $p_L = {}_L \quad {}_L w_s$ and deter the H-types from joining the platform or charge $p_H = {}_H \quad {}_H w_s < p_L$ and attract both types. Notably, if the platform chooses the latter strategy, then it will not invest in auditing, $e = 0$.[62]

The platform will charge p_H and attract the H-types if and only if

$$(p_H \quad {}_H w_p) + (1 \quad)(p_H \quad {}_L w_p) > (1 \quad)(p_L \quad {}_L w_p):$$

Substituting the formulas for p_H and p_L and using the de nition of $\hat{w}$ in equation (5) this condition becomes:

$$({}_H \quad {}_H w) > (1 \quad)({}_H \quad {}_L)(w_s \quad \hat{w}): \tag{11}$$

The left-hand side is the joint platform- rm surplus of attracting the H-type rms on the platform: the fraction of H

2.3 Platform Liability

This subsection explores the social desirability and optimal design of platform liability for harm to user-bystanders, taking the level of firm liability w_s as fixed. We begin by presenting a benchmark where the platform is not liable for the harm, $w_p = 0$.

Proposition 1. *(Firm-Only Liability.) Suppose that the platform is not liable for harm to users, $w_p = 0$, and firm liability is $w_s \; 2 \; [0; d]$. There exists a unique threshold $\hat{w} = \hat{w}(\;) \; 2 \; \underline{w}; \frac{H}{H}$, where $\hat{w}(\;)$ wea H-types, .*[63]

1. *If $w_s \; \underline{w}$ then the platform sets $p_s = \; _L \; _L w_s$, attracts the H-type firms, and does not invest in auditing, $e = 0 < e$. The platform's auditing incentives are socially insufficient.*

2. *If $w_s \; 2 \; (\underline{w}; \hat{w})$ then the platform sets $p_s = \; _H \; _H w_s$, attracts the H-type firms, and does not invest in auditing, $e = 0 < e$. The platform's auditing incentives are socially insufficient.*

3. *If $w_s \; \hat{w}$ then the platform sets $p_s = \; _L \; _L w_s$ and deters the H-type firms. The first-best outcome is achieved.*

Should platforms be held liable for the harm suffered by users? Proposition 1 establishes that platform liability is unnecessary when the firms themselves are held sufficiently liable for harm to the users (case 3 in Proposition 1). In this case, the joint platform-firm surplus of including the H-types is low, so the platform has incentives to deter them by charging a high price. However, when the firms are more judgment proof and the platform faces no liability (cases 1 and 2 in Proposition 1), the private and social incentives diverge. The platform attracts the H-types and does not invest in costly auditing. In such cases, platform liability can be socially desirable, as shown in the next proposition.

Proposition 2. *(Optimal Platform Liability.) Suppose firm liability is $w_s \; 2 \; [0; d]$. The socially-optimal platform liability for harm to users, w_p, is as follows:*

1. *If $w_s \; \underline{w}$ then $w_p = d \;\; w_s \;\; 1 \;\; \frac{L}{H} \; (\underline{w} \;\; w_s) \; 2 \; (0; d \;\; w_s]$ achieves the second-best outcome. The platform sets $p = \; _L \; _L w_s$ and attracts the H-type firms. The platform's auditing incentives are socially efficient, $e = e$.*

2. *If $w_s \; 2 \; (\underline{w}; \hat{w})$ then there exists a threshold $\underline{w}_p \; d > 0$*

Proposition 2 describes how platform liability can be designed to increase social welfare. In case 1, rm liability is below the threshold (w_s $\hat{w}$) and the L-type rms are marginal. From Proposition 1 we know that rm-only liability fails to deter the H-types and gives the platform no incentive to audit and remove the H-types. Imposing liability on the platform motivates the platform to take auditing e ort. If $w_s < \hat{w}$ and the platform was held responsible for the full residual harm, $w_p = d$ w_s, then the platform would *ove r invest* in auditing. Therefore the second-best outcome is achieved when the platform bears some but not all of the residual damage, w_p 2 $(0; d$ $w_s)$. If $w_s = \hat{w}$, then the second-best outcome is achieved when the platform bears full residual liability.

In case 2, the rms' liability is in an intermediate range and the H-type rms are marginal. According to Proposition 1, without platform liability, the platform would attract the H-type rms since the joint platform- rm surplus of including the H-types is larger than the L-type rms' rents. Since the rms' rent is independent of w_p while the joint surplus of keeping the H-types decreases in w_p, the social planner can motivate the platform to raise the price and thus deter the H-types by imposing residual liability on the platform, $w_p = d$ w_s.

Finally, in case 3, platform liability is unnecessary when rm liability is su ciently high. As in Proposition 1, the rst-best outcome is obtained without platform liability.

This section investigated the need for platform liability when the rms that participate on the platform cause harm to homogeneous user-bystanders. If rms have deep pockets and can compensate the user-bystanders for the harm that they cause, then platform liability is unwarranted. If rms are judgment proof or can evade liability in other ways, then platform liability is socially desirable. Holding the platform liable for some or all of the residual harm has two potential bene ts. First, the platform may raise the price that it charges to the rms, which will help to deter rms that pose excessive risks to users. Second, the platform will invest resources to detect and remove risky rms from the platform. How9552pTf7(y)-39(deter-394from)-2053(resoorm- rm03us(5apry)-362(on-245(whenlatfot)1

$f(v) > 0$ for $v \in [0, \infty)$, with cumulative density $F(v)$.[64] As in the baseline model, the platform charges the rms price p per interaction and takes auditing e ort e per rm. Note that there are economies of scale in (per- rm) auditing, so that both the private and the socially optimal incentives for auditing depend on the users' participation rate.[65] Users have the option to join the platform for free. As discussed in the baseline model, many platforms do not charge users in practice and this observation could emerge in equilibrium when there are strong same-side or cross-side network e ects.[66]

We assume that the users cannot directly observe the platform's audit intensity, or equivalently, the platform chooses its audit intensity after the users make their participation decisions.[67] Although the users do not observe the platform's auditing e ort e when making their participation decisions, they observe the liability rule, w_s and w_p, and form correct beliefs about e in equilibrium.

In practice, the public does not directly observe platforms' enforcement e orts or technologies used in improving platform safety. In the words of former Facebook employee and whistleblower Frances Haugen, "Facebook became a $1 trillion company by paying for its pro ts with our safety, including the safety of our children" and "almost no one outside of Facebook knows what happens inside Facebook."[68] The Digital Services Act in the European Union and the PACT Act recently proposed in the US contain many disclosure requirements,[69] which re ects lawmakers' concerns about the lack of transparency on platform safety and e ort.[70]

Consider the rst-best outcome. Assumption A2 implies that it is socially e cient for not join the platform and all the users join the platform.

Next, consider the second-best outcome. As in the baseline model, full deterrence of the H-types may not be possible. If the H-type firms seek to join the platform, then costly auditing is necessary to reduce the social harm. In the second-best benchmark, social welfare is

$$S(e;\hat{v}) = \int_{\hat{v}} [v + \lambda(1-e)(b_H - \delta_H d) + (1-\lambda)(b_L - \delta_L d)] f(v)dv - c(e); \qquad (12)$$

where $\hat{v}$ is the value of the marginal user,

$$\hat{v}(e;w) = (\lambda(1-e)\delta_H + (1-\lambda)\delta_L)(d-w). \qquad (13)$$

Notice that $\hat{v}(e;w)$ is decreasing in e and w for all $d - w > 0$: higher levels of effort and liability stimulate user participation. Holding e constant, the users view w as a "rebate" for joining the platform. Therefore, the social planner would like to set $w = d$ (that is, $w_p = d - w_s$), so that all the users participate. Given full participation by the users, the socially efficient auditing effort is e^*, the same as in the baseline model.

We now characterize the equilibrium and the optimal platform liability. As in the baseline model, the L-type firms are marginal if $w_s \leq \hat{v}$, while the H-types are marginal if $w_s > \hat{v}$. We consider each case in turn.

