Supplemental Appendix
Holding Platforms Liable

Xinyu Hua* Kathryn E. Spier†
HKUST Harvard University

December 8, 2024

This appendix contains the analysis of six additional extensions to the baseline model: (B1) alternative pricing structure with non-refundable fees, (B2) false positives, (B3) litigation costs, (B4) competing platforms, (B5) user participation, and (B6) firm moral hazard.

B1. Alternative Pricing Structure

Our baseline model assumed that the platform could only charge an interaction price to the firms. In this extension, assume that the platform can use two-part tariffs: a non-refundable application fee y and an interaction price p. We will show that platform liability can still be socially beneficial.

When $w_s > \widehat{w}$, the type-b firms are marginal and the platform can – but may not have incentives – to deter them by charging a high interaction price and setting $y = 0$. The analysis is the same as in the baseline model. Therefore, in this extension, we focus on the case with $w_s \leq \widehat{w}$.

Given $w_s \leq \widehat{w}$, the type-g firms are marginal and the platfrom sets $p + y = a_g - \theta_g w_s$. If a type-$b$ firm seeks to join the platform, its expected surplus is

$$\begin{aligned} & (1-e)(a_b - \theta_b w_s - p) - y \\ = \; & (1-e)(\theta_b - \theta_g)(\widehat{w} - w_s) - ey, \end{aligned}$$

*Hong Kong University of Science and Technology. xyhua@ust.hk.
†Harvard Law School and NBER. kspier@law.harvard.edu.

which decreases in e and equals $(\theta_b - \theta_g)(\widehat{w} - w_s)$ when $e = 0$.

Similar to the analysis in the baseline model, the platform will accommodate the type-b firms by setting $y = 0$ and $e = 0$ if and only if the joint benefit for the platform and firms is larger than the type-b firms' surplus.

$$a_b - \theta_b(w_s + w_p) \geq (\theta_b - \theta_g)(\widehat{w} - w_s). \tag{B1}$$

Absent platform liability ($w_p = 0$), as shown in the baseline model, the above condition holds given $w_s \leq \widehat{w}$. Therefore, if $w_p = 0$, the platform would accommodate the type-b firms.

If $w_p = d - w_s$, given $a_b - \theta_b d < 0$, condition (B1) does not hold. Thus, when w_p is sufficiently large, the platform has incentives to block or deter the type-b firms. Note that the type-b firms can be fully deterred if and only if

$$y > \frac{(1-e)(\theta_b - \theta_g)(\widehat{w} - w_s)}{e}. \tag{B2}$$

If the platform sets $y \leq \frac{(1-e)(\theta_b - \theta_g)(\widehat{w} - w_s)}{e}$, then the type-$b$ firms seek to join the platform and the analysis of the equilibrium is the same as in the baseline model.

If the platform sets $y > \frac{(1-e)(\theta_b - \theta_g)(\widehat{w} - w_s)}{e}$, then the type-$b$ firms do not join the platform. However, the platform still needs to commit to some auditing effort, because condition (B2) cannot hold when e is arbitrarily close to 0. Since $y = a_g - \theta_g w_s - p$ and the right-hand side of (B2) decreases in e, to fully deter the type-b firms and minimize the auditing cost, the platform would set $p = 0$, $y = a_g - \theta_g w_s$, and e larger than but arbitrarily close to $\underline{e}$, where $\underline{e}$ satisfies

$$a_g - \theta_g w_s = \frac{(1-\underline{e})(\theta_b - \theta_g)(\widehat{w} - w_s)}{\underline{e}},$$

or, equivalently,

$$\underline{e} = 1 - \frac{a_g - \theta_g w_s}{a_b - \theta_b w_s} > 0.$$

In general, $\underline{e}$ can be larger or smaller than e^{**}, which is the socially optimal auditing effort in the baseline model (when the type-b firms cannot be deterred by the pricing mechanism). If $\underline{e} < e^{**}$, it is socially optimal to deter the type-b firms by using a high non-refundable application fee. Imposing large platform liability (for example, $w_p = d - w_s$) motivates the platform to do so.

Proposition B1. *(Non-Refundable Fees.) Suppose $w_s \leq \widehat{w}$ and $\underline{e} < e^{**}$. If $w_p = 0$, the platform accommodates the type-b firms by choosing $y = 0$, $p = a_g - \theta_g w_s$, and $e = 0$. If $w_p = d - w_s$, the platform deters the type-b firms by choosing $y = a_g - \theta_g w_s$, $p = 0$, and $e = \underline{e} + \varepsilon$ with arbitrarily small $\varepsilon > 0$.*

B2. False Positives (Type-I Errors)

Now we extend the baseline model by considering false positives. Suppose that the auditing effort of the platform may erroneously block the type-g firms with probability ϕe, where $\phi < 1$. If the type-b firms seek to join the platform, social welfare is:

$$S(e) = v + \lambda(1-e)(\alpha_b - \theta_b d) + (1-\lambda)(1-\phi e)(\alpha_g - \theta_g d) - c(e). \tag{B3}$$

The socially optimal auditing effort $\widetilde{e}^{**}$ (if it is positive) satisfies

$$-\lambda(\alpha_b - \theta_b d) - \phi(1-\lambda)(\alpha_g - \theta_g d) - c'(\widetilde{e}^{**}) = 0. \tag{B4}$$

When $w_s > \widehat{w}$, the type-b firms are marginal and the platform would not take auditing effort. There is no type-I error. The analysis is the same as in the baseline model.

