Online Appendix

A General Solution

A.1 Derivation of (6)

The marginal user satisfies $u_t^A(\hat{z}_t) = u_t^B(\hat{z}_t)$. Using (3), the marginal user therefore satisfies

$$Y_t^A - p_t^A - \kappa^A(\hat{z}_t) + I^A\theta^A(q^A - c^A)\mathbb{I}_{\{t=1\}} = Y_t^B - p_t^B - \kappa^B(\hat{z}_t) + I^B\theta^B(q^B - c^B)\mathbb{I}_{\{t=1\}}.$$

Next, we use Y_t^x from (2), and $\kappa^x(z)$ as well as $N_t^A = \hat{z}_t$ and $N_t^B = 1 - \hat{z}_t$ from (1), to obtain

$$\begin{aligned} &K^A + \phi^A D^A \ \ \mathbb{I}_{\{t=2\}} + \gamma^A \hat{z}_t - p_t^A + I^A\theta^A(q^A - c^A)\mathbb{I}_{\{t=1\}} - \hat{\kappa}\hat{z}_t \\ =&K^B + \phi^B D^B \ \ \mathbb{I}_{\{t=2\}} + \gamma^B(1 - \hat{z}_t) - p_t^B + I^B\theta^B(q^B - c^B)\mathbb{I}_{\{t=1\}} - \hat{\kappa}(1 - \hat{z}_t). \end{aligned}$$

Now, we can solve above equation for

$$\hat{z}_t = \frac{1}{2} + \frac{\Delta_K - (p_t^A - p_t^B) + \left[\phi^A D^A - \phi^B D^B\right]\mathbb{I}_{\{t=2\}} + \left[I^A\theta^A(q^A - c^A) - I^B\theta^B(q^B - c^B)\right]\mathbb{I}_{\{t=1\}}}{2\kappa}$$

with

$$\kappa := \hat{\kappa} - \frac{\gamma^A + \gamma^B}{2} \quad \text{and} \quad \Delta_K := K^A - K^B + \frac{\gamma^A - \gamma^B}{2},$$

which was to show.

A.2 Solution in Period 2 — Proof of Lemma 1

Take the marginal user $\hat{z}_2$ from period $t = 2$ in (6), that is,

$$\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B - (p_2^A - p_2^B)}{2\kappa}.$$

Thus, $N_2^A = \hat{z}_2$ and $N_2^B = 1 - \hat{z}_2$. Platform x chooses price p_2^x to maximize $N_2^x p_2^x$. Hence, the objective of platform A becomes

$$\pi_2^A = p_2^A \left(\frac{1}{2} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B - (p_2^A - p_2^B)}{2\kappa}\right).$$

The first-order condition with respect to price p_2^A reads

$$\frac{\partial \pi_2^A}{\partial p_2^A} = \frac{1}{2} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B - (p_2^A - p_2^B)}{2\kappa} - \frac{p_2^A}{2\kappa} = 0,$$

which we solve for

$$p_2^A = \frac{\Delta_K + \kappa + \phi^A D^A - \phi^B D^B + p_2^B}{2}. \tag{A.1}$$

Next, the objective of platform B becomes

$$\pi_2^B = p_2^B \left(\frac{1}{2} + \frac{\phi^B D^B - \phi^A D^A - \Delta_K - (p_2^B - p_2^A)}{2\kappa}\right).$$

The first-order condition with respect to price p_2^B reads

$$\frac{\partial \pi_2^B}{\partial p_2^B} = \frac{1}{2} + \frac{\phi^B D^B - \phi^A D^A - \Delta_K - (p_2^B - p_2^A)}{2\kappa} - \frac{p_2^B}{2\kappa} = 0,$$

which we solve for

$$p_2^B = \frac{\kappa - \Delta_K + \phi^B D^B - \phi^A D^A + p_2^A}{2}. \tag{A.2}$$

It is immediate to see that the second order conditions are satisfied, i.e., $\frac{\partial^2 \pi_2^x}{\partial (p_2^x)^2} < 0$.

Inserting (A.2) into (A.1), we solve

$$p_2^A = \kappa + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{3} = \frac{3\kappa + \Delta_K + \phi^A D^A - \phi^B D^B}{3}. \tag{A.3}$$

We now plug (A.3) into (A.2) to calculate

$$p_2^B = \kappa + \frac{-\Delta_K + \phi^B D^B - \phi^A D^A}{3} = \frac{3\kappa - \Delta_K + \phi^B D^B - \phi^A D^A}{3}. \tag{A.4}$$

Thus, $p_2^A - p_2^B = \frac{2}{3}(\Delta_K + D^A\phi^A - D^B\phi^B)$ and therefore

$$\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{6\kappa} = \frac{3\kappa + \Delta_K + \phi^A D^A - \phi^B D^B}{6\kappa}.$$

Thus, we obtain $p_2^A = 2\kappa\hat{z}_2 = 2\kappa N_1^A$ and $p_2^B = 2\kappa(1-\hat{z}_2) = 2\kappa N_2^B$.

Next, calculate the profit of platform A, i.e.,

$$\pi_2^A = p_2^A \hat{z}_2 = \frac{(3\kappa + \phi^A D^A - \phi^B D^B + \Delta_K)^2}{18\kappa},$$

and the profit of platform B, i.e.,

$$\pi_2^B = p_2^B(1-\hat{z}_2) = \frac{(3\kappa + \phi^B D^B - \phi^A D^A - \Delta_K)^2}{18\kappa}.$$

Or, we can write more compactly

$$\pi_2^x = \frac{(\Delta_K - 2\Delta_K \mathbb{I}_{\{x=B\}} + 3\kappa + \phi^x D^x - \phi^{-x} D^{-x})^2}{18\kappa}, \tag{A.5}$$

for $x, -x \in \{A, B\}$ (with $x = A$ implying $-x = B$ and vice versa).

A.3 Solution in Period 1 — Proof of Proposition 1

The first order conditions (8) and (9) follow from the directly differentiating π_1^x from (5) with respect to p_1^x and I^x respectively. The second order condition with respect to price reads

$$\frac{\partial^2 \pi_1^x}{\partial (p_1^x)^2} = 2\left(\frac{\partial N_1^x}{p_1^x}\right) + \frac{\partial}{\partial p_1^x} \sum_{x'=A,B} \left(\frac{\partial \pi_2^x}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x}\right) < 0.$$

The second order condition with respect to investment can be written as

$$\frac{\partial^2 \pi_1^x}{\partial (I_1^x)^2} = \mathcal{K} - \lambda,$$

where $\mathcal{K}$ is a finite term with generally unknown sign. It follows that $\frac{\partial^2 \pi_1^x}{\partial (I_1^x)^2} = \mathcal{K} - \lambda < 0$ as long as λ is sufficiently large. Thus, the second order condtions are met and the first order conditions are sufficient as long as λ is sufficiently large.

We next prove that inducing $\theta^x = 1$ is optimal for platforms, and strictly so when $I^x > 0$. The

case $\underline{\theta} = 1$ is trivial; hence, consider $\underline{\theta} < 1$. When $I^x = 0$, then the exact value of θ^x is not relevant, so one can induce without loss of generality $\theta^x = 1$. Suppose now to the contrary that in optimum, $\theta^x < 1$ and $I^x > 0$ hold. If $\theta^x = 0$, positive investment, $I^x > 0$, is clearly inefficient; thus, it suffices to consider $\theta^x \in (0, 1)$. Then, every user shares $\chi^x := \theta^x I^x > 0$ units of data with platform x. For users to be willing to share $\theta^x \in (0, 1)$ data with platform x, it must be that users are indifferent between sharing and not sharing data with platform x, which implies — by means of (4) — that $q^x = c^x$. Notice that the stipulation of $q^x = c^x$ can also induce $\theta^x = 1$; intuitively, platform x could raise q^x marginally to break the indifference and to induce $\theta^x = 1$.

As users are indifferent between sharing and not sharing data (i.e., $q^x = c^x$), it follows that I^x and χ^x do not directly affect $\hat{z}_1$ and adoption N_1^x (see (6)). Moreover, the platform pays every user $q^x \chi^x$ dollars for their total data contribution, which depends on I^x and θ^x only via the product $\chi^x = I^x \theta^x$. The platform x can now set $\theta^x = 1$ and reduce investment I^x to $\hat{I}^x < I^x$, whilst keeping $\chi^x = \theta^x I^x = \hat{I}^x$ and $\chi^x q^x$ as well as $\hat{D}^x = N_1^x \chi^x$ and π_2^x (which depends on I^x and θ^x only via χ^x and $\hat{D}^x$) unchanged. This reduces the cost of investment and increases platform payoff by discrete amount $\frac{\lambda((I^x)^2 - (\hat{I}^x)^2)}{2} > 0$, contradicting the optimality of $\theta^x < 1$. As such, $\theta^x = 1$ is optimal.

Next, we show that (conditional on $\theta^x = 1$) the choice of q^x is payoff relevant, in that N_t^x, π_t^x, I^x as well as user welfare do not depend on $q^{x'}$ for $x, x' \in \{A, B\}$. For this sake, we fix the choice of $q^{x'}$, and conjecture (and later verify) that $\frac{\partial p_1^x}{\partial q^x} = I^x$ as well as $\frac{\partial p_1^x}{\partial q^{-x}} = 0$. Also, conjecture (and later verify) that investment I^x does not depend on $q^{x'}$, i.e., $\frac{dI^x}{dq^{x'}} = 0$, for $x, x' \in \{A, B\}$. Under these conjectures, the expression (6) with $\theta^x = 1$ implies that

$$N_1^A = \hat{z}_t = \frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B) + \left[I^A(q^A - c^A) - I^B(q^B - c^B)\right]}{2\kappa}$$

is independent of $q^{x'}$ in a sense that $\frac{dN_1^A}{dq^A} = \frac{\partial N_1^A}{\partial p_1^A}\frac{\partial p_1^A}{\partial q^A} + \frac{\partial N_1^A}{\partial q^A} = 0$ and $\frac{dN_1^A}{dq^B} = \frac{\partial N_1^A}{\partial p_1^B}\frac{\partial p_1^B}{\partial q^B} + \frac{\partial N_1^A}{\partial q^B} = 0$. Analogously, $\frac{dN_1^B}{dq^{x'}} = 0$ for $x' \in \{A, B\}$. As a result and due to $\frac{dI^x}{dq^{x'}} = 0$, $\hat{D}^A = N_1^A I^A$ and $\hat{D}^B = N_1^B I^B$ do not depend on $q^{x'}$ for $x' \in \{A, B\}$, i.e., $\frac{d\hat{D}^x}{dq^{x'}} = 0$. Therefore, D^x, which is a function of $N_1^A I^A$ and $N_1^B I^B$, does not depend on $q^{x'}$, i.e., $\frac{dD^x}{dq^{x'}} = 0$. This, in turn, implies that period-2 payoff $\pi_2^x = N_2^x p_2^x$ does not depend on $q^{x'}$, in that $\frac{d\pi_2^x}{dq^{x'}} = 0$ as well as $\frac{dp_2^x}{dq^{x'}} = \frac{dN_2^x}{dq^{x'}} = 0$.

Then, we can differentiate the payoff in period $t = 1$ to obtain

$$\begin{aligned}
\frac{d}{dq^x}\pi_1^x &= \frac{d}{dq^x}\left(N_1^x p_1^x - q^x N_1^x I^x + N_2^x p_2^x - \frac{\lambda(I^x)^2}{2}\right) = N_1^x I^x - N_1^x I^x = 0 \\
\frac{d}{dq^{-x}}\pi_1^x &= \frac{d}{dq^{-x}}\left(N_1^x p_1^x - q^x N_1^x I^x + N_2^x p_2^x - \frac{\lambda(I^x)^2}{2}\right) = 0,
\end{aligned}$$

where it was used (in the first line) that $N_2^x p_2^x - \frac{\lambda(I^x)^2}{2}$ does not depend on q^x, that N_1^x does not depend on q^x, and that $\frac{\partial p_1^x}{\partial q^x} = I^x$. Thus, platform x's payoff does not depend on $q^{x'}$. As a result, we obtain

$$\frac{d}{dq^{x'}}\frac{\partial \pi_1^x}{\partial I^x} = \frac{\partial}{\partial I^x}\frac{d}{dq^{x'}}\pi_1^x = 0.$$

Since optimal investment I^x solves the first-order condition in (F.40), that is, $\frac{\partial \pi_1^x}{\partial I^x} = 0$, it readily follows that I^x does not depend on $q^{x'}$, which verifies our conjecture that optimal investment I^x does not depend on $q^{x'}$.

Lastly, we solve the first-order condition (regarding period-1 price) (8) for the price

$$p_1^x = I^x q^x - \left(\frac{\partial N_1^x}{\partial p_1^x}\right)^{-1}\left[\sum_{x'=A,B}\left(\frac{\partial \pi_2^x}{\partial D^{x'}}\frac{\partial D^{x'}}{\partial N_1^{x'}}\frac{\partial N_1^{x'}}{\partial p_1^x}\right) + N_1^x\right].$$

Our previous arguments imply that

$$\left(\frac{\partial N_1^x}{\partial p_1^x}\right)^{-1}\left[\sum_{x'=A,B}\left(\frac{\partial \pi_2^x}{\partial D^{x'}}\frac{\partial D^{x'}}{\partial N_1^{x'}}\frac{\partial N_1^{x'}}{\partial p_1^x}\right) + N_1^x\right]$$

does not depend on $q^{x'}$. As a consequence, it readily follows that $\frac{dp_1^x}{dq^x} = \frac{\partial p_1^x}{\partial q^x} = I^x$ and $\frac{dp_1^x}{dq^{-x}} = \frac{\partial p_1^x}{\partial q^{-x}} = 0$, which verifies our initial conjecture.

Taken together, we have shown that $\frac{\partial p_1^x}{\partial q^x} = I^x$, $\frac{\partial p_1^x}{\partial q^{-x}} = 0$ as well as $\frac{dI^x}{dq^{x'}} = 0$ (i.e., investment does not depend on $q^{x'}$) for $x, x' \in \{A, B\}$. That is, period-1 prices can be written in the form

$$p_1^x = \bar{p}_1^x + I^x q^x,$$

where $\bar{p}_1^x$ does not depend on $q^{x'}$, i.e., $\frac{\partial \bar{p}_1{}^x}{\partial q^{x'}} = 0$. It follows (from (6)) that N_1^x does not depend on $q^{x'}$. Because I^x does not depend on q^x or q^{-x}, we have that $\hat{D}^x$ does not depend on q^x or q^{-x}. Thus, the levels of $q^{x'}$ do not affect any period-2 equilibrium quantities, such as N_2^x, π_2^x, p_2^x, or u_2. Finally, it remains to show that q^x does not affect period-1 user welfare u_1. However, this is immediate from the facts that $\frac{\partial p_1^x}{\partial q^x} = I^x$, $\frac{\partial p_1^x}{\partial q^{-x}} = 0$, and the fact that no other equilibrium quantities depend on q^x.