Case 1: $w_s \leq \hat{v}$. In this case, the L-type firms are marginal and the platform charges $p^u = b_L - \delta_L w_s$. The platform's profit function may be written as:

$$\Pi(e;\hat{v}) = S(e;\hat{v}) + \int_{\hat{v}} \Big[\lambda(1-e)(\delta_H - \delta_L)(\hat{v} - w_s) + ((\lambda(1-e)\delta_H + (1-\lambda)\delta_L)(d-w) - v)\Big] f(v)dv; \qquad (14)$$

where $\hat{v}$ is the marginal user defined in (13). Since the platform chooses its auditing effort ex post given $\hat{v}$, the platform's auditing effort e^u (if it is positive) satisfies[71]

$$\frac{\partial \Pi(e^u;\hat{v})}{\partial e} = \frac{dS(e^u;\hat{v})}{de} + \int_{\hat{v}} [-\lambda(\delta_H - \delta_L)(\hat{v} - w_s) - \lambda\delta_H(d-w)] f(v)dv + \lambda\delta_H(d-w)\frac{\partial S(e^u;\hat{v})}{\partial \hat{v}} = 0; \qquad (15)$$

where

$$\frac{\partial S(e^u;\hat{v})}{\partial \hat{v}} = \frac{\partial \Pi(e^u;\hat{v})}{\partial \hat{v}} - \lambda(1-e^u)(\delta_H - \delta_L)(\hat{v} - w_s) f(\hat{v}). \qquad (16)$$

Equation (15) shows that the platform's auditing incentives diverge from the social planner's. The first line of equation (15) is familiar. As in the baseline model, when the platform increases e, the removed H-types lose their information rents, $\lambda(\delta_H - \delta_L)(\hat{v} - w_s)$

[71] See the proof of Proposition 3.

and the users' uncompensated loss is reduced by$_H$ $(d \quad w)$. If $w_p = w_p$ as de ned in Proposition 2, these two e ects o set each other. The last line of equation (15) identi es a new source of divergence: the platp(y)]TJ/B(div)27olti es

2. *If* $w_s = \hat{w}$ *then* $w_p^u = d \quad w_s$ *achieves the second-best outcome. The platform sets* $p^u = \quad _L \quad _L w_s$ *and chooses the e cient auditing e ort* $e^u = e$. *All users participate.*

3. *If* $w_s > \hat{w}$ *then* $w_p^u = d \quad w_s$ *achieves the rst-best outcome. The platform sets* $p^u = \quad _L \quad _L w_s$ *and deters the H-type rms. All users participate.*

To summarize, as in the baseline model, if the rms have deep pockets and can be held fully liable ($w_s = d$), platform liability is unnecessary. However, if the rms are judgment proof, platform liability can motivate the platform to take more auditing e ort or raise the interaction price, which removes or deters the harmful rms. Additionally, platform liability stimulates user participation. So, the optimal level of platform liability is weakly higher than in the baseline model. Note that, when the rms are very judgment proof (Case 1 in Proposition 3), the optimal platform liability leads to excessive auditing.

Remark Lawmakers and commentators have historically expressed concern that the burden of liability might chill economic activity. These concerns were part of the rhetoric for platform immunity to liability in the early years. Section 230 of the Communications Decency Act was adopted to allow the internet to grow and ourish.[75] To be sure, defending against frivolous lawsuits can be costly and distract managers from the core business.[76] However, our analysis shows that platform liability can stimulate user participation, both directly and indirectly.[77]

First, platform liability serves as a \rebate" to attract users. This e ect is unique to the platform market. To see this, consider a non-platform market where a seller sells its product to consumers. Although products liability reduces consumers' uncompensated harm, it raises the seller's costs and leads the seller to raise the price of the product, which can neutralize the impacts on output.[78] By contrast, in a platform market with strong network e ects, users have the option to join the platform for free. The platform does not adjust the price to fully re ect the users' uncompensated harm or the platform's liability costs and cross-side network bene ts (i.e. revenue from the rms). Platform liability stimulates participation by reducing the \e ective" price for users.

Second, platform liability can raise the audit intensity, which attracts users indirectly. For both platform and non-platform markets, when users cannot observe product safety or the platform's audit intensity, liability addresses the moral hazard problem and improves

[75] Section 230 and has been called \the one line of federal code that has created more economic value in this country than any other." See https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change.

[76] Court errors and litigation costs are discussed in Section 4.3.

[77] Some empirical studies observe a positive correlation between liability and innovation. Viscusi and Moore (1993) observe that when products liability is low or moderate, raising liability encouraged rms' investments in innovation. Galasso and Luo (2017) identify a positive correlation between liability and innovation.

[78] If consumers have the same preference for product safety and can observe safety before purchase, then liability is irrelevant to output (Hamada, 1976).

safety. The increased safety reduces the joint costs for the platform and users (or the seller and consumers), thereby stimulating user participation. However, the moral hazard problem is not the only reason for the divergence between the platform's auditing incentive and the social incentive. As shown by equation (15), the divergence occurs also because the platform does not consider the bene t of auditing to the inframarginal users or the impact of increased participation on the rms' rents. Platform liability addresses these externalities and motivates the platform to raise audit intensity.

Rema r Platform liability may be socially bene cial when users observe the platform's auditing e ort e before making their participation decisions. In this setting, the platform's auditing incentives are stronger. Recall that in equation (15), when e ort is not observable, the platform disregards the social bene t of increased participation (the last term). With observable e ort and w_S b, the platform's e ort (if it is positive) satis es:

$$\frac{d\ (\ e^u; b)}{de} = \frac{dS(e^u; b)}{de} + \int_{b} [\ (\ _H \quad _L)(b \quad w_S) \quad _H(d \quad w)] f(v)dv$$
$$+ \quad _H(d \quad w) \quad \frac{\partial S(e^u; b)}{\partial b} \quad \frac{\partial\ (\ e^u; b)}{\partial b} = 0. \quad (17)$$

where

$$\frac{\partial S(e^u; b)}{\partial b} \quad \frac{\partial\ (\ e^u; b)}{\partial b}$$

term in equation (17) drops out and we are left with $d(\ )/de = dS(\ )/de \int_v H(d\ w)f(v)dv = 0$. Private and social incentives diverge because the platform does not consider the safety benefits that accrue to the participating users. Imposing full residual liability on the platform, the platform, but are sophisticated and form beliefs that are, in equilibrium, correct.[95] If the H-type rms seek to join the platform and the platform invests e in auditing, the conditional probability of harm per interaction is

$$E(\;|e) = \frac{(1\quad e)\quad _H + (1\quad\quad)\;_L}{(1\quad e)\quad + (1\quad\quad)}; \tag{19}$$

which is a decreasing functionp1

rents from the L-type rms, $p^r = t^r$ $({}_L w_S + c_L)$. Using (20) and ${}_L =$ ${}_0$ c_L,

$$p^r = {}_L \quad {}_L w_S \quad {}^r (d \quad w): \tag{21}$$

Comparing p^r to its counterpart p (see (6)) in the baseline model reveals an important di erence: the interaction price paid by the rms (21) re ects the user-consumers' expected uncompensated harm, ${}^r (d \quad w)$.

We now explore the platform's auditing incentives. Substituting p^r from (21), $S(e)$ from (2), and $\hat{w}$ from (5) into (7) gives the platform's pro t function

$$\begin{aligned} (e) = S(e) \quad v \quad (1 \quad e) \; ({}_H \quad {}_L)(\hat{w} \quad w_S) \\ + [(1 \quad e) \; ({}_H \quad {}^r) + (1 \quad)({}_L \quad {}^r)](d \quad w): \end{aligned} \tag{22}$$

The platform's pro ts (e) diverge from social welfare $S(e)$ for two0 incremental social bene t of attracting the H

By contrast, when users are *consumers*, the retail price t^r paid by the users to the firms (and the price p^r paid by the firms to the platform) reflects the users' beliefs of the probability of harm. In Proposition 4, when the users are consumers, w_p^r satisfies

$$(\ _H \quad _L)(\bar{w} \quad w_s) = (\ _H \quad)(d \quad w_s \quad w_p^r): \tag{26}$$

Now the right-hand side reflects the users' *uncompensated harm beyond their expectations.* As in the baseline model, when firm liability w_s rises, both sides fall. However, the drop in the firms' rent on the left is *bigger* than the drop in the users' uncompensated harm (beyond their expectations) on the right. Holding w_p fixed, the platform would invest *too little* in auditing. To restore the efficient incentives for auditing, platform liability should be raised. This is why platform liability and firm liability are complements in the retail platform extension.