When $w_s \leq \widehat{w}$, the type-g firms are marginal. The platform sets the interaction price $p^f = \alpha_g - \theta_g w_s$, and its profits can be written as

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) + [(1-e)\lambda\theta_b + (1-\lambda)(1-\phi e)\theta_g](d - w) - v.$$

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

$$\Pi'(e^f) = S'(e^f) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - [\lambda\theta_b + (1-\lambda)\phi\theta_g](d - w) = 0. \tag{B5}$$

Note that the users' (marginal) uncompensated harm, $[\lambda\theta_b + (1-\lambda)\phi\theta_g](d-w)$, is larger than that in the baseline model, while the firms' surplus, $\lambda(\theta_b - \theta_g)(\widehat{w} - w_s)$, remains the same. Thus, the platform's incentives for auditing are weaker than in the baseline model. Hence, the optimal platform liability becomes larger as shown below (the proof is similar to that in the baseline model and omitted).

Proposition B2. *(False Positives.) The socially-optimal platform liability for harm to users, w_p^f, is as follows:*

1. *If* $w_s \le \widehat{w}$ *then* $w_p^f = d - w_s - \frac{\lambda(\theta_b-\theta_g)}{\lambda\theta_b+(1-\lambda)\phi\theta_g}(\widehat{w} - w_s) \ge w_p^*$. *The platform attracts the type-b firms and its auditing incentives are socially efficient,* $e^f = \widetilde{e}^{**}$.

2. *If* $w_s \in (\widehat{w}, \widetilde{w})$ *then there exists a threshold* $\underline{w}_p > 0$ *such that, under any* $w_p^f \in [\underline{w}_p, d - w_s]$, *the platform deters the type-b firms.*

3. *If* $w_s \ge \widetilde{w}$ *then platform liability is unnecessary. Under any* $w_p^f \in [0, d - w_s]$, *the platform deters the type-b firms.*

B3. Litigation Costs

We now 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_u, respectively for the platform, the firm, and the user. Denote $z = z_p + z_s + z_u$. Assume that $z_u \le w_s + w_p$ and $\alpha_g - \theta_g d - z > 0$.[1] So, litigation is credible and it is efficient to have interactions between the type-g firms and users. If the type-b firms seek to join the platform, social welfare is

$$S(e) = v + \lambda(1-e)(\alpha_b - \theta_b(d+z)) + (1-\lambda)(\alpha_g - \theta_g(d+z)) - c(e).$$

The socially optimal auditing effort $\overline{e}^{**} > 0$ satisfies

$$-\lambda(\alpha_b - \theta_b(d+z)) - c'(\overline{e}^{**}) = 0.$$

The two types of firms have the same surplus when:

$$w_s + z_s = \widehat{w} = \frac{\alpha_b - \alpha_g}{\theta_b - \theta_g}. \tag{B6}$$

Case 1: $w_s + z_s \le \widehat{w}$. The platform sets $p^z = \alpha_g - \theta_g(w_s + z_s)$ to extract the type-g firms' surplus. The platform chooses $e > 0$ if and only if $p^z - \theta_b(w_p + z_p) < 0$, which can be rewritten as

$$\alpha_b - \theta_b(w + z_p + z_s) - (\theta_b - \theta_g)(\widehat{w} - w_s - z_s) < 0.$$

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

The platform's profits can be written as

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s) \\ + [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](d + z_u - w) - v.$$

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

$$\text{(B7)} \qquad \Pi'(e^z) = S'(e^z) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s) - \lambda\theta_b(d + z_u - w) = 0.$$

The users' uncompensated loss caused by the type-b firms, $\lambda\theta_b(d+z_u-w)$, increases in z_u; and the firms' surplus, $\lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s)$, decreases in z_s. Therefore, as compared to the baseline model, the platform's auditing incentives are even weaker relative to the social incentives. Moreover, condition (B7) implies that $e^z = \overline{e}^{**}$ if and only if $w_p^z = d + z_u - w_s - (1 - \frac{\theta_g}{\theta_b})(\widehat{w} - w_s - z_s) \geq w_p^*$.

Case 2: $w_s + z_s > \widehat{w}$. The platform's profit-maximizing strategy is to either charge $p = \alpha_g - \theta_g(w_s + z_s)$ and deter the type-b firms from joining the platform or charge $p = \alpha_b - \theta_b(w_s + z_s)$ and attract both types. The platform will charge $p = \alpha_b - \theta_b(w_s + z_s)$ and attract the type-b firms if

$$\text{(B8)} \qquad \lambda(\alpha_b - \theta_b(w + z_s + z_p)) > (1-\lambda)(\theta_b - \theta_g)(w_s + z_s - \widehat{w}),$$

which is less likely to hold when z_s or z_p is larger. That is, the platform is more likely to deter the type-b firms when the litigation costs for the platform or the firms are larger. This also implies that the platform has stronger incentives to deter the type-b firms than in the baseline model.