Another corollary is that it is without loss of generality to set $q^x = c^x$ which, in turn, incentivizes $\theta^x = 1$ regardless of the value of $\underline{\theta}$. It follows that the value of $\underline{\theta}$ does not affect investment I^x, platform payoff π^x, market shares N_t^x, or user welfare.

B Proof and Derivations for Results of Section 2.2

We present proofs and derivations for the results presented in Section 2.2. The proofs for results which assume symmetric platforms are deferred to Appendix D where we present the model solution and solve for the (symmetric) equilibrium in the symmetric platform case in closed form.

B.1 Proof of Proposition 2

Follows from the more general result in Proposition 4 upon setting $\eta = 0$. The proof of Proposition 4 is presented in Appendix D.

B.2 Proof of Proposition 3

Recall that the solution and equilibrium in period $t = 2$ is characterized in Lemma 1. Notice that because there is neither data sharing nor a market for data, we have $D^x = \hat{D}^x = N_1^x I^x$, where we used that in optimum $\theta^x = 1$ (see Proposition 1 which applies in this context). Also, according to Proposition 1, we consider without loss of generality that $q^x = c^x$ for the following arguments.

As a result, realize that the expression for $\hat{z}_t$ from (6) implies for $t = 1$:

$$\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}. \tag{B.6}$$

Noting that $N_1^A = \hat{z}_1$ and $N_1^B = 1 - \hat{z}_1$, we calculate

$$\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^x}{\partial p_1^{-x}} = \frac{1}{2\kappa}.$$

Next, we differentiate the platforms' period-2 payoff (under equilibrium pricing), characterized in (A.5) or Proposition 1, with respect to $D^{x'}$ for $x, x' = A, B$ to obtain

$$\frac{\partial \pi_2^A}{\partial D^A} = \frac{\phi^A\left(3\,\kappa + D^A\phi^A - D^B\phi^B + \Delta_K\right)}{9\kappa} \quad \text{and} \quad \frac{\partial \pi_2^B}{\partial D^B} = \frac{\phi^B\left(3\,\kappa - D^A\phi^A + D^B\phi^B - \Delta_K\right)}{9\kappa}, \tag{B.7}$$

as well as

$$\frac{\partial \pi_2^A}{\partial D^B} = -\frac{\phi^B\left(3\,\kappa + D^A\phi^A - D^B\phi^B + \Delta_K\right)}{9\kappa} \quad \text{and} \quad \frac{\partial \pi_2^B}{\partial D^A} = -\frac{\phi^A\left(3\,\kappa - D^A\phi^A + D^B\phi^B - \Delta_K\right)}{9\kappa}. \tag{B.8}$$

Next, note that because of $D^x = I^x N_1^x$, it follows that

$$\frac{\partial \pi_2^x}{\partial p_1^x} = -\frac{\partial \pi_2^x}{\partial D^x}\frac{I^x}{2\kappa} + \frac{\partial \pi_2^x}{\partial D^{-x}}\frac{I^{-x}}{2\kappa}. \tag{B.9}$$

As a result, the first-order conditions (8) for period-1 prices p_1^x become

$$\begin{aligned} \frac{\partial \pi_1^A}{\partial p_1^A} &= \left(\frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^A - I^A c^A}{2\kappa}\right) - \frac{1}{2\kappa}\left(\frac{\partial \pi_2^A}{\partial D^A} I^A - \frac{\partial \pi_2^A}{\partial D^B} I^B\right) = 0 \\ \frac{\partial \pi_1^B}{\partial p_1^B} &= \left(\frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^B - I^B c^B}{2\kappa}\right) - \frac{1}{2\kappa}\left(\frac{\partial \pi_2^B}{\partial D^B} I^B - \frac{\partial \pi_2^B}{\partial D^A} I^B\right) = 0. \end{aligned} \tag{B.10}$$

Next, we can insert (B.7) and (B.8) as well as $D^x = N_1^x I^x$ into (B.10) (with N_1^x from (B.6)), and subsequently solve the two first-order conditions in (B.10) — which are two linear equations — for prices p_1^x. These price expressions then imply the expression for market share $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K - p_1^A - p_1^B}{2\kappa}$ in $t = 1$. Using the expression for period-2 market share (under equilibrium pricing) from Lemma 1 and $D^A\phi^A = I^A\phi^A\hat{z}_1$ as well as $D^B\phi^B = I^B\phi^B(1 - \hat{z}_1)$, we obtain for $\hat{\phi}^x := \phi^x I^x$:

$$\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + 2\hat{\phi}^A\hat{z}_1 - \hat{\phi}^B}{2\kappa}.$$

Inserting $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K - p_1^A - p_1^B}{2\kappa}$ under period-1 equilibrium prices p_1^x, one obtains, after some algebra omitted here, the following expression for $\hat{z}_2$:

$$\hat{z}_2 = \frac{1}{2} + \frac{3\Delta_K(6\kappa + \hat{\phi}^A + \hat{\phi}^B) + 9\kappa(\hat{\phi}^A - \hat{\phi}^B) - 3c(I^A - I^B)(\hat{\phi}^A + \hat{\phi}^B)}{4\left(27\,\kappa^2 - (\hat{\phi}^A + \hat{\phi}^B)^2\right)}, \tag{B.11}$$

where $\hat{\phi}^x := I^x\phi^x$. Note that $\hat{z}_2$ (partially) increases with $\hat{\phi}^A - \hat{\phi}^B$; it also (partially) increases with $\hat{\phi}^A + \hat{\phi}^B$, when c is sufficiently low.

Next, we show that when c and Δ_K are sufficiently small, and $\phi^A > \phi^B$, then it holds that $I^A > I^B$, $p_1^A < p_1^B$ (when $q^x \leq c^x$), and $p_2^A > p_2^B$, also implying $\hat{\phi}^A > \hat{\phi}^B$. Consider $c = c^x = \Delta_K = 0$ and $q^x = c^x = 0$; the result then follows by continuity once we have establisehd it for $c = \Delta_K = 0$. We conjecture and verify that $N_1^A > N_1^B$ and $D^A\phi^A > D^B\phi^B \iff N_1^A I^A\phi^A > N_1^B I^B\phi^B$. Let

us first prove that $p_1^A < p_1^B$. For this sake, we first show that

$$\sum_{x'=A,B} \left(\frac{\partial \pi_2^A}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x} \right) < \sum_{x'=A,B} \left(\frac{\partial \pi_2^B}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x} \right). \tag{B.12}$$

Because $D^x = N_1^x I^x$ and $\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa}$ as well as $\frac{\partial N_1^x}{\partial p_1^{-x}} = \frac{1}{2\kappa}$, we calculate

$$\sum_{x'=A,B} \left(\frac{\partial \pi_2^x}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x} \right) = -\frac{1}{2\kappa} \left(\frac{\partial \pi_2^x}{\partial D^x} I^x - \frac{\partial \pi_2^x}{\partial D^{-x}} I^{-x} \right). \tag{B.13}$$

To prove inequality (B.12), we therefore need to show that

$$\frac{\partial \pi_2^A}{\partial D^A} I^A - \frac{\partial \pi_2^A}{\partial D^B} I^B > \frac{\partial \pi_2^B}{\partial D^B} I^B - \frac{\partial \pi_2^B}{\partial D^A} I^A.$$

With $\hat{\phi}^x = \phi^x I^x$, we obtain

$$\begin{aligned}
&\frac{\partial \pi_2^A}{\partial D^A} I^A - \frac{\partial \pi_2^A}{\partial D^B} I^B - \left[\frac{\partial \pi_2^B}{\partial D^B} I^B - \frac{\partial \pi_2^B}{\partial D^A} I^A \right] \\
&= \frac{\hat{\phi}^A \left(3\,\kappa + D^A\phi^A - D^B\phi^B\right)}{9\kappa} + \frac{\hat{\phi}^B \left(3\,\kappa + D^A\phi^A - D^B\phi^B\right)}{9\kappa} \\
&\quad - \left[\frac{\hat{\phi}^B \left(3\,\kappa - D^A\phi^A + D^B\phi^B\right)}{9\kappa} + \frac{\hat{\phi}^A \left(3\,\kappa - D^A\phi^A + D^B\phi^B\right)}{9\kappa} \right] \\
&= \frac{2(\hat{\phi}^A + \hat{\phi}^B)(D^A\phi^A - D^B\phi^B)}{9\kappa}.
\end{aligned}$$

As such, given the conjecture $D^A\phi^A > D^B\phi^B$, the inequality (B.12) holds. As prices p_1^x solve the first-order condition (8), inequality (B.12) implies $p_1^A < p_1^B$. It then follows immediately that $N_1^A > N_1^B$ (as $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K - p_1^A - p_1^B}{2\kappa}$).

Next, we show that $I^A > I^B$. Notice that, due to $q^x = c^x$, we have $\frac{\partial N_1^x}{\partial I^x} = 0$ and $\frac{N_1^x}{\partial I^{-x}} = 0$. It becomes apparent from the first-order condition for investment (9) (with $q^x = c^x = 0$) that $I^A > I^B$ when

$$\sum_{x'=A,B} \left(\frac{\partial \pi_2^A}{\partial D^{x'}} \left[\frac{\partial D^{x'}}{\partial I^x} + \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial I^x} \right] \right) > \sum_{x'=A,B} \left(\frac{\partial \pi_2^B}{\partial D^{x'}} \left[\frac{\partial D^{x'}}{\partial I^x} + \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial I^x} \right] \right). \tag{B.14}$$

Because $D^x = N_1^x I^x$, and N_1^x does not (directly) depend on investment (i.e., $\frac{\partial N_1^{x'}}{\partial I^x} = 0$), we have $\frac{\partial D^x}{\partial I^x} = N_1^x$ as well as $\frac{\partial D_2^x}{\partial I^{-x}} = 0$. As such, inequality (B.14) can be rewritten as

$$\frac{\partial \pi_2^A}{\partial D^A} N_1^A > \frac{\partial \pi_2^B}{\partial D^B} N_1^B.$$

Using (B.7) and (B.8), we can calculate

$$\begin{aligned}
&\frac{\partial \pi_2^A}{\partial D^A} N_1^A - \frac{\partial \pi_2^B}{\partial D^B} N_1^B \\
&= \frac{\phi^A \left(3\,\kappa + D^A\phi^A - D^B\phi^B\right)}{9\kappa} N_1^A - \frac{\phi^B \left(3\,\kappa - D^A\phi^A + D^B\phi^B\right)}{9\kappa} N_1^B > 0
\end{aligned}$$

where the last inequality used that $N_1^A > N_1^B$ and $D^A\phi^A > D^B\phi^B$. As a result, inequality (B.14) holds, so that $I^A > I^B$. Because, in addition, $p_1^A < p_1^B$, it follows that $N_1^A > N_1^B$ and $D^A = N_1^A I^B > N_1^B I^B = D^B$. Thus, we have verified the conjecture that $N_1^A > N_1^B$ and $D^A\phi^A > D^B\phi^B \iff N_1^A I^A \phi^A > N_1^B I^B \phi^B$. According to Lemma 1, the fact that $\phi^A D^A > \phi^B D^B$ implies $N_2^A > N_2^B$ (i.e., $\hat{z}_2 > 1/2$) as well as $p_2^A > p_2^B$, which was to show.

B.3 Proof of Lemma 2

Note that $p_2^A = 2\kappa\hat{z}_2$ as well as $p_2^B = 2\kappa(1-\hat{z}_2)$. Thus, we can write user welfare in $t=2$ as

$$u_2 = \hat{z}_2(Y_2^A - 2\kappa\hat{z}_2) + (1-\hat{z}_2)(Y_2^B - 2\kappa(1-\hat{z}_2)) - \bar{\kappa}_2.$$

Next, note that

$$\begin{aligned}
\hat{z}_2 Y_2^A + (1-\hat{z}_2)Y_2^B &= K^B + \phi^B D^B + \hat{z}_2(K^A - K^B + \phi^A D^A - \phi^B D^B + \gamma^A \hat{z}_2) + \gamma^B(1-\hat{z}_2)^2 \\
&= K^B + \phi^B D^B + \hat{z}_2\left(\Delta_K - \frac{\gamma^A - \gamma^B}{2} + \phi^A D^A - \phi^B D^B + \gamma^A\hat{z}_2\right) + \gamma^B(1-\hat{z}_2)^2 \\
&= K^B + \phi^B D^B + \hat{z}_2\left(6\kappa\hat{z}_2 - 3\kappa - \frac{\gamma^A - \gamma^B}{2} + \gamma^A\hat{z}_2\right) + \gamma^B(1-\hat{z}_2)^2,
\end{aligned}$$

where we used

$$\kappa := \hat{\kappa} - \frac{\gamma^A + \gamma^B}{2} \quad \text{and} \quad \Delta_K := K^A - K^B + \frac{\gamma^A - \gamma^B}{2}.$$

Thus,

$$\begin{aligned}
u_2 = K^B &+ \phi^B D^B + \hat{z}_2\left(6\kappa\hat{z}_2 - 3\kappa - \frac{\gamma^A - \gamma^B}{2} + \gamma^A\hat{z}_2\right) + \gamma^B(1-\hat{z}_2)^2 \\
&- 2\kappa\hat{z}_2^2 - 2\kappa(1-\hat{z}_2)^2 - \frac{\hat{\kappa}(\hat{z}_2)^2 + \hat{\kappa}(1-\hat{z}_2)^2}{2} =: \tilde{u}_2 - 2\kappa\hat{z}_2^2 - 2\kappa(1-\hat{z}_2)^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial u_2}{\partial \hat{z}_2} &= 12\kappa\hat{z}_2 - 3\kappa - \frac{\gamma^A - \gamma^B}{2} + 2\gamma^A\hat{z}_2 - 2\gamma^B(1-\hat{z}_2) - \hat{\kappa}\hat{z}_2 + \hat{\kappa}(1-\hat{z}_2) - 4\kappa\hat{z}_2 + 4\kappa(1-\hat{z}_2) \\
&= 4\kappa\hat{z}_2 + \kappa - \frac{\gamma^A - \gamma^B}{2} + 2(\gamma^A + \gamma^B)\hat{z}_2 - 2\gamma^B - \hat{\kappa}(2\hat{z}_2 - 1) \\
&= 4\kappa\hat{z}_2 + 2\kappa + 2(\gamma^A + \gamma^B)\hat{z}_2 - \gamma^B - 2\hat{\kappa}\hat{z}_2 \\
&= 4\kappa\hat{z}_2 + 2\kappa + 2(\gamma^A + \gamma^B)\hat{z}_2 - \gamma^B - 2\left(\kappa + \frac{\gamma^A + \gamma^B}{2}\right)\hat{z}_2 \\
&= 2\kappa\hat{z}_2 + 2\kappa + (\gamma^A + \gamma^B)\hat{z}_2 - \gamma^B.
\end{aligned}$$

As such,

$$\frac{\partial u_2}{\partial \hat{z}_2} < 0 \iff \gamma^B > \frac{2\kappa(1+\hat{z}_2) + \gamma^A\hat{z}_2}{1-\hat{z}_2}.$$

That is, when $\hat{z}_2 \in (0,1)$ and γ^B is sufficiently large relative to γ^A (i.e., when $\gamma^B - \gamma^A$ is sufficiently large), then $\frac{\partial u_2}{\partial \hat{z}_2} < 0$ and user welfare in $t=2$, i.e., u_2, decreases with platform A's market share $N_2^A = \hat{z}_2$, which concludes the argument.