Corollary 1. *Suppose $w_s \quad \bar{w}$. When the users are bystanders, the optimal platform liability decreases in w_s; when the users are consumers, the optimal platform liability increases in w_s.*

Remark The analysis above assumed that the platform removed discovered H-types from the platform. What would happen if the platform is required to disclose the audit results to the consumers, and the consumers decide for themselves whether to interact with the known H-types? Absent platform liability ($w_p = 0$), a rational consumer would decline to interact with a known H-type ex post.[94] Although ex post efficiency would be obtained without platform liability, the platform would have insufficient incentives to audit the sellers ex ante.[95] At the other extreme, with full platform liability ($w_p = d$), a rational consumer would interact with a known H-type.[96] That is, disclosure would not deter harmful interactions. These observations underscore the importance of granting retail platforms the discretion to remove bad actors rather than relying on disclosure alone.[97]

4.2 Platform Competition

We now extend our baseline model (with user-bystanders) by considering two competing platforms, Platform 1 and Platform 2. Users are distributed symmetrically on a Hotelling

[94] The joint surplus for a consumer and an H-type firm from their transaction is $\ _H \quad \ _H$

line with density $g(x) = g(1 - x) > 0$ on $x \in [0,1]$, Platform 1 is located at $x = 0$ while Platform 2 is located at $x = 1$. A user at location $x \in [0,1]$ receives consumption value $v - x$ if they join Platform 1 but $v - (1 - x)$ if they join Platform 2, where 0 reflects the level of differentiation. Assume that v is sufficiently large such that the market is fully covered. The firms can join both platforms, while each user only joins one platform.[98] Thus, the platforms compete for users but not for firms.[99]

In stage 1, the platforms set their prices simultaneously. The timing and the other assumptions are otherwise identical to the baseline model. Denote the platforms' prices and auditing efforts as p_j and e_j, $j = 1,2$. We shall focus on the symmetric equilibrium where $p_1 = p_2$ and $e_1 = e_2$ and, accordingly, each platform serves half of the users. We will show that platform liability can still be socially beneficial in this competitive environment.

Case 1: $w_s \; \hat{w}$. In this case, the L-type firms are marginal and the platforms set $p_1 = p_2 = {}_L \; {}_L w_s > 0$. Although the users do not observe the platforms' auditing efforts directly, they are sophisticated and form rational inferences in equilibrium. In

Now suppose w_s 2 ($b; e$). If $w_p = d \quad w_s$, the users would be fully compensated for any harm and therefore each platform attracts half of the users. Each platform charges p_L if

$$\frac{1}{2}(1 \quad)(p_L \quad {}_L w_p) > \frac{1}{2}[\ (p_H \quad {}_H w_p) + (1 \quad)(p_H \quad {}_L w_p)];$$

which holds given $p_H < p_L$ and $p_H \quad {}_H w_p = \quad {}_H \quad {}_H d < 0$. Hence, imposing full residual liability on the platforms gets the platforms to raise the interaction price and deter the H

3. *If $w_s \geq \overline{w}$, platform liability is unnecessary. The platforms set $p^c = \theta_L w_s$ and deter the H-type firms.*

Comparing Proposition 5 to Proposition 2 reveals how competition changes the socially-optimal level of platform liability. If the firms are very judgment proof, $w_s \leq \underline{w}$, then the socially-optimal level of platform liability is the same as for monopoly, $w_p^c = w_p$. As before, platform liability encourages the platforms to detect and remove the H-type firms from the platforms. If the firms are modestly judgment proof, $w_s \in (\underline{w}, \overline{w})$, then platform liability is socially beneficial when the platforms are sufficiently differentiated (large) but unnecessary when platform competition is fierce (small). By contrast, in the baseline model, platform liability was necessary to induce the platform to raise the interaction price to deter the bad actors. Here, when competition is fierce, the market mechanism gives the platforms the incentive to raise their interaction prices and deter the bad actors from participating.

Regulators across the globe have been focusing efforts on increasing competition and reducing market power in platform markets. For example, the Federal Trade Commission in the U.S. filed a lawsuit against Facebook, asking the court to force it to sell WhatsApp and Instagram.[100] The Digital Services Act and Digital Markets Act in the European Union are geared towards establishing a level playing field (to foster innovation and competitiveness) and creating a safer digital space for users and others.[101] Our analysis shows that policies that encourage platform competition should be complemented by changes in platform liability. When bad actors are judgment proof and undeterred, then platform liability plays an important role of encouraging platforms to invest efficiently to protect users from harm.

4.3 Other Extensions

Firm Moral Hazard. In our baseline model and main extensions, platforms played an instrumental role in solving the adverse selection problem by detecting and removing bad actors from the platforms. As discussed in Section 2.1, adverse selection is empirically relevant: Bad actors, masquerading as legitimate firms, post fraudulent advertisements, steal user data, and sell counterfeit products. Moral hazard is also empirically relevant: Otherwise legitimate app developers may sell user data to others and manufacturers may cut corners to lower costs and raise profit margins. When firms are judgment proof, platform liability can play an instrumental role in solving moral hazard problems, too.

Our baseline model can be easily adapted to reflect a moral hazard problem. Suppose all the firms are identical ex ante but may become either the L-type or H-type ex post. A firm can take (unobservable) care at cost $k > 0$, which reduces the probability of

[100] See https://www.reuters.com/technology/us-ftc-says-court-should-allow-antitrust-lawsuit-against-facebook-go-forward-2021-11-17/

[101] See https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package. A report written by Cremer, et al. and published by the European Commission (2019) raised concerns about increased concentration in platform markets. See https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

becoming an H-type. If the firms are very judgment proof ($w_s \le b\sigma$), then the H-types earn information rents. It follows that ex ante the firms have no incentive to take care and, as in the baseline model, platform liability raises the platform's auditing incentives.[102] When the firms are modestly judgment proof (w_s in a middle range), platform liability motivates the platform to raise the interaction price, which deters the H-type firms and, under certain conditions, motivates the firms to take ex ante effort.[103]

Same-Side Harms. The previous analysis considered a setting with cross-side harms: Firms on one side of the platform harmed the users on the other side of the platform. In practice, some users on platforms may harm other users. For example, some influencers on TikTok create videos that draw attention but may induce children to engage in dangerous activities; celebrities' endorsement of cryptocurrency may persuade investors to buy risky tokens.[104] These influencers can monetize user attention by collaborating with brands or sharing advertising revenue with platforms.[105]

Our model can be adapted to investigate such cases with same-side harms. Consider for example a social networking platform where most user-generated content is perfectly safe but some of it is socially harmful. Suppose further that the advertising revenue that the platform enjoys is proportional to the volume of shared content, both safe and harmful. If the users are judgment proof, and cannot be held accountable for the harmful content that they post, then holding the platform liable may make sense. Without platform liability, the platform has a financial incentive to facilitate the posting and sharing of all content, both safe and harmful; with platform liability, the platform has incentives to that the firms are very judgment proof (w_s $\underline{w}$) so that the L-types are marginal. If the H-types do not join the platform, the platform would not take any auditing effort. But anticipating this, the H-types would deviate to join. In this case, there is no equilibrium where the H-types are fully deterred.[107] Therefore, platform liability can increase the platform's auditing incentives.