Similar to the analysis in the baseline model, we can characterize the optimal platform liability.

Proposition B3. *(Litigation Costs.) There exists a threshold $\widetilde{w}^z \in (\widehat{w}, d)$. The socially-optimal platform liability for harm to users, w_p^z, is as follows:*

1. *If $w_s + z_s \leq \widehat{w}$ then $w_p^z = d + z_u - w_s - (1 - \frac{\theta_g}{\theta_b})(\widehat{w} - w_s - z_s) \geq w_p^*$. The platform attracts the type-b firms and its auditing incentives are socially efficient, $e^z = \overline{e}^{**}$.*

2. *If $w_s + z_s \in (\widehat{w}, \widetilde{w}^z)$ then there exists a threshold $\underline{w}_p^z \in (0, \underline{w}_p)$ such that, under any $w_p^z \in [\underline{w}_p^z, d - w_s]$, the platform deters the type-b firms.*

3. *If* $w_s + z_s \geq \widetilde{w}^z$ *then platform liability is unnecessary. Under any* $w_p^z \in [0, d - w_s]$*, the platform deters the type-b firms.*

When $w_s + z_s \leq \widehat{w}$, as shown earlier, the platform's auditing incentives are even weaker relative to the social incentives, as compared to the baseline model. Hence, the optimal platform liability is larger than that in the baseline model, $w_p^z \geq w_p^*$, where the inequality holds strictly if $z_b > 0$ or $w_s + z_s < \widehat{w}$.

When $w_s + z_s \in (\widehat{w}, \widetilde{w}^z)$, with litigation costs, the platform has stronger incentives to deter the type-b firms than in the baseline model. Hence, the lowest platform liability that motivates the platform to deter the type-b firms is smaller than that in the baseline model, $\underline{w}_p^z < \underline{w}_p$.

B4. Platform Competition

Now consider two competing platforms, Platform 1 and Platform 2. Users are distributed symmetrically on a Hotelling line with density $f^c(x) = f^c(1-x) > 0$ 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 $v - \tau x$ if they join Platform 1 but $v - \tau(1-x)$ if they join Platform 2, where $\tau \geq 0$ reflects the level of differentiation. Assume that v is sufficiently large and τ is not too large such that the market is fully covered. The firms can join both platforms, while each user only joins one platform.[2] Thus, the platforms compete for users but not for firms.

In stage 1, the platforms simultaneously set interaction prices p_j and commit to their audit intensities e_j, $j = 1, 2$. Suppose that the auditing effort is per interaction and the users observe auditing effort before deciding which platform to join.[3] The timing and the other assumptions are otherwise identical to the baseline model. 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 be socially beneficial in this competitive environment.

Case 1: $w_s \leq \widehat{w}$. The platforms set $p^c = \alpha_g - \theta_g w_s$, which attracts the type-b firms.

[2] In practice, many users choose single-homing due to switching costs or same-side network effects.

[3] The results hold qualitatively if auditing costs are per firm and the platforms are sufficiently differentiated (i.e., τ is not too small). With per firm auditing costs, it would be socially efficient to have two platforms if τ is large but efficient to have one platform if τ is small, due to large economies of scale in auditing.

Denote the location of the indifferent user as $\widehat{x}$. If $\widehat{x} \in [0,1]$, then it satisfies

$$\begin{aligned} & v - \tau\widehat{x} - [\lambda(1-e_1)\theta_b + (1-\lambda)\theta_g](d-w) \\ = \; & v - \tau(1-\widehat{x}) - [\lambda(1-e_2)\theta_b + (1-\lambda)\theta_g](d-w), \end{aligned}$$

or equivalently,

$$\widehat{x} = \frac{1}{2} + \frac{\lambda(e_1-e_2)\theta_b(d-w)}{2\tau}.$$

If $w_p = d - w_s$ then $\widehat{x} = \frac{1}{2}$. The users are fully compensated for any harm. Similar to the analysis in the baseline model, the platforms over-invest in auditing.

If $w_p < d - w_s$, given e_2, Platform 1 can attract all the users ($\widehat{x} = 1$) by choosing $e_1 \geq \overline{e}_1$, where

$$\overline{e}_1 = e_2 + \frac{\tau}{\lambda\theta_b(d-w)}.$$

When $\tau \to 0$, $\overline{e}_1 \to e_2$, so Platform 1 would raise its auditing effort slightly to attract all the users as long as its profit is positive. When $\tau \to \infty$, $\overline{e}_1 \to \infty$, so Platform 1 would not be able to capture the whole market. Hence, there exist two thresholds $\underline{\tau}$ and $\overline{\tau}$, with $0 < \underline{\tau} \leq \overline{\tau}$, such that both platforms get positive profits if $\tau > \overline{\tau}$ while they get zero profits if $\tau < \underline{\tau}$. We consider these two cases separately.