C Proofs and Derivations for Results of Section 3

We present proofs and derivations for the results presented in Section 3. The proofs for results which assume symmetric platforms are deferred to Appendix D where we present the model solution and solve for the symmetric equilibrium in the symmetric platform case.

C.1 Proof of Proposition 4

The proof of Proposition 4 is presented in Appendix D where we present the solution for the symmetric platform case.

C.2 Proof of Proposition 5

We already use that $\theta^x = 1$ and set without loss of generality $q^x = c^x$ (see Proposition 1 which applies in this context). As $\eta = 1$, we have $D = D_2^x = N_1^A I^A + N_1^B I^B$, so $\frac{\partial D}{\partial I^x} \geq 0$. Using Lemma 1, we obtain for period-2 platform payoffs (under equilibrium pricing):

$$\pi_2^A = \frac{\left(3\,\kappa + \Delta_K + D(\phi^A - \phi^B)\right)^2}{18\,\kappa} \quad \text{and} \quad \pi_2^B = \frac{\left(3\,\kappa - \Delta_K - D(\phi^A - \phi^B)\right)^2}{18\,\kappa}. \tag{C.15}$$

One can calculate (with $\phi^A \geq \phi^B$)

$$\frac{\partial \pi_2^A}{\partial I^A} = (\phi^A - \phi^B)\left(\frac{3\kappa + \Delta K + D(\phi^A - \phi^B)}{9\kappa}\right)\frac{\partial D}{\partial I^A} \geq 0$$
$$\frac{\partial \pi_2^B}{\partial I^B} = -(\phi^A - \phi^B)\left(\frac{3\kappa - \Delta K - D(\phi^A - \phi^B)}{9\kappa}\right)\frac{\partial D}{\partial I^B} \leq 0,$$

where it was used that — by assumption/parameter condition (7) —- $3\kappa > \Delta_K + \phi^A - \phi^B$ and $D \leq 1$. When $\phi^A > \phi^B$ and $N_1^A > 0$ ($N_1^B > 0$), then $\frac{\partial \pi_2^A}{\partial I^A} > 0 > \frac{\partial \pi_2^B}{\partial I^B}$. As such, the first-order condition with respect to investment (9) readily implies that there exists no interior optimal $I^B \in (0,1)$. In fact, $\frac{\partial \pi_1^B}{\partial I^B} \leq 0$, and $I^B = 0$ is optimal for any $c > 0$ and $\lambda > 0$ (as well as in the limit $c \to 0$ and $\lambda \to 0$).

To derive the expression for platform A's period-2 market share $N_2^A = \hat{z}_2$, first notice that $\frac{\partial N_1^x}{\partial p_1^x} = \frac{-1}{2\kappa}$ and $\frac{\partial N_1^{-x}}{\partial p_1^x} = \frac{1}{2\kappa}$. Then, calculate $\frac{\partial D^x}{\partial p_1^x} = -\frac{I^x - I^{-x}}{2\kappa} = \frac{I^{-x} - I^x}{2\kappa}$ and $\frac{\partial D^{-x}}{\partial p_1^x} = \frac{I^{-x} - I^x}{2\kappa}$, and, due to $D^x = D$, we have $\frac{\partial D}{\partial p_1^x} = \frac{I^{-x} - I^x}{2\kappa}$. Thus,

$$\frac{\partial \pi_2^x}{\partial p_1^x} = \frac{\partial \pi_2^x}{\partial D^x}\left(\frac{I^{-x} - I^x}{2\kappa}\right) = \frac{\partial \pi_2^x}{\partial D^{-x}}\left(\frac{I^{-x} - I^x}{2\kappa}\right) = \frac{\partial \pi_2^x}{\partial D}\left(\frac{I^{-x} - I^x}{2\kappa}\right),$$

where we used $D = D^x = N_1^A I^A + N_1^B I^B = N_1^A I^A$ (due to $I^B = 0$) and (C.15).

Then, the first order conditions with respect to price in period $t = 1$, i.e., $\frac{\partial \pi_1^x}{\partial p_1^x} = 0$ for $x = A, B$, become

$$\frac{\partial \pi_1^A}{\partial p_1^A} = \left(\frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^A - I^A c^A}{2\kappa}\right) - \left(\frac{I^A}{2\kappa}\right)\frac{\partial \pi_2^A}{\partial D} = 0$$
$$\frac{\partial \pi_1^B}{\partial p_1^B} = \left(\frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^B}{2\kappa}\right) + \left(\frac{I^A}{2\kappa}\right)\frac{\partial \pi_2^B}{\partial D} = 0, \tag{C.16}$$

where we used that $I^B = 0$. Using and differentiating the expression for period-2 payoff in (C.15),

we can calculate:

$$\frac{\partial \pi_1^A}{\partial p_1^A} = \left(\frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^A - I^A c^A}{2\kappa}\right) - \left(\frac{I^A(\phi^A - \phi^B)}{2\kappa}\right)\left(\frac{3\,\kappa + \Delta_K + D(\phi^A - \phi^B)}{9\,\kappa}\right)$$
$$\frac{\partial \pi_1^B}{\partial p_1^B} = \left(\frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^B}{2\kappa}\right) - \left(\frac{I^A(\phi^A - \phi^B)}{2\kappa}\right)\left(\frac{3\,\kappa - \Delta_K - D(\phi^A - \phi^B)}{9\,\kappa}\right).$$

Also, recall that $c^A = c^B = (1+\eta)c = 2c$. After some algebra omitted, one obtains equilibrium prices p_1^A and p_1^B by solving (C.16) — which is a system of two linear equations — for p_1^A and p_1^B.

Next, notice that $D = D^x = N_1^A I^A$ and $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K + p_1^A - p_1^B}{2\kappa}$. It follows from Lemma 1 that (under equilibrium pricing) $\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + N_1^A I^A(\phi^A - \phi^B)}{6\kappa}$. Using period-1 equilibrium prices p_1^x, one can then calculate (after some algebra omitted here) the period-2 market share of A:

$$\hat{z}_2 = \frac{1}{2} + \frac{3\Delta_K(6\kappa + I^A\phi^A - I^A\phi^B) + 9\kappa I^A(\phi^A - \phi^B) - 6c(I^A)^2(\phi^A - \phi^B)}{4\,(27\,\kappa^2 - (I^A\phi^A - I^A\phi^B)^2)}, \tag{C.17}$$

which simplifies to $\hat{z}_2 = 1/2$ when $\phi^A = \phi^B$ and $\Delta_K = 0$.

Next, consider that $\phi^A > \phi^B$, and that $\lambda > 0$ and $c > 0$ are sufficiently small. Then, it is clear that, under the baseline without data sharing, we have that $I^A = I^B = 1$. We can use the expression for period-2 presented in the proof of Proposition 3 — that is, (B.11) — to calculate the period-2 market share of platform A (which we denote by $\hat{z}_2'$) by inserting $I^x = 1$:

$$\hat{z}_2' := \frac{1}{2} + \frac{3\Delta_K(6\kappa + \phi^A + \phi^B) + 9\kappa(\phi^A - \phi^B)}{4\,(27\,\kappa^2 - (\phi^A + \phi^B)^2)}.$$

Under data-sharing ($\eta = 1$), we have $I^A = 1 > 0 = I^B$, when c and λ are sufficiently small. We insert $I^A = 1 > I^B = 0$ into (C.17) to get:

$$\hat{z}_2 = \frac{1}{2} + \frac{3\Delta_K(6\kappa + \phi^A - \phi^B) + 9\kappa(\phi^A - \phi^B) - 6c(\phi^A - \phi^B)}{4\,(27\,\kappa^2 - (\phi^A - \phi^B)^2)}$$

Note that (7) implies $27\,\kappa^2 > (\phi^A + \phi^B)^2$. It is then evident that $\hat{z}_2 < \hat{z}_2'$ when $\phi^A > \phi^B$. As such, by continuity, when λ and c are sufficiently small, then data sharing ($\eta = 1$) reduces platform A's period-2 market share relative to the baseline with $\eta = 0$, which was to show.

C.3 Proof of Proposition 6

The proof of Proposition 6 is presented in Appendix D.2 where we present the solution for the symmetric platform case.

C.4 Proof of Proposition 7

Recall that the timing within period-2 is as follows: First, platforms decide how much data to buy from users and, second, they set prices p_2^x, leading to data-dependent continuation payoff π_2^x characterized in Lemma 1. It is clear that platforms pay (per unit) price c for buying data from individual users, because c is the minimal price at which users are willing to sell data to platforms; offering a higher price would only hurt platform payoffs.

To begin with, take the total stock of data D as given and consider platform x's decision how much data to buy (at per unit price c) at the beginning of $t = 2$ before prices p_2^x. Given $D^{x'}$ for $x' = A, B$, platforms' period-2 (continuation) payoff equals π_2^x from Lemma 1. Platform x maximizes

$$\max_{D^x \in [0,D]} \Big(\pi_2^x - cD^x\Big),$$

taking the choice of the other platform D^{-x} as given. Because the period-2 platform payoff under equilibrium pricing from Lemma 1, is strictly convex in D^x, i.e., $\frac{\partial^2 \pi_2^x}{\partial (D^x)^2} > 0$, it follows that (in equilibrium/optimum) $D^x \in \{0, D\}$, i.e., platforms either buy no data or the full amount of data and there is no interior optimum.

Now, let us analyze jointly platforms' (equilibrium) choice of investment I^x and their data acquisition in $t = 2$, D^x. If $D^A = D^B = 0$ in period 2 in equilibrium, then clearly $I^A = I^B = 0$. Next, if $D^B = 0$ in equilibrium, then $I^B = 0$ is clearly optimal.

Now, consider $D^B = D$ as well as $D^A = D$ (otherwise, we could relabel platforms; i.e., relabel A to B and then we are back to the previous case of $I^B = 0$). Notice $D = N_1^A I^A + N_1^B I^B$, so $\frac{\partial D}{\partial I^x} \geq 0$. Using the period-2 platform payoff under equilibrium pricing from Lemma 1 for $D^A = D^B = D$, we obtain

$$\pi_2^A = \frac{\left(3\,\kappa + \Delta_K + D(\phi^A - \phi^B)\right)^2}{18\,\kappa} \quad \text{and} \quad \pi_2^B = \frac{\left(3\,\kappa - \Delta_K - D(\phi^A - \phi^B)\right)^2}{18\,\kappa}. \tag{C.18}$$

One can calculate (with $\phi^A \geq \phi^B$)

$$\begin{aligned} \frac{\partial \pi_2^A}{\partial I^A} &= (\phi^A - \phi^B)\left(\frac{3\kappa + \Delta\kappa + D(\phi^A - \phi^B)}{9\kappa}\right)\frac{\partial D}{\partial I^A} \geq 0 \\ \frac{\partial \pi_2^B}{\partial I^B} &= -(\phi^A - \phi^B)\left(\frac{3\kappa - \Delta\kappa - D(\phi^A - \phi^B)}{9\kappa}\right)\frac{\partial D}{\partial I^B} \leq 0, \end{aligned} \tag{C.19}$$

where it was used that — by assumption (7) —- $3\kappa > \Delta_K + \phi^A - \phi^B$ and $D \leq 1$. Then, the fact that optimal investment — if interior — must solve the first-order condition

$$\frac{\partial \pi_1^x}{\partial I^x} = -\lambda I^x + \frac{\partial \pi_2^x}{\partial I^x} = 0$$

with π_1^x from (10) and $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}$ readily implies that there exists no interior optimal $I^B \in (0, 1)$.[50] In fact, $\frac{\partial \pi_1^B}{\partial I^B} \leq 0$, and $I^B = 0$ is optimal for any $c > 0$ and $\lambda > 0$ (as well as in the limit $c \to 0$ and $\lambda \to 0$). In other words, only one platform — say platform A — undertakes investment, in that $I^A \geq 0 = I^B$.

Take $\phi^A > \phi^B$. Consider that c and λ are sufficiently low (e.g., the limit case $c \to 0$ and $\lambda \to 0$), so that $D^x = D$ is optimal and $I^A = 1$ as well as $I^B = 0$, due to $\frac{\partial \pi_1^x}{\partial I^x} = -\lambda I^x + \frac{\partial \pi_2^x}{\partial I^x}$ and (C.19). Then, $D = D^A = D^B = N_1^A I^A + N_1^B I^B = N_1^A I^A$ (as $I^B = 0$), and the model solution becomes similar to data sharing with $\eta = 1$ from Proposition 5, with different privacy cost $c^x = c$ (instead of $c^x = 2c$). The reason is that under data sharing with $\eta = 1$, platform A incurs effectively privacy cost $2c$, as it must also compensate users for their data being used on B in addition to being used on A. With the market for data, platform A compensates users only for their data being used on A, while B compensates them only for their data being used on B.

Consider that $\phi^A > \phi^B$, and that $\lambda > 0$ and $c > 0$ are sufficiently small. In the baseline, we have $I^A = I^B = 1$. We can use the expression for period-2 presented in the proof of Proposition 3 — that is, (B.11) — to calculate then the period-2 market share of platform A when $I^A = I^B = 1$ (which we denote by $\hat{z}_2'$) by inserting $I^x = 1$:

$$\hat{z}_2' = \frac{1}{2} + \frac{3\Delta_K(6\kappa + \phi^A + \phi^B) + 9\kappa(\phi^A - \phi^B)}{4\,(27\,\kappa^2 - (\phi^A + \phi^B)^2)}.$$

With the market for data, we have $I^A = 1 > 0 = I^B$ as well as $D^x = D = N_1^A I^A$. Next, notice that platform x anticipates, when choosing I^x and p_1^x, that it buys any unit data that is generated

[50] Recall, $N_1^x = \hat{z}_1$ if $x = A$ and $N_1^x = 1 - \hat{z}_1$ if $x = B$.

at per unit price c at time $t=2$. That is, with $D^x = D = N_1^A$, first-period payoff from (10) can be written as

$$\pi_1^A = N_1^A p_1^A + \frac{\left(3\,\kappa + \Delta_K + N_1^A(\phi^A - \phi^B)\right)^2}{18\,\kappa} - cN_1^A$$
$$\pi_1^B = N_1^B p_1^B + \frac{\left(3\,\kappa - \Delta_K - N_1^A(\phi^A - \phi^B)\right)^2}{18\,\kappa} - cN_1^A,$$

with $N_1^A = \frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}$ and $N_1^B = \frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}$, where we used (C.18) for expressions for π_2^x.