False Positives. Our analysis assumed that there were no "false positives." The auditing efforts of the platform did not erroneously remove the L-type firms. Several new insights emerge when the baseline model is extended to include false positives. First, the second-best auditing effort is lower than in our baseline model (since it is socially efficient for L-types to remain on the platform). Second, the platform has weaker incentives to invest in auditing than in the baseline model (since the platform loses revenue when it excludes the L-types). Third, the platform's incentives are even weaker relative to the

reduce their control of online activities, similar to the potential distortion caused by vicarious liability on organizations.[114] Second, our model shows that platform liability may be socially desirable even if auditing is very costly or *completely ineffective* at detecting bad actors. Although platforms would not engage in auditing in this case, liability would force platforms to internalize the social harms and create an incentive for them to use the price mechanism to deter bad actors.

There is active debate over whether platforms may be treated as common carriers.[115] Common carriers, including telephone companies, mail carriers, and transportation systems (e.g., railroads and airlines) have a duty to serve the general public and may not generally exclude users.[116] Common carriers are, however, subject to regulations that ensure public safety and sometimes have discretion or even a duty to exclude parties that may cause harm to others. For example, under federal law, airlines must deny transport to passengers who refuse to be searched for weapons,[117] and airline pilots have "permissive removal" authority to deny service to passengers who appear nervous or potentially disruptive.[118] Although the Digital Millennium Copyright Act limits liability for internet service providers (ISPs), it also requires ISPs to terminate the accounts of repeat infringers.[119] In a lawsuit brought against Western Union, the court opined that the defendant was in fact *obligated* to discontinue service for illegal gambling communications.[120] Common carriers can be held liable if they fail to meet their duties[121] and, in many jurisdictions, the standard of care exceeds "reasonable care."[122]

This article advances the idea that liability can play an instrumental role making platforms safer for users and for society more broadly. An open question is whether civil liability is the best mechanism to accomplish these goals, or whether regulation would prove more e ective. Social media and other platforms share similarities to common carriers and public utilities and so, by analogy, one could in principle regulate them in similar ways. Platform liability arguably has substantial advantages over regulation. Speci cally, given the complexity and diversity of platforms, it would be di cult (and perhaps inadvisable) for regulators to set uniform safety standards.[123] Moreover, given the rapidly changing market conditions, regulators would be chasing a moving target. Platforms, especially big tech platforms, have the relevant information to weigh the social costs and bene ts. Liability would give platforms nancial incentives to use their discretion for the greater good.

[123] This view is shared by many platforms; eBay's 2022 Transparency Report states: \regulatory regimes or technology mandates that are `one size ts all' can actually serve to limit the tools, resources and partnerships necessary to combat bad actors."

[14] Che, Yeon-Koo, and Kathryn E. Spier, \Strategic Judgment Proo ng," *RAND Journal of Economics*, Vol. 39 (2008), pp. 926-948.

[15] Chen, Yongmin and Xinyu Hua, \Ex ante Investment, Ex post Remedies, and Product Liability," *International Economic Review*, Vol 53 (2012), pp. 845-866.

[16] Chen, Yongmin and Xinyu Hua, \Competition, Product Safety, and Product Liability," *Journal of Law, Economics, & Organization*, Vol. 33 (2017), pp. 237-267.

[17] Choi, Albert, and Kathryn E. Spier, \Should Consumers Be Permitted to Waive Products Liability? Product Safety, Private Contracts, and Adverse Selection," *Journal of Law, Economics, & Organization*, Vol. 30 (2014), pp. 734-766.

[18] Choi, Jay Pil, and Arijit Mukherjee, \Optimal Certi cation Policy, Entry, and Investment in the Presence of Public Signals," *RAND Journal of Economics*, Vol. 51 (2020), pp. 989-1013.

[19] Choi, Jay Pil, and Doh-Shin Jeon, \A Leverage Theory of Tying in Two-sided Markets with Nonnegative Price Constraints," *American Economic Journal: Microeconoimcs*, Vol. 13 (2021), pp. 283-337.

[20] Culotta, Aron, Ginger Zhe Jin, Yidan Sun, and Liad Wagman, \Safety Reviews on Airbnb: An Information Tale," (2022), working paper.

[21] Daughety, Andrew F., and Jennifer F. Reinganum, \Product Safety: Liability, R&D, and Signaling," *American Economic Review*, Vol. 85 (1995), pp. 1187-1206.

[22] Daughety, Andrew F. and Jennifer F. Reinganum, \Market, Torts, and Social Inef- ciency," *RAND Journal of Economics*, Vol. 37 (2006), pp. 300-323.

[23] Daughety, Andrew F., and Jennifer F. Reinganum, \Communicating Quality: a Uni ed Model of Disclosure and Signaling," *RAND Journal of Economics*, Vol. 39 (2008b), pp. 973-989.

[24] Daughety, Andrew F., and Jennifer F. Reinganum, \Imperfect Competition and Quality Signaling," *RAND Journal of Economics*, Vol. 39 (2008a), pp. 163-183.

[25] Dari Mattiacci, Giuseppe, and Francesco Parisi, \The Cost of Delegated Control: Vicarious Liability, Secondary Liability and Mandatory Insurance," *International Review of Law and Economics*, Vol. 23 (2003), pp. 453-475.

[26] De Chiara, Alessandro, Ester Manna, Antoni Rubi-Puig and Adrian Segura-Moreiras, \E cient Copyright Filters for Online Hosting Platforms," (2021), working paper.

[27] Dukes, Anthony, and Esther Gal-Or, \Negotiations and Exclusivity Contracts for Advertising," *Management Science*, Vol. 22 (2003), pp. 222-245.

[28] Epple, Dennis, and Artur Raviv, \Product Safety: Liability Rules, Market Structure, and Imperfect Information," *American Economic Review*, Vol. 68 (1978), pp. 80-95.

[29] Farooqi, Shehroze, Maaz Musa, Zubair Sha q, and Fareed Za ar, \Canary-Trap: Detecting Data Misuse by Third-party Apps on Online Social Networks," arXiv:2006.15794v1 [cs.CY], (2020), https://arxiv.org/pdf/2006.15794.pdf.

[30] Fu, Qiang, Jie Gong, and Ivan Png, \Law, Social Responsibility, and Outsourcing," *International Journal of Industrial Organization*, Vol. 57 (2018), pp. 114-146.

[31] Galasso, Alberto, and Hong Luo, \Tort Reform and Innovation," *Journal of Law and Economics*, Vol. 60 (2017), pp. 385-412.

[32] Galeotti, Andrea, and Jose Luis Moraga-Gonzalez, \Platform Intermediation in a Market for Di erentiated Products," *European Economic Review*, Vol. 53 (2009), pp. 417-428.

[33] Gans, Joshua S., \The Specialness of Zero," *Journal of Law and Economics*, Vol. 65 (2022), pp. 157-176.

[34] Gomes, Renato, \Optimal Auction Design in Two-Sided Markets," *RAND Journal of Economics*, Vol. 45 (2014), pp. 248-272.

[35] Hagiu, Andrei, \Pricing and Commitment by Two-Sided Platforms," *RAND Journal of Economics*, Vol. 37 (2006), pp. 720-737.

[36] Hagiu, Andrei, \Quantity vs. Quality and Exclusion by Two-Sided Platforms," (2009), working paper.

[37] Hagiu, Andrei, and Julian Wright, \Marketplace or Reseller?" *Management Science*, Vol. 61 (2015), pp. 184-203.

[38] Hagiu, Andrei, and Julian Wright, \Controlling vs. Enabling," *Management Science*, Vol. 65 (2018), pp. 577-595.

[39] Hamada, Koichi, \Liability Rules and Income Distribution in Product Liability," *American Economic Review*, Vol. 66 (1976), pp. 228-234.

[40] Hamdani, Assaf, \Who is Liable for Cyberwrongs?" *Cornell Law Review*, Vol. 87 (2002), pp. 901-957.

[41] Hamdani, Assaf, \Gatekeeper Liability," *Southern California Law Review*, Vol. 77 (2003), pp. 53-122.

[42] Harsanyi, John C., and Reinhard Selten, *A General Theory of Equilibrium Selection in Games*, MIT Press (1988).