First, suppose $\tau > \overline{\tau}$. In this case, competition is not fierce and $\widehat{x} \in (0,1)$. Platform 1 chooses e_1 to maximize its profit

$$F^c(\widehat{x})[(1-e_1)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e_1)],$$

where $F^c(\widehat{x})$ is the number of users choosing Platform 1. The profit-maximizing auditing effort by Platform 1, e_1^c (if it is positive), satisfies

$$\begin{aligned} 0 \;=\; & -F^c(\widehat{x})[\lambda(p^c - \theta_b w_p) + c'(e_1^c)] \\ & + f^c(\widehat{x})\frac{\lambda\theta_b(d-w)}{2\tau}[(1-e_1^c)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e_1^c]. \end{aligned} \tag{B9}$$

In the symmetric equilibrium with $F^c(\widehat{x}) = \frac{1}{2}$ and $e_1^c = e_2^c = e^c$, this can be rewritten as

$$\begin{aligned} 0 \;=\; & \frac{1}{2}S'(e^c) + \frac{1}{2}[\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - \lambda\theta_b(d-w)] \\ & + f^c(\widehat{x})\frac{\lambda\theta_b(d-w)}{2\tau}[(1-e^c)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e^c)], \end{aligned} \tag{B10}$$

where the last term captures the competition effect. If $w_p > w_p^*$, as shown in the baseline model, the second term on the right-hand side of (B10) is positive while the last term is non-negative, so the platforms over-invest in auditing, $e^c > e^{**}$. If $w_p = w_p^*$, the second term becomes 0 while the last term is positive if $e^c = e^{**}$, so the platforms over-invest in auditing, $e^c > e^{**}$. Finally, if $w_p = 0$ and $\tau \to \infty$, similar to the analysis in the baseline model, $e^c \to 0$. By continuity, there exists a unique threshold $\widehat{\tau} \geq \overline{\tau}$ such that $e^c < e^{**}$ if $\tau > \widehat{\tau}$ and $w_p = 0$. These observations imply that, given $\tau > \widehat{\tau}$, there exists $\widehat{w}_p \in (0, w_p^*)$ under which $e^c = e^{**}$. Hence, the optimal platform liability is $w_p^c = \widehat{w}_p < w_p^*$, which motivates the platform to choose the socially efficient auditing effort. Competition raises the platforms' auditing incentives, so that the optimal platform liability is less than in the baseline model.

Next, suppose $\tau < \underline{\tau}$. Given fierce competition, the platforms invest to the point where profits are dissipated,

$$(1 - e^c)\lambda(p^c - \theta_b w_p) + (1 - \lambda)(p^c - \theta_g w_p) - c(e^c) = 0. \tag{B11}$$

If $w_p = 0$ then platform safety is socially excessive, $e^c > e^{**}$. Absent platform liability, the platforms take too much auditing effort. Equation (B11) also implies $\frac{de^c}{dw_p} < 0$. Therefore, if $\tau < \underline{\tau}$, platform liability mitigates the over-investment problem and raises social welfare.

Case 2: $\boldsymbol{w_s > \widehat{w}}$. In this case, the type-b firms are marginal. The platforms have a choice: they can either charge the firms $p = \alpha_g - \theta_g w_s$ and deter the type-b firms or charge the firms $p = \alpha_b - \theta_b w_s < \alpha_g - \theta_g w_s$ and attract both types. As shown in the baseline model, when $w_s \geq \widetilde{w} > \widehat{w}$, a monopoly platform has incentives to charge the high price and deter the type-b firms. With competition, a platform can attract more users by deterring the type-b firms, because the users observe the prices and prefer to join a safer platform. Therefore, given $w_s \geq \widetilde{w}$, both platforms deter the type-b firms. As in the baseline model, platform liability is unnecessary.

Now suppose $w_s \in (\widehat{w}, \widetilde{w})$. If $w_p = d - w_s$, the users would be fully compensated for any harm and therefore each platform attracts half of the users. Each platform charges the high price and deter the type-b firms if

$$\frac{1}{2}(1 - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) > \frac{1}{2}[\alpha_b - \theta_b w_s - (\lambda\theta_b + (1 - \lambda)\theta_g)w_p],$$

which holds given $\alpha_b - \theta_b d < 0$. Hence, imposing full residual liability on the platforms gets the platforms to raise the interaction price and deter the type-b firms.