Now, notice that $\frac{\partial D}{\partial p_1^x} = \frac{I^{-x} - I^x}{2\kappa}$, as $\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa}$ and $\frac{\partial N_1^x}{\partial p_1^{-x}} = \frac{1}{2\kappa}$ the first order conditions with respect to price in period $t=1$ become (with $D = N_1^A$ and $I^A = 1 > I^B = 0$)

$$\frac{\partial \pi_1^A}{\partial p_1^A} = \left(\frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^A - c}{2\kappa}\right) - \frac{1}{2\kappa}\left(\frac{\partial \pi_2^A}{\partial D}\right) = 0 \tag{C.20}$$
$$\frac{\partial \pi_1^B}{\partial p_1^B} = \left(\frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^B + c}{2\kappa}\right) + \frac{1}{2\kappa}\left(\frac{\partial \pi_2^B}{\partial D}\right) = 0,$$

which — using the expressions for π_2^x with $D^x = D$ — simplifies to

$$\frac{\partial \pi_1^A}{\partial p_1^A} = \left(\frac{1}{2} + \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^A - c}{2\kappa}\right) - \frac{\phi^A - \phi^B}{2\kappa}\left(\frac{3\,\kappa + \Delta_K + D(\phi^A - \phi^B)}{9\,\kappa}\right) = 0$$
$$\frac{\partial \pi_1^B}{\partial p_1^B} = \left(\frac{1}{2} - \frac{\Delta_K - (p_1^A - p_1^B)}{2\kappa}\right) - \left(\frac{p_1^B + c}{2\kappa}\right) - \frac{\phi^A - \phi^B}{2\kappa}\left(\frac{3\,\kappa - \Delta_K - D(\phi^A - \phi^B)}{9\,\kappa}\right) = 0.$$

To obtain equilibrium prices p_1^A and p_1^B, one then solves (C.20) — which is a system of two linear equations — for p_1^A and p_1^B. Solving for period-1 equilibrium prices (after some algebra omitted here), using $I^A = 1$, $\hat{z}_1 = \frac{1}{2} + \frac{\Delta_K + p_1^B - p_1^A}{2\kappa}$, and $D^x = D = N_1^A I^A = I^A \hat{z}_1$ as well as the expression for $\hat{z}_2$ from Lemma 1, one can calculate (after some algebra omitted here) the market share of A in $t=2$ (in the limit when $c \to 0$):[51]

$$\hat{z}_2 = \frac{1}{2} + \frac{3\Delta_K(6\kappa + \phi^A + \phi^B) + 9\kappa(\phi^A - \phi^B)}{4\,(27\,\kappa^2 - (\phi^A - \phi^B)^2)}.$$

It is evident that $\hat{z}_2 < \hat{z}_2'$, i.e., the market for data (in which users sell their data) reduces market concentration relative to the baseline.

C.5 Proofs of Lemma 3 and Details for Section 3.4

We now prove the results of Lemma 3, and provide details for the solution of the model variant in which there is a market for data and the platforms own data.

C.5.1 Proof of Lemma 3

We solve for optimal data sharing at the onset of period $t=2$ to maximize total continuation surplus, i.e., $\pi_2^A + \pi_2^B$, whereby π_2^x is characterized in (A.5). As a preparation, we differentiate

[51] The general expression would be $\hat{z}_2 = \frac{1}{2} + \frac{3\Delta_K(6\kappa+\phi^A+\phi^B)+9\kappa(\phi^A-\phi^B)}{4\left(27\,\kappa^2-(\phi^A-\phi^B)^2\right)} + o(c)$, where the remainder term $o(c)$ tends to zero as $c \to 0$.

(A.5) with respect to $D^{x'}$ for $x, x' = A, B$ to obtain

$$\frac{\partial \pi_2^A}{\partial D^A} = \frac{\phi^A \left(3\,\kappa + D^A\phi^A - D^B\phi^B + \Delta_K\right)}{9\kappa} \quad \text{and} \quad \frac{\partial \pi_2^B}{\partial D^B} = \frac{\phi^B \left(3\,\kappa - D^A\phi^A + D^B\phi^B - \Delta_K\right)}{9\kappa},$$

as well as

$$\frac{\partial \pi_2^A}{\partial D^B} = -\frac{\phi^B \left(3\,\kappa + D^A\phi^A - D^B\phi^B + \Delta_K\right)}{9\kappa} \quad \text{and} \quad \frac{\partial \pi_2^B}{\partial D^A} = -\frac{\phi^A \left(3\,\kappa - D^A\phi^A + D^B\phi^B - \Delta_K\right)}{9\kappa}.$$

Above expressions are identical to (B.7) and (B.8).

Next, notice that $D^x \geq \hat{D}^x$, whereby $\hat{D}^x$ is platform x's stock of data at the beginning of period $t = 2$ *before* data sharing. Using (B.7) and (B.8), we obtain

$$\begin{aligned}\left(\frac{\partial(\pi_2^A + \pi_2^B)}{\partial D^A}\right) &= \phi^A \left(\frac{\left(3\,\kappa + D^A\phi^A - D^B\phi^B + \Delta_K\right) - \left(3\,\kappa - D^A\phi^A + D^B\phi^B - \Delta_K\right)}{9\kappa}\right) \\ &= \frac{2\phi^A(D^A\phi^A - D^B\phi^B + \Delta_K)}{9\kappa}.\end{aligned}$$

Consider $\hat{D}^A\phi^A + \Delta_K \geq \hat{D}^B\phi^B$. Thus, for any $D^A \geq \hat{D}^A$ and $D^B = \hat{D}^B$, we have $\left(\frac{\partial(\pi_2^A+\pi_2^B)}{\partial D^A}\right) \geq 0$, where the inequality is strict if $D^A > \hat{D}^A$.

Symmetrically, we can combine (B.7) and (B.8) to calculate

$$\left(\frac{\partial(\pi_2^A + \pi_2^B)}{\partial D^B}\right) = \frac{2\phi^B(D^B\phi^B - D^A\phi^A - \Delta_K)}{9\kappa}.$$

Thus, for $D^A \geq \hat{D}^A$ and $D^B = \hat{D}^B$, we have $\left(\frac{\partial(\pi_2^A+\pi_2^B)}{\partial D^B}\right) \leq 0$, where the inequality is strict if $D^A > \hat{D}^A$.

It follows that, given $\hat{D}^A\phi^A + \Delta_K \geq \hat{D}^B\phi^B$ and for $\phi^A \geq \phi^B$, total surplus is maximized upon implementing $D^A = \hat{D}^A + \hat{D}^B$ and $D^B = \hat{D}^B$, which was to show.

C.5.2 Solution of Model Variant with Market for Data and Details for Section 3.4

At the beginning of time period $t = 2$, before choosing prices p_2^x, platforms A and B trade data with each other, with endogenous price p_2^D. We assume throughout that (in equilibrium) $\hat{D}^A\phi^A + \Delta_K \geq \hat{D}^B\phi^B$ as well as $\phi^A \geq \phi^B$. Without loss of generality, it suffices to consider that platforms A and B choose the optimal allocation of data through Nash Bargaining at the beginning of period $t = 2$, with equal bargaining weights. The price for data p_2^D is then chosen to implement the split of resulting surplus. According to Lemma 3, total surplus is maximized when one platform x (labeled A) shares data with the other platform $-x$ (labeled B), but not the other way around.

We now derive the Nash bargaining solution. Now, B is the platform that sells data, and A is the platform that buys data. Then, data trade at the beginning of period $t = 2$ implies $D^A = \hat{D}^A + \hat{D}^B$ while $D^B = \hat{D}^B$, whereby $\hat{D}^x = N_1^x I^x \theta^x$ and $\theta^x = 1$ in optimum. As a result, using the expressions from Lemma 1 for period-2 platform payoffs π_2^x under equilibrium pricing, platforms derive the following payoff (just after Nash bargaining and trade and under period-2 equilibrium prices):

$$\pi_2^A = \frac{(3\kappa + \Delta_K + \phi^A\hat{D}^A)^2}{18\kappa} \quad \text{and} \quad \pi_2^B = \frac{(3\kappa - \Delta_K - \phi^A\hat{D}^A)^2}{18\kappa}.$$

On the other hand, absent data trade, we would have $D^x = \hat{D}^x$ and platform (equilibrium) payoffs

in period $t = 2$ would be

$$\hat{\pi}_2^A = \frac{(3\kappa + \Delta_K + \phi^A \hat{D}^A - \phi^B \hat{D}^B)^2}{18\kappa} \quad \text{and} \quad \hat{\pi}_2^B = \frac{(3\kappa - \Delta_K + \phi^B \hat{D}^B - \phi^A \hat{D}^A)^2}{18\kappa}.$$

As such, the total surplus created for the platforms from data trade equals

$$S := \pi_2^A + \pi_2^B - (\hat{\pi}_2^A + \hat{\pi}_2^B).$$

We denote by $\tilde{\pi}_2^x$ platform x's payoff at the beginning of period $t = 2$ just before Nash bargaining takes place. At this moment, platform x owns a stock of data $\hat{D}^x = N_1^x I^x$. Just after the data trade, we have $D^A = \hat{D}^A + \hat{D}^B$ and $D^B = \hat{D}^B$ and platform's continuation payoff becomes π_2^x.

It is well-known that, as per Nash bargaining protocol with equal bargaining weights 1/2, platform x's payoff (just before data trade and Nash bargaining) then reads

$$\tilde{\pi}_2^x = \hat{\pi}_2^x + \frac{1}{2} S = \frac{\hat{\pi}_2^x - \hat{\pi}_2^{-x} + \pi_2^A + \pi_2^B}{2}. \tag{C.21}$$

That is, platforms' payoff in $t = 2$ before data trade $\tilde{\pi}_2^x$ is the sum of the "reservation value" $\hat{\pi}_2^x$, which would obtain absent data trade, and half of the surplus generated from data trade S (because the bargaining weights of platforms are 1/2 each). In the context of Nash bargaining and the implementation of the optimal data allocation and payoffs, platform B receives a lump-sum transfer from A equal to $\tilde{\pi}_2^B - \pi_2^B$ as $\tilde{\pi}_2^B$ is the payoff just before data trade/Nash bargaining and π_2^B the payoff just after data trade/Nash bargaining. This transfer would imply a (per unit) price for $p_D = (\tilde{\pi}_2^B - \pi_2^B)/D^B$; this price for data does not play a role in what follows.

Formally, anticipating the continuation equilibrium in period $t = 2$ with Nash bargaining, platforms $x = A, B$ now maximize at time $t = 1$:

$$\max_{p_1^x, I^x, q^x} \left(N_1^x p_1^x - \frac{\lambda (I^x)^2}{2} - q^x \theta^x N_1^x I^x + \tilde{\pi}_2^x \right), \tag{C.22}$$

taking the choice of the other competing platform as given and subject to $N_1^A = \hat{z}_1$ and $N_1^B = 1 - \hat{z}_1$ with $\hat{z}_1$ characterized in (6). As in the baseline, setting $\theta^x = 1$ is optimal, and the choice of q^x is not payoff-relevant (so one can set $q^x = c^x$). Also note that $c^A = c < 2c = c^B$, because users anticipate that their data is used by both platforms when they contribute it to platform B.

Next, we present first order conditions with respect to period-1 prices and investment. Differentiating the objective in (C.22) with respect to p_1, we obtain (for $q^x = c^x$)

$$\frac{\partial \pi_1^x}{\partial p_1^x} = N_1^x + \frac{\partial N_1^x}{\partial p_1^x}(p_1^x - I^x c^x) + \sum_{x'=A,B} \left(\frac{\partial \tilde{\pi}_2^x}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x} \right) = 0.$$

The first order condition with respect to investment becomes (due to $\frac{\partial N_1^x}{\partial I^x} = 0$ when $q^x = c^x$):

$$\frac{\partial \pi_1^x}{\partial I^x} = \sum_{x'=A,B} \left(\frac{\partial \tilde{\pi}_2^x}{\partial D^{x'}} \left[\frac{\partial D^{x'}}{\partial I^x} + \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial I^x} \right] \right) - \lambda I^x - N_1^x c^x = 0,$$

which holds when investment I^x satisfies $I^x \in (0, 1)$. Provided its existence, we focus on a subgame perfect equilibrium in pure strategies in which $\hat{D}^A \phi^A + \Delta_K \geq \hat{D}^B \phi^B$.

By Lemma 1, platform A's market share in period $t = 2$ (under equilibrium pricing) becomes

$$\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + D\phi^A - D^B \phi^B}{6\kappa}.$$

When $I^A > 0$, then $\hat{z}_2 > 1/2$ under all parameters and, in particular, even when $\Delta_K = 0$ and $\phi^A = \phi^B$. As a result, when $\phi^A - \phi^B$ and Δ_K are sufficiently small and $I^x > 0$, then market concentration $\hat{z}_2 > 1/2$ is necessarily higher than under the baseline (where market concentration is 1/2 under platform symmetry).

D Solution for Symmetric platforms — Proofs of Propositions 2, 4, 6, 8, and 9

We now solve the model in the symmetric platform case with data sharing $\eta^x = \eta$, i.e., platform x must share fraction η of its data with its competitor $-x$ and vice versa. We solve for a symmetric equilibrium with $p_t^x = p_t^{-x}$, $I^x = I^{-x}$, $N_1^x = N_1^{-x} = 1/2$, $\theta^x = \theta^{-x}$, and $q^x = q^{-x}$. In this Section, we prove all Lemmata and Propositions which assume symmetric platforms, that is, Propositions 2, 4, 6, 8, and 9.