[43] Hay, Bruce, and Kathryn E. Spier, \Manufacturer Liability for Harms Caused by Consumers to Others," *American Economic Review*, Vol.95 (2005), pp. 1700-1711.

[44] Hua, Xinyu and Kathryn E. Spier, \Product Safety, Contracts, and Liability," *RAND Journal of Economics*, Vol. 51 (2020), pp. 233-259.

[45] Jeon, Doh-Shin, Yassine Lefouili, and Leonardo Madio, \Platform Liability and Innovation," working paper, 2022.

[46] Julien, Bruno, and Alessandro Pavan, \Information Management and Pricing in Platform Markets," *Review of Economic Studies*, Vol. 86 (2019), pp. 1666-1703.

[47] Kraakman, Reinier H., \Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy," *Journal of Law, Economics, & Organization*, Vol. 2 (1986), pp. 53-104.

[48] Karle, Heiko, Martin Peitz, and Markus Reisinger, \Segmentation versus Agglomeration: Competition between Platforms with Competitive Sellers," *Journal of Political Economy*, Vol. 128 (2020), pp. 2329-2374.

[49] Landes, William M., and Richard A. Posner, \Joint and Multiple Tortfeasors: An Economic Analysis," *Journal of Legal Studies*, Vol. 9 (1980), pp. 517-555.

[50] Lefouili, Yassine, and Leonardo Madio, \The Economics of Platform Liability," *European Journal of Law and Economics*, Vol. 53 (2022), pp. 319-351.

[51] Maroni, Marta, \Mediated Transparency: The Digital Services Act and the Legitimisation of Platform Power," (2023) working paper.

[52] Nocke, Volker, Martin Peitz, and Konrad Stahl, \Platform Ownership," *Journal of the European Economic Association*, Vol. 5 (2007), pp. 1130-1160.

[53] Pitchford, Rohan, \How Liable Should a Lender Be? The Case of Judgment-Proof Firms and Environment Risk," *American Economic Review*, Vol. 85 (1995), pp. 1171-1186.

[54] Polinsky, A. Mitchell and William P. Rogerson, \Products Liability, Consumer Misperceptions, and Market Power." *Bell Journal of Economics*, Vol. 14 (1983), pp. 581-589.

[55] Rahman, K. Sabeel, \Regulating Informational Infrastructure: Internet Platforms as. Rogerson,or

al o6 Td [(the)-350(Eu)30(r)50(opean)-350(Econom)50(ic)-350(Asso)50(ciation)]TJ/F34 11.9252 0 46.488 0 Td

, Vol37 5 (1986), pp645-667581-589.

[58] Simon, Marilyn J., \Imperfect Information, Costly Litigation, and Product Quality," *Bell Journal of Economics*, Vol. 12 (1981), pp. 171-184.

[59] Sitaraman, Ganesh, \Deplatforming," *Yale Law Journal*, forthcoming.

[60] Shavell, Steven, \The Judgment Proof Problem," *International Review of Law and Economics*, Vol. 6 (1986), pp. 45-58.

[61] Spence, A. Michael, \Consumer Misperceptions, Product Failure, and Producer Liability," *Review of Economic Studies*, Vol. 44 (1977), pp. 561-572.

[62] Spence, A. Michael, \Monopoly, Quality, and Reputation," *Bell Journal of Economics*, Vol. 6 (1975), pp. 417-429.

[63] Tadelis, Steven, \Reputation and Feedback Systems in Online Platform Markets," *Annual Review of Economics*, Vol. 8 (2016), pp. 321-340.

[64] Tan, 1-4]

Appendix A

An Example of the Coordination Game. This example illustrates the idea that, given the same-side network effects, the platform finds it optimal to set a sufficiently small price (or even zero price) for the users.

Suppose that there are two potential users, 1 and 2, who independently choose whether to join the platform or not. Each user receives a private benefit, v, if and only if both users join the platform. In addition, when joining the platform, a user incurs costs, x v=2, which can include entry costs, opportunity costs, and the expected harm caused by the firms on the platform. The platform charges the same membership fee, Now we prove the remaining results in the lemma. Using the de nition of $r_H(w_s)$ in the lemma, (8) implies $e > 0$ if and only if $(\ _H \ _H w) \ (\ _H \ _L)(\bar{w} \ w_s) < 0$. This gives the condition for cases 1 and 2. Totally di erentiating (10), and using the fact the social welfare function is concave, gives $de = dw_s = \ _L = S^{00}(e) > 0$ and $de = dw_p = \ _H = S^{00}(e) > 0$. When $e > 0$ (an interior solution), increasing the level of liability for either the rm or the platform increases the platform's auditing e ort. Equation (10) implies $e > e$ if and only if $r_H(w_s) \ _H(d \ w) > 0$. This gives the condition for subcases 2(a), 2(b) and 2(c).

Proof of Proposition 1. Note that $\bar{w} < d < \frac{L}{L}$ by Assumption A1. Suppose $w_p = 0$ and $w_s \ \bar{w}$. From Lemma 1, a necessary and su cient condition for $e = 0$ is (8) or

$$_H \ _H w_s > (\ _H \ _L)(\bar{w} \ w_s):$$

Substituting for $\bar{w}$ from (5),

$$_H \ _H w_s > (\ _H \ _L) \ (\ _H \ _L)w_s;$$

which is equivalent to $w_s < \frac{L}{L}$. Since $w_s \ _{H} \ _{L}$

$^{H} \ _{L} \ _{L} \ W$

(either tt1.95 370(set)]TJ/F52 Tf 6.257 1.793 Td TdTJ/F34 11.9552 Tf8.368 -1.793 Td [(s)]TJ/F37 11.9552 Tf 8

Proof of Proposition 3. We start by showing that the platform does not charge the users (i.e. $m = 0$) if $\ {}_L \quad (\ {}_H + (1 \quad)\ {}_L)d$ is sufficiently large. To see this, first consider the scenario where the L-type firms are marginal ($w_s \quad \hat{w}$). Given the belief e and damage award $w = w_s + w_p$, a user will participate when

$$v \quad m + [\ (1 \quad e)\ {}_H + (1 \quad)\ {}_L](d \quad w):$$

The platform's equilibrium price charge to the firms is the same as in the baseline model (see Lemma 1). Thus, the platform's profits are

$$[1 \quad F(m + (\ (1 \quad e)\ {}_H + (1 \quad)\ {}_L)(d \quad w))][\hat{\phi}(\ e) + \ m] \quad c(e);$$

where $1 \quad F(\ )$ is the users' participation rate and

$$\hat{\phi}(\ e) = (1 \quad e)\ (\ {}_L \quad {}_L w_s \quad {}_H w_p) + (1 \quad)(\ {}_L \quad {}_L w): \qquad (30)$$

When $e = 0$, $w_s = 0$ and $w_p = d$, $\hat{\phi}(\ e)$ achieves the lowest value

$$\ {}_L \quad (\ {}_H + (1 \quad)\ {}_L)d;$$

which is positive by Assumption A2. Taking differentiation of the profit function with respect to m, we have

$$[1 \quad F(\ )] \quad f(\ )[\hat{\phi}(\ e) + \ m];$$

which is negative if $\hat{\phi}(\ e)$ is sufficiently large. Hence, if $\ {}_L \quad (\ {}_H + (1 \quad)\ {}_L)d$ is sufficiently large, the platform would set $m = 0$.

Next, consider the scenario where the H-type firms are marginal ($w_s > \hat{w}$). If the platform accommodates all the H-type firms, a user will participate when

$$v \quad m + [\ {}_H + (1 \quad)\ {}_L](d \quad w):$$

If the platform deters all the H-types firms by charging a larger price, a user will participate when

$$v \quad m + (1 \quad)\ {}_L(d \quad w):$$

Similar to the earlier analysis, we can show that, if $\ {}_L \quad {}_L d$ is sufficiently large, the platform would set $m = 0$.