We now show that platform liability is necessary when $w_s \in (\widehat{w}, \widetilde{w})$ and τ is sufficiently large. Suppose to the contrary that, under $w_p = 0$, the platforms charge $p = \alpha_g - \theta_g w_s$ and deter the type-b firms. Each platform's profit is $(1-\lambda)(\alpha_g - \theta_g w_s)/2$. If Platform 1 deviates to $p = \alpha_b - \theta_b w_s$, the indifferent user's location $\widehat{x}$ satisfies

$$\tau\widehat{x} + [\lambda\theta_b + (1-\lambda)\theta_g](d - w_s) = \tau(1-\widehat{x}) + (1-\lambda)\theta_g(d - w_s),$$

that is,

$$\widehat{x} = \frac{1}{2} - \frac{\lambda\theta_b(d - w_s)}{2\tau}.$$

Accordingly, Platform 1's profit from deviation is

$$F^c\Big(\max\Big\{0, \frac{1}{2} - \frac{\lambda\theta_b(d - w_s)}{2\tau}\Big\}\Big)(\alpha_b - \theta_b w_s), \tag{B12}$$

which goes to 0 when $\tau \to 0$ and goes to $(\alpha_b - \theta_b w_s)/2$ when $\tau \to \infty$. Note that $(1-\lambda)(\alpha_g - \theta_g w_s) < (\alpha_b - \theta_b w_s)$ given $w_s \in (\widehat{w}, \widetilde{w})$. Hence, there exists a threshold $\widetilde{\tau} > 0$ such that, absent platform liability, both platforms deter the type-b firms if and only if $\tau \leq \widetilde{\tau}$. If $\tau > \widetilde{\tau}$, platform liability is socially desired. If $\tau \leq \widetilde{\tau}$, platform liability is unnecessary. Since the price that the platforms charge is observed by users, and the platforms are not highly differentiated, the users will prefer to join a platform that completely deters the harmful type-b firms.

Proposition B4. *(Platform Competition with Observable Effort.) The socially-optimal liability for the competing platforms, w_p^c, is as follows.*

1. *If $w_s \leq \widehat{w}$, there exist $\widehat{\tau}$ and $\underline{\tau}$ with $0 < \underline{\tau} \leq \widehat{\tau}$: when $\tau > \widehat{\tau}$, $w_p^c \in (0, w_p^*)$ motivates the platforms to choose the socially efficient auditing effort; when $\tau < \underline{\tau}$, $w_p^c > 0$ mitigates the over-investment problem and raises social welfare.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$, there exists $\widetilde{\tau} > 0$: when $\tau > \widetilde{\tau}$, $w_p^c = d - w_s$ motivates the platforms to deter the type-b firms; when $\tau \leq \widetilde{\tau}$, platform liability is unnecessary and the platforms deter the type-b firms under any $w_p^c \in [0, d - w_s]$.*

3. *If $w_s > \widetilde{w}$, platform liability is unnecessary. Under any $w_p^c \in [0, d - w_s]$, the platforms deter the type-b firms.*

B5. User Participation

Suppose that the users' valuations of the quasi-public good are drawn from density $f^u(v) > 0$ for $v \in [0, \infty)$, with cumulative density $F^u(v)$.[4] As in the baseline model, the platform charges the firms price p per interaction and takes auditing effort e per firm. The users have the option to join the platform for free.[5]

Assumption A2 implies that it is socially efficient for all users to participate and assumption A1 implies that it is socially inefficient for the type-b firms to participate. As in the baseline model, full deterrence of the type-b firms may not be possible. If the type-b firms seek to join the platform, social welfare is

$$S(e, \widehat{v}) = \int_{\widehat{v}}^{\infty} [v + \lambda(1-e)(\alpha_b - \theta_b d) + (1-\lambda)(\alpha_g - \theta_g d)] f^u(v)dv - c(e), \tag{B13}$$

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

$$\widehat{v}(e, w) = (\lambda(1-e)\theta_b + (1-\lambda)\theta_g)(d - w). \tag{B14}$$

Notice that $\widehat{v}(e, w)$ 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.

Case 1: $\boldsymbol{w_s \leq \widehat{w}}$. The type-$g$ firms are marginal and the platform charges $p^u = \alpha_g - \theta_g w_s$. The platform's profit function can be written as:

$$\begin{aligned}\Pi(e, \widehat{v}) = S(e, \widehat{v}) + \int_{\widehat{v}}^{\infty} \Big\{ &-(1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) \\ &+ ((1-e)\lambda\theta_b + (1-\lambda)\theta_g)(d-w) - v \Big\} f^u(v)dv,\end{aligned} \tag{B15}$$

[4] This framework is equivalent to the model where users decide how much time (T) to spend on the platform. The user's marginal value decreases in T. At each moment, the user is randomly matched with a firm and may be harmed. Intuitively, when platform liability increases and/or the platform raises audit intensity, the user spends more time.

[5] The platform might also charge a membership fee $m \geq 0$ to each user. However, it can be shown that $m = 0$ in equilibrium if $\alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)d$ is sufficiently large (that is, if cross-side network effects are strong). We maintain the assumption that $\alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)d$ is sufficiently large such that the platform does not charge the users.

Since users observe the auditing effort, the platform's effort (if it is positive) satisfies

$$\text{(B16)}\quad \frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}}^{\infty} [\lambda(\theta_b-\theta_g)(\widehat{w}-w_s) - \lambda\theta_b(d-w)] f^u(v)dv$$
$$- \lambda\theta_b(d-w)[\lambda(1-e^u)(\theta_b-\theta_g)(\widehat{w}-w_s)] f^u(\widehat{v}) = 0$$

where $\widehat{v} \equiv \widehat{v}(e,w)$.