D.1 Proofs of Propositions 2 and 4

The arguments below therefore prove Proposition 4. The statements from Proposition 2 follow upon setting $\eta = 0$, in that Proposition 2 is a special case of Proposition 4. When platform x must share fraction $\eta \in [0, 1]$ of its data with the competitor platform $-x$, then we have $c^x = c(1 + \eta)$ where c is a constant. We solve for a symmetric equilibrium with $p_t^x = p_t^{-x}$, $I^x = I^{-x}$, $N_1^x = N_1^{-x} = 1/2$, $\theta^x = \theta^{-x}$, and $q^x = q^{-x}$.

D.1.1 Period $t = 2$

In the symmetric platform case, we have $\Delta_K = 0$, $\phi^A = \phi^B$ and $D^A = D^B$, so that — by Lemma 1 — $p_2^x = \kappa$ and $N_2^x = \frac{1}{2}$. As such, $\pi_2^x = \frac{\kappa}{2}$. And, average user utility is

$$u_2 = N_2^A(Y_2^A - p_2^A) + N_2^B(Y_2^B - p_2^B) - \bar{\kappa}_2 = K^x + \phi^x D^x + \frac{\gamma^x}{2} - \kappa - \frac{\hat{\kappa}}{4}.$$

As can be seen, in equilibrium, $\pi_2^x = \kappa/2$ does not depend on D^x and ϕ^x, whereas u_2 increases with D^x and ϕ^x.

In addition, we can use the expressions for π_2^x, i.e., the period-2 platform payoff under equilibrium pricing from Lemma 1, (or alternatively (B.7) and (B.8) for $\Delta_K = 0$ and $D^A = D^B$ as well as $\phi^A = \phi^B$) to derive

$$\frac{\partial \pi_2^x}{\partial D^x} = -\frac{\partial \pi_2^x}{\partial D^{-x}} = \frac{\phi^x}{3}. \tag{D.23}$$

D.1.2 Period $t = 1$

It follows that $\theta^A = \theta^B = 1$ (see Proposition 1). We note that $D^A = N_1^A I^A + \eta N_1^B I^B$ and $D^B = \eta N_1^A I^A + N_1^B I^B$ for $\eta \in [0, 1]$ as well as $c^x = (1 + \eta)c$. Also observe that

$$\pi_1^x = N_1^x \left[p_1^x - q^x I^x\right] + N_2^x p_2^x - \frac{\lambda (I^x)^2}{2}.$$

The two platforms $x = A, B$ solve

$$\max_{p_1^x, I^x, q^x} \pi_1^x,$$

taking the choice of the other platform $(p_1^{-x}, I^{-x}, q^{-x})$ as given.

Equilibrium prices in $t = 1$

We now calculate equilibrium prices in period $t = 1$, i.e., p_1^x for $x = A, B$, taking investments I^x as given. Recall $N_1^A = \hat{z}_1$ and $N_1^B = 1 - \hat{z}_1$ with $\hat{z}_1$ characterized in (6). That is,

$$N_1^A = \frac{1}{2} + \frac{-(p_1^A - p_1^B) + \left[I^A(q^A - c^A) - I^B(q^B - c^B)\right]}{2\kappa} \tag{D.24}$$
$$N_1^B = \frac{1}{2} + \frac{(p_1^A - p_1^B) - \left[I^A(q^A - c^A) - I^B(q^B - c^B)\right]}{2\kappa}.$$

Next, we calculate

$$\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^{-x}}{\partial p_1^x} = \frac{1}{2\kappa}$$

as well as

$$\frac{\partial D^{-x}}{\partial p_1^x} = \frac{I^{-x} - \eta I^x}{2\kappa} \quad \text{and} \quad \frac{\partial D^x}{\partial p_1^x} = \frac{\eta I^{-x} - I^x}{2\kappa}.$$

Thus, the first-order condition with respect to price p_1^x reads:

$$\frac{\partial \pi_1^x}{\partial p_1^x} = N_1^x - \frac{p_1^x}{2\kappa} + \frac{I^x q^x}{2\kappa} + \frac{\partial \pi_2^x}{\partial D^x}\left(\frac{\eta I^{-x} - I^x}{2\kappa}\right) + \frac{\partial \pi_2^x}{\partial D^{-x}}\left(\frac{I^{-x} - \eta I^x}{2\kappa}\right) = 0. \tag{D.25}$$

Using $N_1^x = \frac{1}{2}$, $I^x = I^{-x}$, and $\frac{\partial \pi_2^x}{\partial D^x} = \frac{\phi^x}{3} = -\frac{\partial \pi_2^x}{\partial D^{-x}}$ (see (D.23)), we can solve

$$p_1^x = \kappa + I^x\left(q^x - \frac{2(1-\eta)\phi^x}{3}\right). \tag{D.26}$$

The equilibrium price expression from Proposition 2 follows upon setting $\eta = 0$.

Equilibrium investments

We now calculate equilibrium investments $I^x = I^{-x}$, given the optimal period-1 pricing from (D.26). To start with, recall (D.24) and calculate the partial derivative of N_1^x with respect to investments/investments I^x, I^{-x} (holding p_1^x and p_1^{-x} fixed):

$$\frac{\partial N_1^x}{\partial I^x} = \frac{q^x - c^x}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^{-x}}{\partial I^x} = -\frac{q^x - c^x}{2\kappa}. \tag{D.27}$$

Thus,

$$\frac{\partial D^x}{\partial I^x} = N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x + \eta\left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}$$
$$\frac{\partial D^{-x}}{\partial I^x} = \eta\left(N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x\right) + \left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}.$$

Hence, the partial derivative of period-1 payoff π_1^x with respect to investment I^x becomes

$$\frac{\partial \pi_1^x}{\partial I^x} = \left(\frac{\partial N_1^x}{\partial I^x}\right)(p_1^x - I^x q^x) + \frac{\partial \pi_2^x}{\partial D^x}\left(N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x + \eta\left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}\right)$$
$$+ \frac{\partial \pi_2^x}{\partial D^{-x}}\left(\eta\left(N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x\right) + \left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}\right) - N_1^x q^x - \lambda I^x$$

If interior, i.e., $I^x \in (0, 1)$, optimal investment/effort solves the first-order condition $\frac{\partial \pi_1^x}{\partial I^x} = 0$.

Recall the optimal price p_1^x from (D.26) so that

$$p_1^x - I^x q^x = \kappa - \frac{2(1-\eta)\phi^x I^x}{3}.$$

Next, note that (in a symmetric equilibrium), $N_1^x = N_1^{-x} = 1/2$, $q^x = q^{-x}$, $I^x = I^{-x}$ as well as $\frac{\partial \pi_2^x}{\partial D^x} = \frac{\phi^x}{3} = -\frac{\partial \pi_2^x}{\partial D^{-x}}$ (see (D.23)), and $\frac{\partial N_1^x}{\partial I^x} = -\frac{\partial N_1^{-x}}{\partial I^x}$ (see (D.27). Using these relations, we obtain after simplifications:

$$\begin{aligned}
\frac{\partial \pi_1^x}{\partial I^x} &= \frac{\partial N_1^x}{\partial I^x}\left(\kappa - \frac{2(1-\eta)\phi^x I^x}{3}\right) - \frac{q^x}{2} \\
&\quad + \frac{\phi^x}{3}\left(\frac{1}{2} + \frac{\partial N_1^x}{\partial I^x} I^x (1-\eta)\right) - \frac{\phi^x}{3}\left(\frac{\eta}{2} - \frac{\partial N_1^x}{\partial I^x} I^x (1-\eta)\right) - \lambda I^x \\
&= \frac{\partial N_1^x}{\partial I^x}\left(\kappa - \frac{2(1-\eta)\phi^x I^x}{3}\right) - \frac{q^x}{2} + \frac{\phi^x}{3}\left(\frac{(1-\eta)}{2} + \left(\frac{\partial N_1^x}{\partial I^x}\right)(2I^x(1-\eta))\right) - \lambda I^x \\
&= \kappa\left(\frac{\partial N_1^x}{\partial I^x}\right) - \frac{q^x}{2} + \frac{\phi^x(1-\eta)}{6} - \lambda I^x.
\end{aligned}$$

Inserting $\frac{\partial N_1^x}{\partial I^x} = \frac{(q^x - c^x)}{2\kappa}$ into above expression for $\frac{\partial \pi_1^x}{\partial I^x}$, we obtain

$$\frac{\partial \pi_1^x}{\partial I^x} = \frac{1}{2}\left(\frac{\phi^x(1-\eta)}{3} - c^x - 2\lambda I^x\right).$$

As a result, equilibrium investment/effort satisfies — if it is interior and solves $\frac{\partial \pi_1^x}{\partial I^x} = 0$ —

$$I^x = \frac{\phi^x(1-\eta) - 3c^x}{6\lambda}.$$

Overall, we therefore obtain

$$I^x = \min\left\{1, \left[\frac{\phi^x(1-\eta) - 3c^x}{6\lambda}\right]^+\right\}. \tag{D.28}$$

The equilibrium investment expression from Proposition 2 follows upon setting $\eta = 0$. The equilibrium investment expression from Proposition 4 follows upon inserting $c^x = c(1+\eta)$.

D.1.3 Payoffs

We now calculate the payoff of platform x, using the derived expressions for prices and investment. Thus,

$$\pi_1^x = N_1^x\left[p_1^x - q^x I^x\right] + N_2^x p_2^x - \frac{\lambda (I^x)^2}{2} = \kappa - \frac{I^x \phi^x (1-\eta)}{3} - \frac{\lambda (I^x)^2}{2},$$

where we used that $N_1^x = \frac{1}{2}$ and the price expression $p_1^x = \kappa + I^x\left(q^x - \frac{2(1-\eta)I^x\phi^x}{3}\right)$ as well as $\pi_2^x = \kappa/2$.

Next, we calculate user welfare. We know from earlier results that users' total payoff in period $t = 2$ reads

$$\begin{aligned}
u_2 &= K^x + \phi^x D^x + \frac{\gamma^x}{2} - \kappa - \frac{\hat{\kappa}}{4} \\
&= K^x + \frac{\phi^x(1+\eta)I^x}{2} + \frac{\gamma^x}{2} - \kappa - \frac{\hat{\kappa}}{4},
\end{aligned}$$

where we used $D^x = (1+\eta)\hat{D}^x = (1+\eta)N_1^x I^x = 0.5(1+\eta)I^x$. Next, total welfare reads

$$\begin{aligned} u_1 &= K^x + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4} - p_1^x + u_2 + I^x(q^x - c^x) \\ &= K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4} + \frac{2I^x(1-\eta)\phi^x}{3} - I^x q^x + I^x(q^x - c^x) + u_2 \\ &= \underbrace{2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right)}_{\equiv const} + I^x\left(\frac{\phi^x(7-\eta)}{6} - c^x\right), \end{aligned} \tag{D.29}$$

whereby investment I^x was previously derived and is characterized in (D.28) and $c^=c(1+\eta)$. The first term is simply a model constant, which we denote by "*const*." As a result, we have that $I^x\left(\frac{\phi^x(7-\eta)}{6} - c^x\right) \geq 0$, with the inequality being strict if $I^x > 0$. Since I^x decreases with η, it also follows that total user welfare decreases with η, i.e., $\frac{\partial u_1}{\partial \eta} \leq 0$, and does so strictly when investment $I^x > 0$ is positive. We therefore conclude that required data sharing ($\eta > 0$) reduces user welfare relative to the baseline with $\eta = 0$, and does so strictly when $\phi^x > 3c$.

Finally, also note that any platform's available data in period $t = 2$ (after data sharing) reads $D^x = \frac{(1+\eta)I^x}{2}$, so that

$$\frac{\partial D^x}{\partial \eta} = \frac{I^x}{2} + \frac{1+\eta}{2}\frac{\partial I^x}{\partial \eta}$$

When $I^x \in (0,1)$, then

$$2\left(\frac{\partial D^x}{\partial \eta}\right) = \frac{\phi^x(1-\eta) - 3c(1+\eta)}{6\lambda} - \frac{(1+\eta)(\phi^x+3)}{6\lambda} < 0.$$

As such, when investment is interior (i.e., $I^x \in (0,1)$), then data sharing decreases the amount of data that platforms use in period $t = 2$.

D.2 Market for Data when Users own Data — Proof of Proposition 6

We start by solving the model at the beginning of state $t = 2$, given a total stock of data $D = N_1^A I^A + N_1^B I^B$. We solve for a symmetric equilibrium: Symmetry in equilibrium implies $N_1^x = 1/2$, $D^A = D^B$, and $I^A = I^B = I^x$, so $D = I^x$. At the beginning of period $t = 2$ (before prices are chosen), the two platforms simultaneously choose D^x to maximize

$$\max_{D^x \in [0,D]} \pi_2^x - cD^x, \tag{D.30}$$

taking the choice of the other platform, i.e. D^{-x}, as given. Here, the payoff π_2^x is characterized in Proposition 1. Recall (B.7) and (B.8), and observe that $\frac{\partial^2 \pi_2^x}{\partial (D^x)^2} > 0$. Thus, there exists no interior maximum to the optimization (D.30), and we therefore conjecture $D^x = D^{-x} = D$.

Next, we characterize optimal choice of D^x, given $D = \hat{D}^A + \hat{D}^B$. For this purpose, we use (B.7) and $D^x = D^{-x}$ — which holds in symmetric equilibrium — to take the derivative with respect to D^x in (D.30), yielding

$$\frac{\partial}{\partial D^x}(\pi_2^x - cD^x) = \frac{\phi^x}{3} - c$$

under $D^x = D^{-x}$. We now consider two distinct cases.

First, suppose that $c > \frac{\phi^x}{3}$. Then, for any level of $D \geq 0$, platforms optimally choose $D^x = 0$. Anticipating $D^x = 0$ for any levels of $D = \hat{D}^A + \hat{D}^B$, it is clear that platforms optimally do not exert any investment I^x to collect data, so $I^x = D = 0$. The platform payoff in period $t = 2$ then reads $\pi_2^x = \frac{\kappa}{2}$.

Second, suppose that $\phi^x \geq 3c$. In this case, $D_2^x = D$ and both platforms acquire all available data D. The platform payoff in period $t = 2$ then reads $\pi_2^x = \frac{\kappa}{2} - cD$. Moreover, using $D_2^x = D = I^x/2 + I^{-x}/2$ as well as the expression for π_2^x from Lemma 1, we have

$$\pi_2^x = \frac{\left[3\kappa + \phi^x(I^x + I^{-x})/2 - \phi^{-x}(I^x + I^{-x})/2\right]^2}{18\kappa}$$

so that $\frac{\partial \pi_2^x}{\partial I^x} = 0$ (since $\phi^x = \phi^{-x}$). It is immediate from (5) that $\frac{\partial \pi_2^x}{\partial I^x} = 0$ implies $\frac{\partial \pi_1^x}{\partial I^x} \leq 0$ with the inequality being strict for $I^x > 0$. As such, optimal data collection investment satisfies $I^A = I^B = 0$, so $D = D^x = 0$. In either case, we have established $I^x = D = D^x = 0$ in symmetric equilibrium.