In the remaining analysis, we maintain the assumption that $\ {}_L \quad (\ {}_H + (1 \quad)\ {}_L)d$ is sufficiently large, which also implies $\ {}_L \quad {}_L d$ is sufficiently large, such that the platform does not charge the users.

Now we prove condition (15), which highlights the potential divergence between the private and social incentives for auditing. Given w, (12) implies

$$\frac{dS(e;\hat{w})}{de} = \frac{\partial S(e;\hat{w})}{\partial e} \quad \frac{\partial S(e;\hat{w})}{\partial \hat{w}} \quad {}_H(d \quad w):$$

Using (14), if the equilibrium auditing effort is positive, then e^u satisfies

$$
\begin{aligned}
\frac{\partial (\ e^u; \hat{v})}{\partial e} &= \frac{\partial S(e^u; \hat{v})}{\partial e} + \int (\ _H \quad _L)(\hat{v} \quad w_s) \quad _H(d \quad w) \ f(v)dv \\
&= \frac{dS(e^u; \hat{v})}{de} + \int_{\hat{v}}^{\hat{v}} (\ _H \quad _L)(\hat{v} \quad w_s) \quad _H(d \quad w) \ f(v)dv + \quad _H(d \quad w)\frac{\partial S(e^u; \hat{v})}{\partial \hat{v}} \\
&= 0.
\end{aligned}
$$

Next, we show that, if $w_s < \hat{v}$, then $w_p^u > w_p$. Totally differentiating (12) with respect to w_p gives

$$
\frac{dS(e^u; \hat{v})}{dw_p} = \left[\frac{\partial S(e^u; \hat{v})}{\partial e} \quad \frac{\partial S(e^u; \hat{v})}{\partial \hat{v}} \quad _H(d \quad w)\right] \frac{\partial e^u}{\partial w_p} + \frac{\partial S(e^u; \hat{v})}{\partial \hat{v}} \frac{\partial \hat{v}}{\partial w_p}; \qquad (31)
$$

where $\frac{\partial S(e^u; \hat{v})}{\partial \hat{v}} < 0$ and $\frac{\partial \hat{v}}{\partial w_p} < 0$. Similar to the analysis in the baseline model, we can show that, given

Proof of Proposition 4. We prove two claims respectively for $w_s \quad \bar{w}$ and $w_s > \bar{w}$.

Claim 1 *: Suppose* $w_s \quad \bar{w}$*. The platform sets* $p^r = \quad _L \quad _L w_s \quad ^r(d \quad w)$ *and attracts the* H*-type rms where* $e^r = E(\ |e^r)$ *are the equilibrium posterior beliefs. Let* $e = E(\ |e\ )$*,* $^0 = E(\ |0)$*, and* $r_H(w_s) = (\ _H \quad _L)(\bar{w} \quad w_s)$*.*

1. *If* $(\ _H \quad _H d) + (\ _H \quad ^0)(d \quad w) \quad r_H(w_s)$ *then the platform does not audit,* $e^r = 0 < e$ *.*
2. *If* $(\ _H \quad _H d) + (\ _H \quad ^0)(d \quad w) < r_H(w_s)$ *then* $e^r > 0$*. The platform's auditing e ort decreases in rm liability* $de^r/dw_s < 0$ *and increases in platform liability* $de^r/dw_p > 0$*.*
   - (a) *If* $(\ _H \quad)(d \quad w) > r_H(w_s)$ *then* $0 < e^r < e$ *.*
   - (b) *If* $(\ _H \quad)(d \quad w) = r_H(w_s)$ *then* $0 < e^r = e$ *.*
   - (c) *If* $(\ _H \quad)(d \quad w) < r_H(w_s)$ *then* $0 < e \quad < e^r$*.*

Proof of Claim 1 : Since $w_s \quad \bar{w}$, it is not possible for the platform to deter the H-types without deterring the L-types, too. If the L-type is willing to participate, then the H-type also prefers to participate.

To begin,we construct values $\{e^r; p^r; t^r\}$ that maximize the platform's pro ts subject to the platform's incentive compatibility constraint and the participation constraints of the consumers and the L-type rms (as the L-type rm is marginal). Then, we will verify that these values are an equilibrium of the game.

$$\max_{\{e;p;t\}} \quad (\ e;p) = (1 \quad e) \ (p \quad _H w_p) + (1 \quad)(p \quad _L w_p) \quad c(e) \tag{36}$$

subject to

$$e = \arg\max_{e^0 \quad 0} \ (\ e^0; p) \tag{37}$$

$$_0 \quad t \quad E(\ |e)(d \quad w_s \quad w_p) \quad 0 \tag{38}$$

$$t \quad (\ _L w_s + c_L) \quad p \quad 0: \tag{39}$$

(37) is the platform's incentive compatibility constraint, (38) is the consumer's participation constraint, and (39) is the L-type rm's participation constraint. [124]

The L-type's participation constraint (39) must bind. To see this, consider two cases. First, suppose that neither (38) nor (39) binds. Then the platform would increase the price p which would increase the platform's pro ts in (36) and maintain the consumer's participation constraint (38). Second, suppose that (38) binds while (39) does not. Again, the platform would increase the price p marginally. The direct e ect of increasing p is

[124] The H -type's participation constraint is satis ed if (39) holds, and is therefore not included in the program.

that the platform's pro ts in (36) increase. Since $@^2$ ($e; p$)=$@e@p$ = < 0; increasing p also (weakly) decreases the platform's e ort e in (37), which in turn raises E(je) and, since (38) binds, reduces t. However, since t is not in (36), the platform's pro ts still increase.

Since the L-type's constraint (39) binds, $p = t$ ($_L w_s + c_L$) and we can rewrite the optimand (36) as a function of e and t:

$$(1 \quad e) \; (t \quad (\; _L w_s + c_L) \quad _H w_p) + (1 \quad)(t \quad (\; _L w_s + c_L) \quad _L w_p) \quad c(e): \qquad (40)$$

Next, we show that the consumer's participation constraint (38) binds. Suppose not. Then, the platform would increase t and its pro ts would rise. Since both participation constraints (38) and (39) bind, we have

$$p = \; _0 \quad E(\; \mathrm{j}e)(d \quad w_s \quad w_p) \quad (\; _L w_s + c_L): \qquad (41)$$

Since $_L = \; _0 \quad c_L$ and $w = w_s + w_p$ the solution to the platform's optimization problem is:

$$e^r = \arg\max_{e \; 0} \; (\; e; p^r) \qquad (42)$$

$$t^r = \; _0 \quad E(\; \mathrm{j}e^r)(d \quad w) \qquad (43)$$

$$p^r = \; _L \quad _L w_s \quad E(\; \mathrm{j}e^r)(d \quad w): \qquad (44)$$

We now verify that the values f $e^r; p^r; t^r$ g de ned in (42), (43), and (44) are an equilibrium of the game. Suppose that the platform charges p^r in (44), and that the rms and consumers believe that the probability of harm is $^r = E(\; \mathrm{j}e^r)$ where e^r is de ned in (42). The consumers are (just) willing to pay t^r in (43) and the L-type rms are (just) willing to pay p^r in (44). If the consumers and the rms all participate, the platform exerts e ort e^r in (42). Therefore the equilibrium beliefs $^r = E(\; \mathrm{j}e^r)$ are consistent.

Next, we verify that Assumption A2 guarantees that the platform's pro ts are positive. To do this, we will show that the platform's pro ts are positive even if consumers and the rms believed that the platform is not auditing at all, so $E(\; \mathrm{j}0) = \; ^0$.[125] In this scenario, the most that consumers would be willing to pay is $t = \; _0 \quad ^0(d \quad w)$ from (38). The most that the L-type rms would be willing to pay is $p = \; _L \quad _L w_s \quad ^0(d \quad w)$ from (39). The platform's pro ts can be rewritten as

$$(0) = \; _L \quad ^0 d + \; (\; _H \quad _L) w_s:$$

Therefore, (0) > 0 for any w_s 0 if Assumption $A2$ holds.[126]

[125] The platform is better o if the consumers believe that the product is safer. If consumers perceive the product to be safer, they will pay a higher price t for the product which means that the platform can charge the rms a higher price p.