When $w_s = \widehat{w}$, $\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de}$ if and only if $w_p^u = d - w_s$. Therefore, imposing full residual liability on the platform motivates the platform to choose $e^u = e^{**}$ and attracts all the users to join the platform.

When $w_s < \widehat{w}$, the last term on the right-hand side of equation (B16) is negative. Moreover, if $w_p \leq w_p^*$, where $w_p^* \in (0, d-w_s)$ is the optimal platform liability in the baseline model, then the second term on the right-hand side of equation (B16) is non-positive. Therefore, $\frac{dS(e^u,\widehat{v})}{de} > 0$, that is, the platform's auditing incentive is socially insufficient. The social planner chooses w_p to maximize social welfare:

$$\text{(B17)}\qquad \frac{dS(e^u,\widehat{v})}{dw_p} = \frac{dS(e^u,\widehat{v})}{de}\frac{de^u}{dw_p} + \frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p},$$

where $\frac{\partial \widehat{v}}{\partial w_p} = -(\lambda(1-e^u)\theta_b + (1-\lambda)\theta_g) < 0$. Since $\frac{\partial S(\cdot)}{\partial \widehat{v}} < 0$, the last term in (B17), $\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}$, is non-negative. Intuitively, given the auditing effort, platform liability stimulates user participation and therefore raises social welfare. Moreover, as shown earlier, $\frac{dS(e^u,\widehat{v})}{de} > 0$ if $w_p \leq w_p^*$. Hence, as long as $\frac{de^u}{dw_p} > 0$, it is socially optimal to set $w_p^u > w_p^*$.

Case 2: $\boldsymbol{w_s > \widehat{w}}$. In this case, type-$b$ firms are marginal. First, suppose $w_s \geq \widetilde{w}$, where $\widetilde{w}$ is defined in the baseline model. The platform charges $p^u = \alpha_g - \theta_g w_s$, which deters all of the type-b firms. Anticipating that the type-b firms are fully deterred, the users participate if $v \geq (1-\lambda)\theta_g(d-w)$. Hence, all the users participate when $w_p = d - w_s$. Second, suppose $w_s \in (\widehat{w}, \widetilde{w})$. As shown in the baseline model, given $w_p \geq \underline{w}_p$, the platform charges $p^u = \alpha_g - \theta_g w_s$, which deters all of the type-b firms. Again, setting $w_p = d - w_s$ attracts all the users.

Proposition B5. *(User Participation with Observable Effort.) The socially-optimal platform liability for harm to users, w_p^u, is as follows:*

1. *If $w_s < \widehat{w}$, then $w_p^u > w_p^*$ as long as $\frac{de^u}{dw_p} > 0$. The platform's auditing effort is not socially optimal.*

2. *If* $w_s = \widehat{w}$, *then* $w_p^u = d - w_s$. *The platform chooses the socially optimal auditing effort* $e^u = e^{**}$ *and all users participate.*

3. *If* $w_s > \widehat{w}$, *then* $w_p^u = d - w_s$. *The platform deters the type-b firms and all users participate.*

As in the baseline model, platform liability motivates the platform to take auditing effort or set high interaction prices to block or deter risky firms. When users are heterogeneous, platform liability has the additional benefit in stimulating user participation.

Example: Uniform Distribution. In case 1 of Proposition B5, the socially optimal platform liability is larger than that in the baseline model as long as the equilibrium auditing effort increases in w_p. Recall that, in the baseline model, the equilibrium effort always increases in w_p. However, in this extension, the equilibrium effort may increase or decrease in w_p. For illustration, suppose that v follows the uniform distribution on $[0, \overline{v}]$. Then with observable effort, the platform's effort (if it is positive) satisfies

$$\begin{aligned}\frac{d\Pi(e^u, \widehat{v})}{de} &= -c'(e^u) - \lambda(\alpha_g - \theta_g w_s - \theta_b w_p)\Big[1 - \frac{\widehat{v}}{\overline{v}}\Big] \\ &\quad +\lambda\theta_b(d - w)\Big[\lambda(1 - e^u)(\alpha_g - \theta_g w_s - \theta_b w_p) + (1 - \lambda)(\alpha_g - \theta_g w)\Big]\frac{1}{\overline{v}} \\ &= 0,\end{aligned}$$

which implies

$$\begin{aligned}\frac{d^2\Pi(e^u, \widehat{v})}{dedw_p} &= \frac{\lambda}{\overline{v}}\Big\{\overline{v} - (\lambda(1 - e^u)\theta_b \\ &\quad +(1 - \lambda)\theta_g)\Big[(1 + \beta)\theta_b(d - w) + \alpha_g - \theta_g w_s - \theta_b w_p\Big] \\ &\quad -\theta_b\Big[(1 - e^u)\lambda(\alpha_g - \theta_g w_s - \theta_b w_p) + (1 - \lambda)(\alpha_g - \theta_g w)\Big]\Big\}.\end{aligned}$$

If $\overline{v}$ is very small and $w_p = 0$ then $\frac{d^2\Pi(e^u,\widehat{v})}{dedw_p} < 0$ and, accordingly, $\frac{de^u}{dw_p} < 0$. By contrast, if $\overline{v}$ is sufficiently large then for any $w_p \leq w_p^*$ we have $\frac{d^2\Pi(e^u,\widehat{v})}{dedw_p} > 0$ and, accordingly, $\frac{de^u}{dw_p} > 0$. Intuitively, given the participation threshold, an increase in platform liability raises the marginal profit from auditing effort; at the same time, the increase in platform liability decreases the participation threshold, which in turn reduces the marginal profit from auditing effort. The former effect dominates when $\overline{v}$ is sufficiently large.