Given that $I^x = 0$ in a symmetric equilibrium, the platform x solves in each period $t = 1, 2$:

$$\max_{p_t^x} N_t^x p_t^x,$$

taking the choice of the other platform, i.e., p_t^{-x}, as given. We have that

$$\frac{\partial N_t^x}{\partial p_t^x} = -\frac{1}{2\kappa} \quad \text{and} \quad \frac{\partial N_t^{-x}}{\partial p_t^x} = \frac{1}{2\kappa}.$$

Thus, price p_1^x solves the first-order condition $N_t^x + \frac{\partial N_t^x}{\partial p_t^x} p_t^x = 0$, which we can solve — using $N_t^x = 1/2$ — for $p_t^x = \kappa$ with $t = 1, 2$.

Platform x's total payoff at $t = 1$ reads $\pi_1^x = \kappa$, while $\pi_2^x = \kappa/2$. Next, we calculate user welfare. We know that users total payoff in period $t = 2$ reads

$$u_2 = K^x - \kappa - \frac{\hat{\kappa}}{4} + \phi^x D_2 = K^x - \kappa - \frac{\hat{\kappa}}{4}.$$

As such, total welfare reads

$$u_1 = 2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right),$$

and it is clear that u_1 is lower than under the baseline, which is characterized in (D.29) for $\eta = 0$, that is,

$$2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right) + I^x\left(\frac{7\phi^x}{6} - c^x\right) \geq u_1.$$

The inequality above is strict if and only if it holds in the baseline that $I^x > 0 \iff \phi^x > 3c$.

E Proofs and Derivations for Section 4

E.1 Proof of Proposition 8

According to the objective (11), the user union chooses at the beginning of time $t = 1$ (before investments and $t = 1$ prices are chosen) f to maximize $u_1 - fI^x$, whereby $c^x = c - f$ and $\eta = 0$. Next, recall the expression for user welfare from (D.29), that is,

$$u_1 = const + I^x\left(\frac{\phi^x(7-\eta)}{6} - c^x\right),$$

where we define for convenience

$$const := 2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right).$$

We consider $\eta = 0$ (no data sharing), and we rewrite the objective (11) as

$$\hat{u}_1 := u_1 - fI^x = const + I^x \left(\frac{7\phi^x}{6} - c + f \right) - fI^x = const + I^x \left(\frac{7\phi^x}{6} - c \right), \tag{E.31}$$

where I^x is from Proposition 2, that is,

$$I^x = \min \left\{ 1, \left[\frac{\phi^x - 3(c - f)}{6\lambda} \right]^+ \right\}. \tag{E.32}$$

If $6c \geq 7\phi^x$, then (E.31) immediately implies that the user union optimally implements $I^x = 0$, which is achieved by setting $f = 0$.

On the other hand, if $6c < 7\phi^x$, the user union optimally implements $I^x > 0$ and, in fact, the objective $\hat{u}_1 = u_1 - fI^x$ strictly increases with I^x, i.e., $\frac{\partial(u_1 - fI^x)}{\partial I^x} > 0$. As such, the relation (E.31) reveals that the user union optimally chooses f to maximize investment subject to $I^x \leq 1$ (exogenous upper bound of investment); thus, $I^x = 1$. Solving $I^x = 1$ (using (E.32)) for f yields

$$f = f^* = \frac{6\lambda + 3c - \phi^x}{3}.$$

which concludes the proof. The user welfare then reads

$$\hat{u}_1 = 2 \left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4} \right) + \frac{7\phi^x}{6} - c,$$

which is strictly larger than under the baseline (with $\eta = 0$) or data sharing (with $\eta > 0$) yielding user welfare:

$$2 \left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4} \right) + I^x \left(\frac{\phi^x(7 - \eta)}{6} - c(1 + \eta) \right).$$

Likewise, since we have shown in Proposition 6, that user welfare under the baseline is higher than when users own their data and sell their data in a market for data, it readily follows that welfare under user union is higher than under that scenario too.

E.2 Proof of Proposition 9

To begin with, note that $D^x = N_1^x I^x$ as well as $\theta^x = 1$, where by symmetry $N_1^x = 1/2$ and $I^x = I^{-x}$ so $D^x = D^{-x}$.

E.2.1 Prices

It holds that $N_1^x = N_1^{-x} = \hat{z}_1$, with

$$N_1^x = \frac{1}{2} - \frac{(p_1^x - p_1^{-x})}{2\kappa}.$$

As such, we can calculate

$$\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^{-x}}{\partial p_1^x} = \frac{1}{2\kappa}$$

and

$$\frac{\partial D^{-x}}{\partial p_1^x} = \frac{I^{-x}}{2\kappa} \quad \text{and} \quad \frac{\partial D^x}{\partial p_1^x} = \frac{-I^x}{2\kappa}$$

Thus, the first-order condition with respect to price p_1^x reads:

$$\frac{\partial \pi_1^x}{\partial p_1^x} = N_1^x - \frac{p_1^x}{2\kappa} + \frac{I^x q^x}{2\kappa} - \frac{\partial \pi_2^x}{\partial D^x}\left(\frac{I^x}{2\kappa}\right) + \frac{\partial \pi_2^x}{\partial D^{-x}}\left(\frac{I^{-x}}{2\kappa}\right) = 0. \tag{E.33}$$

Using $N_1^x = \frac{1}{2}$, $I^x = I^{-x}$, and $\frac{\partial \pi_2^x}{\partial D^x} = \frac{\phi^x}{3} = -\frac{\partial \pi_2^x}{\partial D^{-x}}$, we can solve

$$p_1^x = \kappa + I^x\left(q - \frac{2\phi^x}{3}\right).$$

E.2.2 Investment

To start with, note that because of

$$N_1^x = \frac{1}{2} - \frac{(p_1^x - p_1^{-x})}{2\kappa},$$

we have $\frac{\partial N_1^x}{\partial I^{x'}} = 0$ for all $x, x' \in \{A, B\}$. Thus,

$$\frac{\partial D^x}{\partial I^x} = N_1^x = \frac{1}{2} \quad \text{and} \quad \frac{\partial D^{-x}}{\partial I^x} = 0.$$

Hence, the derivative with respect to investment I^x becomes

$$\frac{\partial \pi_1^x}{\partial I^x} = \frac{\partial \pi_2^x}{\partial D^x}\left(\frac{\partial D^x}{\partial I^x}\right) - N_1^x q - \lambda I^x$$

If interior, i.e., $I^x \in (0,1)$, optimal investment solves the first-order condition $\frac{\partial \pi_1^x}{\partial I^x} = 0$. Using $N_1^x = 1/2$ and $\frac{\partial \pi_2^x}{\partial D^x} = \frac{\phi^x}{3}$, we obtain $I^x = \frac{\phi^x - 3q}{6\lambda}$, so optimal investment satisfies

$$I^x = \max\left\{1, \left[\frac{\phi^x - 3q}{6\lambda}\right]^+\right\}.$$

E.2.3 User Welfare

Using the expression for u_1 from (D.29) with $\eta = 0$ and $c^x = c$, total user welfare becomes

$$\hat{u}_1 = u_1 - qI^x = 2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right) + I^x\left(\frac{7\phi^x}{6} - c\right),$$

and depends on q^x only via investment I^x. When $6c \geq 7\phi^x$, then it is optimal to stipulate $I^x = 0$ which is achieved by setting $q = 0$.

When, on the other hand, $6c < 7\phi^x$, $\hat{u}_1$ strictly increases with I^x. As such, optimal investment maximizing user welfare $\hat{u}_1$ is either at the boundary 1 (exogenous upper boundary for investment) or such that the constraint $\pi_1^x \geq 0$ binds (platform participation constraint) and platforms just break even. Using arguments analogous to the ones used in the proof of Proposition 8, this leads to optimal investment $I^x = 1$. The price for data is the same for the two platforms and satisfies

$$q = \frac{\phi^x}{3} - 2\lambda I^x = \frac{\phi^x}{3} - 2\lambda.$$

Total user welfare then reads

$$\hat{u}_1 = u_1 - qI^x = 2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right) + \frac{7\phi^x}{6} - c,$$

which is strictly larger than under the baseline (with $\eta = 0$) or data sharing (with $\eta > 0$) yielding user welfare: $2\left(K^x - \kappa + \frac{\gamma^x}{2} - \frac{\hat{\kappa}}{4}\right) + I^x\left(\frac{\phi^x(7-\eta)}{6} - c(1+\eta)\right)$. Likewise, since we have shown in Proposition 6, that user welfare under the baseline is higher than when users own their data and sell their data in a market for data, it readily follows that welfare under user union is higher than under that scenario too.

E.3 Details for Section 4.3

Take the reward level of the user union f as given, and note $c^x = c - f$. As argued in the main text, given the reward f, the equilibrium from $t = 1$ onward is then characterized in Proposition 1 and Lemma 1. To induce $\theta^x = 1$, platform x sets $q^x \geq c^x$. Assume that given f and c^x, the continuation equilibrium exists and is unique (up to q^x), i.e., $\hat{z}_t$, I^x, p_2^x and $\bar{p}_1^x = p_1^x - I^x q^x$ are unique and do not depend on the exact level of q^x (as long as $\theta^x = 1$ is induced).

We now analyze under what conditions user union participation is privately optimal (i.e., incentive compatible) for any individual user z. To do so, suppose that individual user z deviates by not joining the user union at the beginning of time $t = 1$ while all other users $[0, 1] - \{z\}$ join the union; without loss of generality, consider that $z \leq \hat{z}_1$. Thus, user z saves the membership fee $m(z)$, but does not receive the reward for contributing data. Also, user z is quoted the same service prices p_t^x as other users (i.e., there is no service price discrimination), but faces potentially different data prices $\hat{q}^x$ than other users.

More formally, sser z's utility from the deviation is

$$u^{Dev}(z) = \max_{x\in\{A,B\},\hat{\theta}^x\in[0,1],\rho\in\{0,1\}} \rho\big(Y_1^x - p_1^x + \hat{\theta}^x I^x(\hat{q}^x - c) - \kappa^x(z)\big) + \max_{x\in\{A,B\}} \big(Y_2^x - p_2^x - \kappa^x(z)\big),$$

where ρ denotes z's decision to consume at any platform x and $\hat{\theta}^x \in [0, 1]$ is z's choice of contributing data at the platform x she joins. For simplicity, we already imposed that z participates in $t = 2$, as — per assumption — the entire market is covered ($N_t^A + N_t^B = 1$) and the deviation of z (and union membership) does not affect period-2 payoff $\max_{x\in\{A,B\}} \big(Y_2^x - p_2^x - \kappa^x(z)\big)$.

In contrast, user z's payoff from joining the union is

$$u^{Union}(z) = -m(z) + \max_{x\in\{A,B\},\hat{\theta}^x\in[0,1]} \big(Y_1^x - p_1^x + \hat{\theta}^x I^x(q^x - c + f) - \kappa^x(z)\big) + \max_{x\in\{A,B\}} \big(Y_2^x - p_2^x - \kappa^x(z)\big),$$

We already impose the assumption that in equilibrium the entire market is covered; thus, if z does not deviate, it must be that she adopts at least one platform. Notice that, being part of the union, z optimally shares data with the platform x she joins, so it is optimal to set $\hat{\theta}^x = 1$ in above maximization.

Without loss of generality, suppose that $z \leq \hat{z}_1$, so $m(z) = I^A f$ and

$$A = \arg\max_{x\in\{A,B\}} \big(Y_1^x - p_1^x + I^x(q^x - c + f) - \kappa^x(z)\big)$$

Then, the gain from deviating equals

$$\begin{aligned} \Delta^z &:= u^{Dev}(z) - u^{Union}(z) \\ &= -\big[Y_1^A - p_1^A + I^A(q^A - c) - \kappa^A(z)\big] + \max_{x\in\{A,B\},\rho\in\{0,1\},\hat{\theta}^x\in\{0,1\}} \rho\big(Y_1^x - p_1^x - \kappa^x(z) + \hat{\theta}^x I^x(\hat{q}^x - c)\big) \end{aligned} \tag{E.34}$$

Next, notice that by Proposition 1, we have $\frac{\partial p_1^x}{\partial q^x} = I^x$. That is, period-1 prices can be written in the form

$$p_1^x = \bar{p}_1^x + I^x q^x,$$

where $\bar{p}_1^x$ is unique (by assumption that the continuation equilibrium from Proposition 1 is unique)

and does not depend on $I^{x'}$ or $q^{x'}$, i.e., $\frac{\partial \bar{p_1}^x}{\partial I^{x'}} = \frac{\partial \bar{p_1}^x}{\partial q^{x'}} = 0$. Thus, inserting $p_1^x = \bar{p}_1^x + I^x q^x$, we obtain

$$\Delta^z = -\big[Y_1^A - \bar{p}_1^A - I^A c - \kappa^A(z)\big] + \max_{x\in\{A,B\},\rho\in\{0,1\},\hat{\theta}^x\in[0,1]} \rho\big(Y_1^x - p_1^x - \kappa^x(z) + \hat{\theta}^x I^x(\hat{q}^x - c)\big). \tag{E.35}$$

We now need to show $\Delta^z \leq 0$.

If user z is not member of the union, platform x can offer z a potentially different data price $\hat{q}^x$ (e.g., $\hat{q}^x < q^x$) than the data price q^x that it offers to union members. That is, we assume that the platform x can make the price for data that it pays user z contingent on union membership. We assume that the timing underlying the choice of $\hat{q}^x$ is as follows. First, after observing q^x, I^x, p_1^x, the user z joins a platform x; then, x observes whether z is user union member; if yes, z is quoted data price q^x and, if not, x can alter the data price that it quotes to z to any level $\hat{q}^x$ (i.e., x quotes a new price $\hat{q}^x$ to the deviant). When user z joins platform x and x observes that z is not member, it is optimal for x to minimize payments $\hat{q}^x$ whilst inducing z to set $\hat{\theta}^x = 1$. As such, it is optimal for x to choose $\hat{q}^x \leq c$, i.e., not to pay the deviant z more than her required cost of sharing data c. Thus, if x induces z to share data, then $\hat{q}^x = c$. Overall, it follows that $\max_{\hat{\theta}^x\in[0,1]} \hat{\theta}^x I^x(\hat{q}^x - c) = 0 = I^x[\hat{q}^x - c]^+ = 0$ (recall that $[\cdot]^+ = \max\{\cdot, 0\}$). In other words, it is optimal for the platform not to leave rents to the deviant by stipulating $\hat{q}^x = c$.