[126] If $e = 1$ then $E(\; \mathrm{j}1) = \; _L$. One can verify that (1) > 0 if and only if $_L \quad _L d > \frac{c(1)}{1}$: This condition is independent of w_s and w_p. It may hold even if A2 is not satis ed (that is, $_L \quad _L d \quad (\; _H \quad _L) d$). When this condition holds, even if A2 is not satis ed, the platform may still be active. That is, A2 is a su cient but not necessary condition for the platform to be active.

We now show that the algebraic condition in case 1 is necessary and su cient for a corner solution, $e^r = 0$. We rst show the condition is necessary. If $e^r = 0$ then $E(\ j0) = \ ^0$. Since the consumer's participation constraint (38) binds we have $t^r = {}_0 \ ^0(d \ w)$; since the L-type rm's participation constraint (39) binds we have $p^r = {}_L \ {}_L w_s \ ^0(d \ w)$. Finally, for $e^r = 0$ to satisfy the platform's IC constraint (37) we need $@ (\ e; p)=@e \ 0$ or equivalently $p^r \ {}_H w_p \ 0$. Substituting p^r, this condition becomes

$$ {}_L \ {}_L w_s \ ^0(d \ w) \ {}_H w_p \ 0: \qquad (45)$$

Adding and subtracting terms this becomes

$$({}_H \ {}_H d) \ ({}_H \ {}_L) \ {}_L w_s \ {}_H w_p + {}_H w + ({}_H \ ^0)(d \ w) \ 0; \qquad (46)$$

and rearranging this expression gives

$$({}_H \ {}_H d) + ({}_H \ ^0)(d \ w) \ ({}_H \ {}_L) \ ({}_H \ {}_L) w_s: \qquad (47)$$

The right-hand side is $r_H(w_s)$. This con rms that the condition in case 1 is necessary.

Next, we show that the condition in case 1 is su cient. Suppose the condition holds and $e^r > 0$. Since $E(\ je^r) < \ ^0$, $t^r > {}_0 \ ^0(d \ w)$ and $p^r > {}_L \ {}_L w_s \ ^0(d \ w)$. Assumption A2 implies $p^r \ {}_H w_p > 0$, so the platform does not audit, $e^r = 0$.

Now consider case 2. The condition implies $p^r \ {}_H w_p < 0$ so the platform is losing money from each H-type transaction. The equilibrium e ort $e^r > 0$ and consumers' equilibrium beliefs $^r = E(\ je^r)$ satisfy equation (23). the platform charges $p^r = {}_L \ {}_L w_s \ ^r(d \ w)$ and consumers believe that the platform will exert e ort e^r beli8671dt[(+)] audit, the rms charge the consumers $t^r = \ _0 \quad {}^0(d \quad w)$. The platform's price extracts the marginal H-type rm's surplus, that is, $p^r = t^r \quad (\ _H w_s + c_H)$ or

$$p^r = \ _H \quad _H w_s \quad {}^0(d \quad w): \qquad (50)$$

The platform's pro ts are

$$\begin{aligned} p^r \quad {}^0 w_p &= (1 \quad)(\ _L \quad _L d) + \quad (\ _H \quad _H d) + (1 \quad)[\ _H \quad _L \quad (\ _H \quad _L) w_s] \\ &= (1 \quad)(\ _L \quad _L d) + \quad (\ _H \quad _H d) + (1 \quad)(\ _H \quad _L)(\hat{w} \quad w_s) \\ &< (1 \quad)(\ _L \quad _L d) \end{aligned}$$

where the inequality follows from Assumption A1 and $w_s > \hat{w}$: Therefore, if $w_s > \hat{w}$, the platform charges $p^r = \ _L \quad _L(d \quad w_p)$ and deters the H-types.

We now proceed to proof Proposition 4. Suppose $w_s \quad \hat{w}$, so the L-type is marginal. From Claim 1, we have $e^r = e$ if and only if

$$(\ _H \quad _L)(\hat{w} \quad w_s) \quad (\ _H \quad)(d \quad w) = 0: \qquad (51)$$

Substituting that $w = w_p + w_s$ and isolating w_p on the left-hand side establishes the result. Suppose $w_s > \hat{w}$. The results follow from Claim 2.

Online Appendix B

This appendix contains the analysis of four additional extensions: heterogeneous users with observable e ort, rm moral hazard, false positives and litigation costs.

B1. Heterogeneous Users with Observable E ort

Section 3 shows that platform liability can be socially desired when heterogenous users make participation decisions but do not observe the platform's auditing e ort. Now we consider the setting where the platform can commit to its auditing e ort before the users make participation decisions.

If $w_s > \hat{v}$ then the analysis is the same as case 2 in Section 3. As shown in Section 3, if $w_s > \hat{v}$, the platform would not take any auditing e ort, and imposing full residual liability on the platform implements the rst-best outcome. The following analysis examines case 1 where $w_s \quad \hat{v}$.

When auditing e ort is observable, equation (14) implies that the platform's e ort (if it is positive) satis es

$$\frac{d\ (e^u;\hat{v})}{de} = \frac{dS(e^u;\hat{v})}{de} + \int_{\hat{v}} [\ (\ _H \quad _L)(\hat{v} \quad w_s) \quad _H(d \quad w)]\, f(v)dv$$
$$_H(d \quad w)[\ (1 \quad e^u)(\ _H \quad _L)(\hat{v} \quad w_s)]\, f(\hat{v}) = 0 \quad (52)$$

where $\hat{v} \quad \hat{v}(e;w)$.

When $w_s = \hat{v}$, $\frac{d\ (e^u;\hat{v})}{de} = \frac{dS(e^u;\hat{v})}{de}$ if and only if $w_p^u = d \quad w_s$. Therefore, imposing full residual liability on the platform implements the second-best outcome: the platform chooses $e^u = e$ and all the users join the platform.

When $w_s < \hat{v}$, the last term on the right-hand side of equation (52) is negative. Moreover, if $w_p \quad w_p$, where $w_p \; 2\; (0; d \quad w_s)$ is the optimal platform liability in Proposition 2 of the baseline model, then the second term on the right-hand side of equation (52) is non-positive. Therefore, $\frac{dS(e^u;\hat{v})}{de} > 0$; that is, the platform's auditing incentive is socially insu cient. The social planner chooses w_p to maximize social welfare:

$$\frac{dS(e^u;\hat{v})}{dw_p} = \frac{dS(e^u;\hat{v})}{de}\frac{de^u}{dw_p} + \frac{@S(e^u;\hat{v})}{@\hat{v}}\frac{@\hat{v}}{@w_p}; \quad (53)$$

where $\frac{@\hat{v}}{@w_p} = \quad (\ (1 \quad e$

1. *If $w_s < w_b$, then $w_p^u > w_p$ as long as $\frac{de^u}{dw_p} > 0$. The platform sets $p^u = {}_L \quad {}_L w_s$. The second-best outcome is not achieved.*

2. *If*

may become either the L-type or H-type ex post. If a firm takes (unobservable) care with cost $k > 0$, the probability of becoming an H-type is . If the firm does not take care, the probability of being an H-type rises to $\hat{\ } >$: The platform commits to its price p before the firms decide to take care or not. The firms privately learn their realized types and decide whether to join the platform.

For simplicity, we maintain the following assumption

$$k < (\hat{\ } \quad)(\ _L \quad _L d) + \quad (\ _H \quad _H d). \tag{54}$$

Assumption (54) leads to several implications.

First, since $\ _H \quad _H d < 0$, $k < (\hat{\ } \quad)(\ _L \quad _L d)$. If the H-types never join the platform, it is socially efficient for the (ex ante identical) firms to invest k.

Second, Assumption (54) implies

$$k < (\hat{\ } \quad)[(\ _L \quad _L d) \quad (\ _H \quad _H d)] = (\hat{\ } \quad)(\ _H \quad _L)(d \quad \underline{w}).$$

Even if both types join the platform, it is efficient for the firms to invest k.