To summarize, even if the heterogeneous users observe the auditing effort and choose whether to join the platform or not, platform liability can be socially desired. The optimal platform liability is (weakly) larger than in the baseline model, as long as the equilibrium effort increases in w_p, which holds when v follows the uniform distribution on $[0, \overline{v}]$ with sufficiently large $\overline{v}$.

B6. Firm Moral Hazard

The baseline model assumes that the firms' types are exogenously given. Platform liability can still be socially beneficial if the firms' types are endogenous and the firms can take effort to improve safety. In this section, suppose all the firms are identical ex ante but may become either the type-g or type-b ex post. If a firm takes (unobservable) care with cost $c > 0$, the probability of becoming type-b is λ. If the firm does not take care, the probability of being type-b rises to $\widehat{\lambda} > \lambda$. 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

$$c < (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g d) + \lambda(\alpha_b - \theta_b d). \tag{B18}$$

Assumption (B18) leads to several implications.

First, since $\alpha_b - \theta_b d < 0$, $c < (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g d)$. If the type-$b$ firms never join the platform, it is socially efficient for the (ex ante identical) firms to invest c.

Second, Assumption (B18) implies

$$c < (\widehat{\lambda} - \lambda)[(\alpha_g - \theta_g d) - (\alpha_b - \theta_b d)] = (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(d - \widehat{w}).$$

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

Finally, Assumption (B18) implies

$$\lambda(\alpha_b - \theta_b d) + (1 - \lambda)(\alpha_g - \theta_g d) - c > (1 - \widehat{\lambda})(\alpha_g - \theta_g d),$$

that is, social welfare is larger if all the firms invest c and join the platform than if no firm invests and only the type-g firms join the platform.

In the first-best benchmark, all the firms invest c ex ante and only the type-g firms join the platform. Given c, there exists $w^m \in (\widehat{w}, d)$ such that, if and only if $w_s > w^m$,

$$c < (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w}).$$

Case 1: $w_s \leq \widehat{w}$. The type-g firms are marginal. The platform charges $p = \alpha_g - \theta_g w_s$. Since the type-$g$ firms do not have any surplus, ex ante the firms have no incentive to take care. As in the baseline model, $w_p^m = w_p^* \in (0, d - w_s]$ motivates the platform to choose the socially optimal auditing effort.

Case 2: $w_s > \widehat{w}$. The type-b firms are marginal. Consider three scenarios.

Case 2.1: $w_s > \frac{\alpha_b}{\theta_b}$. Then the type-$b$ firms would never join the platform. The platform either charges $p_g = \alpha_g - \theta_g w_s$, under which the firms would not invest c, or charges p_0, where

$$p_0 = \alpha_g - \theta_g w_s - c/(\widehat{\lambda} - \lambda) > 0,$$

under which the firms would invest c. Social welfare is larger if the platform charges p_0. The platform's profit under p_g is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p);$$

while its profit under p_0 is

$$\Pi^0 = (1 - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1 - \lambda)/(\widehat{\lambda} - \lambda).$$

The profit difference,

$$\Pi^0 - \Pi^g = (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1 - \lambda)/(\widehat{\lambda} - \lambda),$$

decreases in w_p. That is, the platform has stronger incentives to charge p_0 if w_p is lower. When $c > \frac{(\widehat{\lambda}-\lambda)^2}{(1-\lambda)}(\alpha_g - \theta_g w_s)$, then the platform never charges p_0, so platform liability is unnecessary. When $c \leq \frac{(\widehat{\lambda}-\lambda)^2}{(1-\lambda)}(\alpha_g - \theta_g w_s)$, then $\Pi^0 - \Pi^g \geq 0$ 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: $w_s \in (w^m, \frac{\alpha_b}{\theta_b})$**.** Given $w_s < \frac{\alpha_b}{\theta_b}$, the type-$b$ firms may have incentives to join the platform. Moreover, given $w_s > w^m$, we have $c < (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, which implies $p_0 > p_b = \alpha_b - \theta_b w_s > 0$. If the platform charges p_g, the firms would not invest c and the platform's profit is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p).$$

If the platform charges p_b, the type-g firms' surplus is $(\theta_b - \theta_g)(w_s - \widehat{w})$. Since $c <$