We now show under what circumstances $\Delta^z \leq 0$ and user union is incentive compatible. We do so separately when $I^A, I^B > 0$ in equilibrium (see Part I) and when $I^A \cdot I^B = 0$ in equilibrium (see Part II)

Part I

We now establish under what circumstances $\Delta^z \leq 0$ when $I^A, I^B > 0$. First, when $\rho = 0$ (i.e., the deviant stays out of the market in $t = 1$), then $\Delta^z = -\big[Y_1^A - \bar{p}_1^A - I^A c - \kappa^A(z)\big]$, which must be negative by the assumption that, conditional on union membership, user z derives positive payoff and participates.[52]

When $\rho = 1$ and

$$A \in \arg\max_{x\in\{A,B\}} \max_{\rho\in\{0,1\},\hat{\theta}^x\in[0,1]} \rho\big(Y_1^x - p_1^x - \kappa^x(z) + \hat{\theta}^x I^x(\hat{q}^x - c)\big)$$

then (according to (E.35))

$$\Delta^z = -I^A(q^A - c) + I^A[\hat{q}^A - c]^+ = -I^A(q^A - c),$$

which is negative for $q^A \geq c$.

When on the other hand

$$B \in \arg\max_{x\in\{A,B\}} \max_{\rho\in\{0,1\},\hat{\theta}^x\in[0,1]} \rho\big(Y_1^x - p_1^x - \kappa^x(z) + \hat{\theta}^x I^x(\hat{q}^x - c)\big)$$

and $\rho = 1$, then (according to (E.35))

$$\Delta_z = -\big[Y_1^A - \bar{p}_1^A - I^A c - \kappa^A(z)\big] + \big(Y_1^B - \bar{p}_1^B - q^B I^B - \kappa^B(z)\big),$$

which is negative for

$$q^B \geq \frac{(Y_1^B - Y_1^A) - (\bar{p}_1^B - \bar{p}_1^A) - \big(\kappa^B(z) - \kappa^A(z)\big) + I^A c}{I^B}.$$

[52] This must be the case because we assume that the entire market is covered in equilibrium.

Thus, we have $\Delta^z \leq 0$ for $z \leq \hat{z}_1$ if

$$q^A \geq c \quad \text{and} \quad q^B \geq \frac{(Y_1^B - Y_1^A) - (\bar{p}_1^B - \bar{p}_1^A) - \big(\kappa^B(z) - \kappa^A(z)\big) + I^A c}{I^B}.$$

Analogously, we have $\Delta^z \leq 0$ for $z > \hat{z}_1$ if

$$q^B \geq c \quad \text{and} \quad q^A \geq \frac{(Y_1^A - Y_1^B) - (\bar{p}_1^A - \bar{p}_1^B) - \big(\kappa^A(z) - \kappa^B(z)\big) + I^B c}{I^A}.$$

At the same time, to incentivize $\theta^x = 1$, we must have $q^x \geq c - f$.

Altogether, $\Delta^z \leq 0$ and $q^x \geq c - f$ hold for all $z \in [0,1]$ if

$$\begin{aligned} q^A &\geq c + \max_{z\in[0,1]}\left[\max\left\{-f, \frac{(Y_1^A - Y_1^B) - (\bar{p}_1^A - \bar{p}_1^B) - \big(\kappa^A(z) - \kappa^B(z)\big) + I^B c}{I^A}\right\}\right] \\ q^B &\geq c + \max_{z\in[0,1]}\left[\max\left\{-f, \frac{(Y_1^B - Y_1^A) - (\bar{p}_1^B - \bar{p}_1^A) - \big(\kappa^B(z) - \kappa^A(z)\big) + I^A c}{I^B}\right\}\right]. \end{aligned} \tag{E.36}$$

As such, if (E.36) holds in equilibrium, then union membership is incentive compatible. Note that (E.36) can be seen as incentive compatibility condition for user union membership.

Recall that the exact values of q^x are not payoff-relevant in a sense made precise in Proposition 1, i.e., the value of q^x does not affect $\hat{z}_t$, p_2^x, $\bar{p}_1^x$, I^x, or u_1. Notice that the right-hand-side of both inequalities in (E.36) does not depend on q^A or q^B and, in particular, involves equilibrium quantities only depending on model parameters. As such, we can always find ("large enough") equilibrium levels of q^A and q^B that satisfy (E.36). That is, provided $I^A, I^B > 0$, there exists an equilibrium (unique up to q^x) in which all users are members of the union and user union participation is incentive compatible in equilibrium (i.e., (E.36) holds). Because, given a level of f and $c^x = c - f$, the continuation equilibrium (outlined in Proposition 1 and Lemma 1) exists with unique $\hat{z}_t$, $\bar{p}_1^x$, p_2^x, and I^x, the user union equilibrium is unique too up to the level of q^x.

E.3.1 Part II

We now establish under what circumstances $\Delta^z \leq 0$ when $I^A \cdot I^B = 0$, i.e., when $I^A = 0$ or $I^B = 0$ or both.

First, when $\rho = 0$ (i.e., the deviant stays out of the market in $t = 1$), then $\Delta^z = -\big[Y_1^A - \bar{p}_1^A - I^A c - \kappa^A(z)\big]$, which must be negative by the assumption that, conditional on union membership, user z derives positive payoff and participates.[53]

Next, when $I^A = 0$, then $p_1^A = \bar{p}_1^A$ as well as

$$A = \arg\max_{x\in\{A,B\}} \big(Y_1^x - p_1^x + I^x(q^x - c + f) - \kappa^x(z)\big),$$

which — owing to $I^B(q^B - c + f) \geq 0$ — implies (for any $z \leq \hat{z}_1$)

$$A = \arg\max_{x\in\{A,B\}} \max_{\hat{\theta}^x\in[0,1]} \big(Y_1^x - p_1^x - \kappa^x(z) + \underbrace{\hat{\theta}^x I^x(\hat{q}^x - c)}_{=0}\big).$$

Inserting this relation into (E.35), we obtain $\Delta^z \leq 0$.

Next, consider $I^B = 0$. Then,

$$A = \arg\max_{x\in\{A,B\}} \big(Y_1^x - p_1^x + I^x(q^x - c + f) - \kappa^x(z)\big),$$

[53] This must be the case because we assume that the entire market is covered in equilibrium.

implies

$$A = \arg \max_{x \in \{A,B\}} \max_{\hat{\theta}^x \in [0,1]} \left(Y_1^x - p_1^x - \kappa^x(z) + \underbrace{\hat{\theta}^x I^x(\hat{q}^x - c)}_{=0} \right)$$

for any $z \leq \hat{z}_1$ if $q^A - c + f = 0$. When $I^B = 0$ and $q^A = c - f$, then $\Delta^z = I^A f \leq 0$ if $f \leq 0$ (see (E.34)).

Analogously, when $I^A = 0 < I^B$, then $\Delta^z \leq 0$ for any $z > \hat{z}_1$ holds if $f \leq 0$ and $q^B = c - f$. As such, when $I^A = 0$ or $I^B = 0$, user union is incentive compatible if $q^x = c - f$ as well as $f \leq 0$. Under these circumstances, the continuation equilibrium (outlined in Proposition 1 and Lemma 1) exists with unique $\hat{z}_t$, $\bar{p}_1^x$, p_2^x, and I^x.

F Extended Discussions

F.1 Market for Data When Platforms own Data — Model Variant with "Symmetric" Equilibrium

Assume ex-ante symmetry, i.e., $\Delta_K = 0$ as well as $\phi^A = \phi^B$. We now introduce a model variant with a market for data, in which platforms own and trade user-generated data. Notably, we make additional assumptions such that the model variant features a symmetric equilibrium in the subgame in period $t = 1$.

F.1.1 Stage 2

At the beginning of time period $t = 2$, platform A and B trade data with each other, with endogenous price p_2^D. Without loss of generality, it suffices to consider that platform A and B choose the optimal allocation of data through Nash Bargaining at the beginning of period $t = 2$, with equal bargaining weights $1/2$. The price for data p_2^D is then chosen to implement the split of resulting surplus. According to Lemma 3, total surplus is maximized when one platform x shares data with platform $-x$, but not the other way around.

We now derive the Nash bargaining solution. For this sake, call B the platform that shares data with the other one. That is, B shares data with platform A, but not the other way around, with $\hat{D}^A \phi^A \geq \hat{D}^B \phi^B$. Thus, $D^A = \hat{D}^A + \hat{D}^B$, whereas $D^B = \hat{D}^B$. Using the period-2 platform payoff under equilibrium pricing from Lemma 1 with $D^A = \hat{D}^A + \hat{D}^B$ and $D^B = \hat{D}^B$, we obtain the following platform payoffs just after Nash bargaining but before choosing price p_2^A and p_2^B:

$$\pi_2^A = \frac{(3\kappa + \phi^A \hat{D}^A)^2}{18\kappa} \quad \text{and} \quad \pi_2^B = \frac{(3\kappa - \phi^A \hat{D}^A)^2}{18\kappa}.$$

Then, just after the data trade, total surplus of both platforms (excluding user surplus) reads

$$S := \pi_2^A + \pi_2^B = \frac{(3\kappa + \phi^A \hat{D}^A)^2 + (3\kappa - \phi^A \hat{D}^A)^2}{18\kappa}$$

If there were no data sharing, then $D^x = \hat{D}^x$ and platforms' period-2 payoff under equilibrium pricing would be according to Lemma 1:

$$\hat{\pi}_2^x = \frac{(3\kappa + \phi^x \hat{D}^x - \phi^{-x} \hat{D}^{-x})^2}{18\kappa}.$$

Under these circumstances, total surplus would be

$$\hat{\pi}_2^x + \hat{\pi}_2^{-x} = \frac{(3\kappa + \phi^x \hat{D}^x - \phi^{-x} \hat{D}^{-x})^2 + (3\kappa + \phi^{-x} \hat{D}^{-x} - \phi^x \hat{D}^x)^2}{18\kappa}.$$

Thus, surplus generated through the efficient (ex-post) allocation of data is $S - (\hat{\pi}_2^A + \hat{\pi}_2^B)$. As per

Nash bargaining protocol (among the two parties $x = A$ and $x = B$) with equal bargaining weights $1/2$, the payoff of platform x becomes

$$\tilde{\pi}_2^x := \frac{1}{2}(S - \hat{\pi}_2^A - \hat{\pi}_2^B) + \hat{\pi}_2^x,$$

which is the payoff at the beginning of period $t = 2$ before data trade happens.

As such, the payoff of platform A at inception of period $t = 2$ reads

$$\begin{aligned}\tilde{\pi}_2^A &:= \left(\frac{(3\kappa + \phi^A \hat{D}^A)^2 + (3\kappa - \phi^A \hat{D}^A)^2 + (3\kappa + \phi^A \hat{D}^A - \phi^B \hat{D}^B)^2 - (3\kappa + \phi^B \hat{D}^B - \phi^A \hat{D}^A)^2}{36\kappa}\right)\\ &= \frac{(\phi^A \hat{D}^A)^2 + 6\kappa(\phi^A \hat{D}^A - \phi^B \hat{D}^B)}{18\kappa}.\end{aligned}$$

Likewise, the payoff of platform B at inception of period $t = 2$ becomes

$$\begin{aligned}\tilde{\pi}_2^B &:= \left(\frac{(3\kappa + \phi^A \hat{D}^A)^2 + (3\kappa - \phi^A \hat{D}^A)^2 + (3\kappa + \phi^B \hat{D}^B - \phi^A \hat{D}^A)^2 - (3\kappa + \phi^A \hat{D}^A - \phi^B \hat{D}^B)^2}{36\kappa}\right)\\ &= \frac{(\phi^A \hat{D}^A)^2 + 6\kappa(\phi^B \hat{D}^B - \phi^A \hat{D}^A)}{18\kappa}.\end{aligned}$$

There does not exist a fully symmetric equilibrium. We therefore look for an equilibrium in which both platforms enter period $t = 2$ symmetrically, i.e., $N_1^x = 1/2$ and $\hat{D}^x = I^x/2$, and, with equal probability of $1/2$, A and B are the data "buyers" and "sellers" respectively in the trade of data. Formally, at the beginning of period $t = 2$, nature determines which platform is buyer and seller of data, where each platform is selected by nature with equal probability. After that draw, we possibly relabel platforms such that A is data buyer and B is data seller. This assumption greatly simplifies the analysis. Notice that in the general case, platform A as the "stronger" platform will act as buyer of data with probability one, while B is seller with probability one. However, in the case of this section, not the strength but nature determines who buys/sells data; in the equilibrium we consider, both platforms enter stage 2 symmetrically, so the choice of nature is consistent with the result of Lemma 3 and its implications for the optimal data trade.

From the perspective of platform x, at the very beginning of period $t = 1$, there is a draw (by nature) resulting into $x = A$ (i.e., $-x = B$) and $x = B$ (i.e., $-x = A$) with equal probability $1/2$. Then, above expressions imply that platform x's expected payoff just before this draw becomes

$$\begin{aligned}\bar{\pi}_2^x &= \frac{1}{2}\left(\frac{(\phi^x \hat{D}^x)^2 + 6\kappa(\phi^x \hat{D}^x - \phi^{-x} \hat{D}^x)}{18\kappa} + \frac{(\phi^{-x} \hat{D}^{-x})^2 + 6\kappa(\phi^x \hat{D}^x - \phi^{-x} \hat{D}^{-x})}{18\kappa}\right)\\ &= \frac{(\phi^x \hat{D}^x)^2}{36\kappa} + \frac{\phi^x \hat{D}^x - \phi^{-x} \hat{D}^x}{3} + \frac{(\phi^{-x} \hat{D}^{-x})^2}{36\kappa}. \qquad \text{(F.37)}\end{aligned}$$

We focus on a subgame perfect equilibrium in which both platforms choose in period $t = 1$ the triple (q^x, p_1^x, I^x) simultaneously to maximize $\bar{\pi}_2^x$. In period $t = 2$, each platform is selected with probability $1/2$ to buy data from the other one, data trade occurs, and after that platforms simultaneously set prices p_2^x. In equilibrium in the subgame in period $t = 1$, the allocation is indeed symmetric and, as it will turn out, platforms x will exert symmetric level of investment and set symmetric prices in stage $t = 1$. The equilibrium in subgame in $t = 2$ — as previously discussed — is no more symmetric.

F.2 Stage $t = 1$

We now study platforms' optimal choice of prices and investments, given the continuation payoff in period $t = 2$, which is characterized in (F.37) and is denoted $\bar{\pi}_2^x$. For this sake, platform x

maximizes

$$\pi_1^x := N_1^x p_1^x - q^x N_1^x I^x + \bar{\pi}_2^x - \frac{\lambda (I^x)^2}{2}, \tag{F.38}$$

where $\bar{\pi}_2^x$ is from (F.37).