Finally, Assumption (54) implies

$$(\ _H \quad _H d) + (1 \quad)(\ _L \quad _L d) \quad k > (1 \quad \hat{\ })(\ _L \quad _L d);$$

that is, social welfare is larger if all the firms invest k and join the platform than if no firm invests and only the L-types join the platform.

In the first-best benchmark, all the firms invest k ex ante and only the L-types join the platform. Given k, there exists $w^k \in (\underline{w}; d)$ such that, if and only if $w_s > w^k$,

$$k < (\hat{\ } \quad)(\ _H \quad _L)(w_s \quad \underline{w}).$$

The pro t di erence,

$$ {}^{0} \quad {}^{L} = (b \quad)(\quad _{L} \quad _{L} w_s \quad _{L} w_p) \quad k(1 \quad)=(b \quad); $$

decreases in w_p. That is, the platform has stronger incentives to charge p_0 if w_p is lower. When $k > \frac{(b \quad)^2}{(1 \quad)}(\quad _{L} \quad _{L} w_s)$, then the platform never charges p_0, so platform liability is unnecessary. When $k \quad \frac{(b \quad)^2}{(1 \quad)}(\quad _{L} \quad _{L} w_s)$, then ${}^{0} \quad {}^{L} \quad 0$ if $w_p = 0$ but may become negative if w_p is large, so it is optimal to set $w_p = 0$.

Case 2.2: w_s 2 $(w^k; \frac{H}{H})$: Given $w_s < \frac{H}{H}$, the H-types may have incentives to join the platform. Moreover, given $w_s > w^k$, we have $k < (b \quad)(\quad _{H} \quad _{L})(w_s \quad b)$, which implies $p_0 > p_H = \quad _{H} \quad _{H} w_s > 0$. If the platform charges p_L, the rms would not invest k and the platform's pro t is

$$ {}^{L} = (1 \quad b)(\quad _{L} \quad _{L} w_s \quad _{L} w_p): $$

If the platform charges p_H, the L-types receive information rent $(\quad _{H} \quad _{L})(w_s \quad b)$. Since $k < (b \quad)(\quad _{H} \quad _{L})(w_s \quad b)$, the rms would invest k and always join the platform. Then the platform's pro t is

$$ {}^{H} = \quad (\quad _{H} \quad _{H} w_s \quad _{H} w_p) + (1 \quad)(\quad _{H} \quad _{H} w_s \quad _{L} w_p): $$

If the platform charges

w_s 2 $(w^k; \overline{w})$, only under a non-empty set of $w_p > 0$, the platform charges p_0 and the rst-best outcome is achieved.[27] That is, if w_s 2 $(w^k; \overline{w})$, platform liability is socially desired.

If $w_s = \overline{w}$, ^{0} ^{H} 0 and ^{0} ^{L} 0 only under $w_p = 0$, so it is optimal to set $w_p = 0$. If w_s 2 $(\overline{w}; \frac{H}{H})$, the platform never charges p_0. Since it is e cient for all the rms to invest k and the pro t di erence ^{H} ^{L} decreases in w_p, it is optimal to set $w_p = 0$, under which the platform charges p_H and the rms invest k.

Case 2.3: w_s 2 $(b w; w^k)$. Given $w_s < w^k$, we have $k > (b \quad)(_H \quad _L)(w_s \quad b w)$, which implies $p_0 < p_H$. If the platform charges p_L, the rms would not invest k and the platform's pro t is

$$^L = (1 \quad b)(_L \quad _L w_s \quad _L w_p):$$

If the platform charges p_H, the L-types receive information rent $(_H \quad _L)(w_s \quad b w)$. Since $k > (b \ (1 \ ))(_H \quad _L)(w$

.875 -1.793 Td[39 (L)]TJ/F34 11.9552 Tf 25.487 0 Td [518050

B3. False Positives (Type-I Errors)

Now we extend the baseline model by considering false positives. Suppose that the auditing e ort of the platform may erroneously remove the L-type rms with probability e, where < 1. The rst-best benchmark is the same as in the baseline model. For the second-best benchmark, suppose that the H-type rms seek to join the platform. Social welfare is:

$$S(e) = v + \quad (1 \quad e)($$

B4. Litigation Costs

We extend the baseline model by considering litigation costs. When a user gets harmed by a firm and files a lawsuit, the litigation costs are $z_p; z_s; z_b$, respectively for the platform, the firm, and the user. Denote $z = z_p + z_s + z_b$. Assume that $z_b \quad w_s + w_p$ and $\quad_L \quad _L d \quad z >$ 0.[128] So, litigation is credible and it is efficient to have interactions between the L-type firms and users. If the H-type firms seek to join the platform, social welfare is

$$S(e) = v + \quad (1 \quad e)(\quad_H \quad _H (d + z)) + (1 \quad)(\quad_L \quad _L (d + z)) \quad c(e):$$

The socially optimal auditing effort $\bar{e} \quad > 0$ satisfies

$$(\quad_H \quad _H (d + z)) \quad c^0(\bar{e} \quad) = 0:$$

The two types of firms have the same rent when:

$$w_s + z_s = \quad = \frac{\quad_H \quad _L}{\quad_H \quad _L}: \tag{58}$$

Case 1: $w_s + z_s \quad$. The platform sets $p^z = \quad_L \quad _L (w_s + z_s)$ to extract the L-type firms' rent. The platform chooses $e > 0$ if and only if $p^z \quad _H (w_p + z_p) < 0$, which can be rewritten as

$$\quad_H \quad _H (w + z_p + z_s) \quad (\quad_H \quad _L)(\quad w_s \quad z_s) < 0:$$

The platform's profits can be written as

$$(e) = S(e) \quad (1 \quad e) \; (\quad_H \quad _L)(\quad w_s \quad z_s)$$
$$+ [(1 \quad e) \quad _H + (1 \quad) \quad _L](d + z_b \quad w) \quad v:$$

Denote the equilibrium auditing effort as e^z. If $e^z > 0$, the first-order condition is

$$^0(e^z) = S^0(e^z) + \quad (\quad_H \quad _L)(\quad w_s \quad z_s) \quad _H (d + z_b \quad w) = 0: \tag{59}$$

The users' uncompensated loss caused by the H-types, $\quad_H (d + z_b \quad w)$, increases in z_b; and the firms' information rent, $\quad (\quad_H \quad _L)(\quad w_s \quad z_s)$, decreases in z_s. Therefore, as compared to the baseline model, the platform's auditing incentives are even weaker relative to the social incentives. We can show the following results.

Lemma 3. *Suppose* $w_s + z_s \quad$. *The platform sets* $p^z = \quad_L \quad _L (w_s + z_s)$ *and attracts the* H*-types. Let* $r^z_H(w_s) \quad (\quad_H \quad _L)(\quad w_s \quad z_s)$ *denote the* H*-types' information rents.*

1. *If* $\quad_H \quad _H (w + z_p + z_s) \quad r^z_H(w_s)$ *then the platform does not audit,* $e^z = 0 < \bar{e}$.
2. *If* $\quad_H \quad _H (w + z_p + z_s) < r^z_H(w_s)$ *then* $e^z > 0$.
   (a) *If* $\quad_H (d + z_b \quad w) > r^z_H(w_s)$ *then* $0 < e^z < \bar{e}$.

[128] We also assume that z is lower than the benefit of improved platform incentives.

(b) *If* $_H(d + z_b \quad w) = r^z_H(w_s)$ *then* $0 < e^z = \bar{e}$.

(c) *If* $_H(d + z_b \quad w) < r^z_H(w_s)$ *then* $0 < \bar{e} \quad < e^z$.

Case 2: $w_s + z_s > wb$. The platform's pro t-maximizing strategy is to either charge $p = \; _L \quad _L(w_s + z_s)$ and deter the H-types from joining the platform or charge $p = \; _H \quad _H(w_s + z_s)$ and attract both types. The platform will charge $p = \; _H \quad _H(w_s + z_s)$ and attract the H