$(\widehat{\lambda}-\lambda)(\theta_b-\theta_g)(w_s-\widehat{w})$, the firms would invest c and always join the platform. Then the platform's profit is

$$\Pi^b = \lambda(\alpha_b - \theta_b w_s - \theta_b w_p) + (1-\lambda)(\alpha_b - \theta_b w_s - \theta_g w_p).$$

If the platform charges p_0, the firms would invest c but the type-b firms would not join the platform. Then the platform's profit becomes

$$\Pi^0 = (1-\lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda).$$

Note that

$$\Pi^0 - \Pi^b = (1-\lambda)(\theta_b-\theta_g)(w_s-\widehat{w}) - \lambda(\alpha_b - \theta_b w_s - \theta_b w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda)$$

increases in w_p, while

$$\Pi^0 - \Pi^g = (\widehat{\lambda}-\lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda)$$

decreases in w_p. It can be verified that, when $w_s = w^m$, $\Pi^0 - \Pi^b \geq 0$ if and only if $w_p \geq (\alpha_b - \theta_b w_s)/\theta_b > 0$, and $\Pi^0 - \Pi^g \geq 0$ if $w_p = (\alpha_b - \theta_b w_s)/\theta_b$ and

$$\Big(1 - \frac{\widehat{\lambda}-\lambda}{1-\lambda}\Big)(\alpha_g - \theta_g w_s) \leq \Big(1 - \frac{\theta_g(\widehat{\lambda}-\lambda)}{\theta_b(1-\lambda)}\Big)(\alpha_b - \theta_b w_s),$$

which holds if θ_g is close to 0 and $\widehat{\lambda}$ is close to 1. Moreover, given $w_s \in (w^m, \frac{\alpha_b}{\theta_b})$, if there exists $w_p > 0$ under which $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$, then for any $w_s' = w_s + \varepsilon$ with arbitrarily small $\varepsilon > 0$, $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$ if platform liability is set at $w_p' = w_p - \varepsilon > 0$. Hence, there exists a unique threshold $\overline{w} \in [w^m, \frac{\alpha_b}{\theta_b}]$ such that, given $w_s \in (w^m, \overline{w})$, only under a non-empty set of $w_p > 0$, the platform charges p_0 and the first-best outcome is achieved.[6] That is, if $w_s \in (w^m, \overline{w})$, platform liability is socially desired.

If $w_s = \overline{w}$, $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$ only under $w_p = 0$, so it is optimal to set $w_p = 0$. If $w_s \in (\overline{w}, \frac{\alpha_b}{\theta_b})$, the platform never charges p_0. Since it is efficient for all the firms to invest c and the profit difference $\Pi^b - \Pi^g$ decreases in w_p, it is optimal to set $w_p = 0$, under which the platform charges p_b and the firms invest c.

[6] Note that $\overline{w}$ may equal w^m or $\frac{\alpha_b}{\theta_b}$ under certain parameter values.

Case 2.3: $w_s \in (\widehat{w}, w^m)$. Given $w_s < w^m$, we have $c > (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, which implies $p_0 < p_b$. If the platform charges p_g, the firms would not invest c and the platform's profit is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p).$$

If the platform charges p_b, the type-g firms' surplus is $(\theta_b - \theta_g)(w_s - \widehat{w})$. Since $c > (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, the firms would not invest c but always join the platform. The platform's profit is

$$\Pi^b = \widehat{\lambda}(\alpha_b - \theta_b w_s - \theta_b w_p) + (1 - \widehat{\lambda})(\alpha_b - \theta_b w_s - \theta_g w_p).$$

If the platform charges $p_0 < p_b$, the firms would invest c and join the platform, so the platform's profit becomes

$$\Pi^0 = \alpha_g - \theta_g w_s - c/(\widehat{\lambda} - \lambda) - [\lambda\theta_b + (1 - \lambda)\theta_g]w_p.$$

When $w_p = 0$, it can be verified that $\Pi^b > \Pi^g$ and $\Pi^b > \Pi^0$, that is, the platform would charge p_b and the firms do not invest c but join the platform. Similar to the analysis in the baseline model, with full residual liability ($w_p = d - w_s$), the platform's profit is larger under p_g than under p_b, so the platform may charge either p_0 or p_g. Under either price, social welfare is larger than under p_b. Hence, given $w_s \in (\widehat{w}, w^m)$, platform liability is socially desired.

Summarizing the above analysis, we have

Proposition B6. *(Firm Moral Hazard.) Suppose that firm liability is $w_s \in [0, d]$ and the firms can take effort with costs c. The socially-optimal liability, w_p^m, is as follows:*

1. *If $w_s \leq \widehat{w}$, it is optimal to set $w_p^m = w_p^* \in (0, d - w_s]$. The platform charges $p^m = \alpha_g - \theta_g w_s$ and takes auditing effort e^{**}. The firms do not invest c.*
2. *If $w_s \in (\widehat{w}, \overline{w})$, it is optimal to set $w_p^m > 0$. The firms invest c if $w_s \in (w^m, \overline{w})$.*
3. *If $w_s \geq \overline{w}$, either platform liability is unnecessary or it is optimal to set $w_p^m = 0$.*