F.2.1 Prices

We set $\theta^x = 1$. We start by analyzing optimal pricing in $t = 1$. First, note that $\hat{D}^x = I^x N_1^x = I^x/2$. One calculates using $\bar{\pi}_2^x$ from (F.37):

$$\frac{\partial \bar{\pi}_2^x}{\partial \hat{D}^x} = \frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3} \quad \text{and} \quad \frac{\partial \bar{\pi}_2^x}{\partial \hat{D}^{-x}} = -\left(\frac{(\phi^{-x})^2}{36\kappa} + \frac{\phi^{-x}}{3}\right). \tag{F.39}$$

Next, we calculate (using (D.24)):

$$\frac{\partial N_1^x}{\partial p_1^x} = -\frac{1}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^{-x}}{\partial p_1^x} = \frac{1}{2\kappa}.$$

Thus, the first-order condition of π_1^x from (F.38) with respect to price p_1^x reads:

$$\frac{\partial \pi_1^x}{\partial p_1^x} = N_1^x - \frac{p_1^x}{2\kappa} + \frac{I^x q^x}{2\kappa} + \frac{\partial \pi_2^x}{\partial \hat{D}^x}\frac{\partial \hat{D}^x}{\partial N_1^x}\frac{\partial N_1^x}{\partial p_1^x} + \frac{\partial \pi_2^x}{\partial \hat{D}^{-x}}\frac{\partial \hat{D}^{-x}}{\partial N_1^{-x}}\frac{\partial N_1^{-x}}{\partial p_1^x} = 0. \tag{F.40}$$

From the above relations, we know that — with $I^x = I^{-x}$, $\phi^x = \phi^{-x}$, $\frac{\partial \pi_2^x}{\partial \hat{D}^x} = -\frac{\partial \pi_2^x}{\partial \hat{D}^{-x}}$ and $\frac{\partial N_1^x}{\partial p_1^x} = -\frac{\partial N_1^{-x}}{\partial p_1^x}$ as well as $\frac{\partial \hat{D}^x}{\partial N_1^x} = I^x$.

Thus, using $N_1^x = \frac{1}{2}$, $I^x = I^{-x}$, the first-order condition (F.40) becomes

$$\frac{\partial \pi_1^x}{\partial p_1^x} = \frac{1}{2} - \frac{p_1^x}{2\kappa} + \frac{I^x q^x}{2\kappa} - \frac{I^x}{\kappa}\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right) = 0.$$

We can solve above equation for period-1 price

$$p_1^x = \kappa + I^x q^x - 2I^x\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right). \tag{F.41}$$

F.2.2 Investment

Next, we turn to solving for investment I^x. For this sake, we calculate

$$\frac{\partial N_1^x}{\partial I^x} = \frac{(q^x - 1.5c)}{2\kappa} \quad \text{and} \quad \frac{\partial N_1^{-x}}{\partial I^x} = -\frac{(q^x - 1.5c)}{2\kappa},$$

noting that $c^x = 1.5c$. Observe that when z shares data with platform x, then, with probability $1/2$, platform x does not sell data at $t = 2$ and the user's privacy cost is c and otherwise, with probability $1/2$, x sells all the data to $-x$ and the realized privacy cost is $2c$. Thus, on average, the user's privacy cost for sharing one unit of data is $c^x = 1.5c$.

We recall that $\hat{D}^x = I^x N_1^x$, so that

$$\frac{\partial \hat{D}^x}{\partial I^x} = N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x \quad \text{and} \quad \frac{\partial \hat{D}^{-x}}{\partial I^x} = \left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}.$$

Hence, the derivative of payoff in period $t = 1$ with respect to investment I^x becomes

$$\frac{\partial \pi_1^x}{\partial I^x} = \left(\frac{\partial N_1^x}{\partial I^x}\right)(p_1^x - I^x q^x) + \frac{\partial \bar{\pi}_2^x}{\partial \hat{D}^x}\left(N_1^x + \frac{\partial N_1^x}{\partial I^x} I^x\right)$$
$$+ \frac{\partial \bar{\pi}_2^x}{\partial D^{-x}}\left(\left(\frac{\partial N_1^{-x}}{\partial I^x}\right) I^{-x}\right) - N_1^x q^x - \lambda I^x$$

If interior, i.e., $I^x \in (0,1)$, optimal investment solves the first-order condition $\frac{\partial \pi_1^x}{\partial I^x} = 0$.

Recall the price p_1^x from (F.41) so that

$$p_1^x - I^x q^x = \kappa - 2I^x\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right) = \kappa - 2I^x\left(\frac{\partial \bar{\pi}_2^x}{\partial \hat{D}^x}\right),$$

where the last equality uses (F.39). Next, note that $N_1^x = N_1^{-x} = 1/2$, $I^x = I^{-x}$ as well as $\frac{\partial \pi_2^x}{\partial D^x} = \frac{\phi^x}{3} = -\frac{\partial \pi_2^x}{\partial D^{-x}}$, and $\frac{\partial N_1^x}{\partial I^x} = -\frac{\partial N_1^{-x}}{\partial I^x}$. Using these relations, we obtain after simplifications:

$$\frac{\partial \pi_1^x}{\partial I^x} = \frac{\partial N_1^x}{\partial I^x}\left(\kappa - 2I^x\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right)\right) - \frac{q^x}{2}$$
$$+ \left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right)\left(\frac{1}{2} + \frac{\partial N_1^x}{\partial I^x} I^x\right) + \left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right)\left(\frac{\partial N_1^x}{\partial I^x} I^x\right) - \lambda I^x$$
$$= \kappa\left(\frac{\partial N_1^x}{\partial I^x}\right) - \frac{q^x}{2} + \frac{1}{2}\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right) - \lambda I^x$$

Inserting

$$\frac{\partial N_1^x}{\partial I^x} = \frac{(q^x - 1.5c)}{2\kappa}$$

into above expression for π_1^x, we obtain:

$$\frac{\partial \pi_1^x}{\partial I^x} = \frac{1}{2}\left(\left(\frac{(\phi^x)^2}{36\kappa} + \frac{\phi^x}{3}\right) - 1.5 - 2\lambda I^x\right).$$

As a result, equilibrium investment/effort satisfies — if it is interior and solves the first-order condition $\frac{\partial \pi_1^x}{\partial I^x} = 0$ —

$$I^x = \frac{\phi^x\left(1 + \frac{\phi^x}{12\kappa}\right) - 4.5c}{6\lambda}.$$

That is, optimal investment reads

$$I^x = \min\left\{1, \left[\frac{\phi^x\left(1 + \frac{\phi^x}{12\kappa}\right) - 4.5c}{6\lambda}\right]^+\right\}.$$

Propositions 2 and 4 readily imply that, with a market for data, the investment I^x above is larger than under the baseline and under data sharing (for $\eta > 0$) when c is sufficiently small or ϕ^x is large. Under these circumstances, the service price p_1^x from (F.41) is strictly lower than in the baseline (see Proposition 2) or with data sharing (see Proposition 4), holding q^x fixed in the comparison.

F.3 Commitment Solutions

Commitment to data sharing. Suppose the platform can commit to a future data sharing policy, $\eta^x \in [0,1]$ in period $t = 1$. Formally, in $t = 2$, platforms maximize (as before) $\max_{p_2^x} \pi_2^x$

whereby $D^x = N_1^x \theta^x I^x + \eta^{-x} N_1^{-x} \theta^{-x} I^{-x}$. In $t = 1$, platforms choose simultaneously η^x, p_1^x, I^x, q^x to maximize π_1^x, i.e., they solve $\max_{\eta^x, p_1^x, I^x, q^x} \pi_1^x$. We now show that, when $c^x = c(1 + \eta^x) \geq 0$, then $\eta^x = 0$ holds in equilibrium. For this sake, we consider without loss of generality $q^x = c^x$ so that $\theta^x = 1$, and calculate

$$\frac{\partial \pi_1^x}{\partial \eta^x} = -N_1^x I^x + \frac{\partial \pi_2^x}{\partial D_2^{-x}} N_1^x I^x,$$

which is (strictly) negative (when $I^x > 0$), because the expression for period-2 payoff π_2^x (under equilibrium pricing) in Lemma 1 readily imply $\frac{\partial \pi_2^x}{\partial D_2^{-x}} < 0$. Thus, platforms optimally choose $\eta^x = 0$.

Commitment to product/service quantities. Suppose that the platforms can commit to a minimum future quantity at the beginning. Then compared to the baseline, (i) both the platforms generates higher payoff; (ii) the platform that makes commitment take a larger market share; (iii) the platform that does not commit determines a higher price. We prove these next:

Proposition 10. *In equilibrium, one and only one platform commits to quantity, i.e., the solution involves $\{C, NC\}$ or $\{NC, C\}$. Without loss of generality, suppose platform A chooses 'C' and B chooses 'NC', then*

$$\begin{aligned}
&N_2^A = \hat{z}_2 = \frac{5}{8} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{8\kappa}, \\
&p_2^A = \frac{5\kappa}{4} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{4}, \qquad p_2^B = \frac{3\kappa}{2} - \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{2}, \\
&\pi_2^A = \frac{(5\kappa + \Delta_K + \phi^A D^A - \phi^B D^B)^2}{32\kappa}, \qquad \pi_2^B = \frac{(3\kappa - \Delta_K - \phi^A D^A + \phi^B D^B)^2}{16\kappa}.
\end{aligned} \tag{F.42}$$

Proof. We first prove that when platform $-x$ chooses 'NC', i.e. not to make a commitment, then the best response for platform x is to commit.

Without loss of generality, suppose platform B chooses 'NC'. If platform A also chooses 'NC', then the case is the same as the baseline, i.e. $p_2^A = \kappa + \frac{\Delta_K}{3} + \frac{D^A\phi^A - D^B\phi^B}{3}$, $N_2^A = \hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + D^A\phi^A - D^B\phi^B}{6\kappa}$. Then consider the case that platform A chooses 'C'. Given andy possible p_2^B, platform A need to decide N_2^A in period $t = 1$ to maximize π_2^A. To fullfill the commitment in period $t = 2$, platform A's product pricing, p_2^A, is restricted to satisfy

$$\begin{aligned}
\hat{z}_2 &= \frac{1}{2} + \frac{\Delta_K - p_2^A + p_2^B + D^A\phi^A - D^B\phi^B}{2\kappa}, \\
\Rightarrow p_2^A &= \Delta_K + p_2^B + D^A\phi^A - D^B\phi^B + \kappa(1 - 2N_2^A).
\end{aligned} \tag{F.43}$$

Then platform A's optimization problem is

$$\max_{N_2^A} \pi_2^A = p_2^A N_2^A = [\Delta_K + p_2^B + D^A\phi^A - D^B\phi^B + \kappa(1 - 2N_2^A)]N_2^A.$$

Thus we obtain that the optimal response is

$$N_2^{A*} = N_2^{A*}(p_2^B) = \frac{1}{4} + \frac{\Delta_K + p_2^B + D^A\phi^A - D^B\phi^B}{4\kappa}. \tag{F.44}$$

Consider platform B. In period $t = 2$, given platform A's commitment N_2^{A*}, then $N_2^B = 1 - N_2^{A*}$ is also given. Platform B foresees that platform A will make quality commitments based on p_2^B.[54]

[54] Note that the commitment only limits the minimum level of service on platform A. Thus platform B does not have unlimited access to higher prices.

Therefore, platform B faces the following problem:

$$\max_{p_2^B} \pi_2^B = p_2^B(1 - N_2^{A*}(p_2^B)) = p_2^B \left(\frac{3}{4} - \frac{\Delta_K + p_2^B + D^A\phi^A - D^B\phi^B}{4\kappa} \right).$$

We then obtain that the optimal pricing for platform B is

$$p_2^B = \frac{3\kappa}{2} - \frac{\Delta_K}{2} - \frac{\phi^A D^A - \phi^B D^B}{2}. \tag{F.45}$$

Plug (F.45) into (F.43) and (F.44), we have

$$\begin{aligned}
N_2^A &= \hat{z}_2 = \frac{5}{8} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{8\kappa}, \\
p_2^A &= \frac{5\kappa}{4} + \frac{\Delta_K + \phi^A D^A - \phi^B D^B}{4}, \\
\pi_2^A &= \frac{(5\kappa + \Delta_K + \phi^A D^A - \phi^B D^B)^2}{32\kappa} > \frac{(3\kappa + \Delta_K + \phi^A D^A - \phi^B D^B)^2}{18\kappa} = \pi_{2\ baseline}^A, \\
\pi_2^B &= \frac{(3\kappa - \Delta_K - \phi^A D^A + \phi^B D^B)^2}{16\kappa} > \frac{(3\kappa - \Delta_K - \phi^A D^A + \phi^B D^B)^2}{18\kappa} = \pi_{2\ baseline}^B.
\end{aligned} \tag{F.46}$$

Therefore, the best response for platform A is to make a quantity commitment, i.e. the best response to 'NC' is 'C'. Interestingly, both platforms increase payoff, implying that the best response to 'C' is 'NC'. Moreover, under the sufficient condition that $3\kappa > \max\{\Delta_K + \phi^A - \phi^B, -\Delta_K - \phi^A + \phi^B\}$, N_2^A is larger than the baseline equilibrium N_2^A, and p_2^B is larger than the baseline equilibrium. □

F.4 Institutional Background on Privacy Protection Policy and Open Data Initiatives

Privacy protection policies such as GDPR strengthen individual ownership rights over personal data by granting rights to access, correct, and delete personal data held by firms. While GDPR is a sweeping initiative implemented by the European Union, the U.S. system is piecemeal and multi-layered, with regional initiatives such as CCPA. Generally speaking, firms must minimize personal data processing and can only process personal data under limited and specific circumstances. One such circumstance is an individual's explicit opt-in consent.[55]

While policies such as GDPR have focused on data ownership rights, open data initiatives, e.g., in the form of data sharing initiatives, have emphasized data access. Private data ownership rights should not be confused with access. Privately owned data can allow open access while non-proprietary assets can be de facto closed for access (Merges, 2008). The question of whether or not data should be openly accessible (purely open access to data) is a debated issue in academia (e.g.,

[55] Federal laws on privacy protection tend to be industry and region specific in the United States, with the Department of Health Human Resources (DHHS) enforcing the Health Insurance Portability and Accountability Act of 1996 (HIPPA) in healthcare, the Federal Communication Commission (FCC) regulating telecommunication services, the federal reserve systems monitoring the financial sector through Gramm-Leach-Bliley Act (GLBA), the Security and Exchange Commission (SEC) focusing on public firms and financial exchanges, and the Department of Homeland Security (DHS) dealing with terrorism and cybercrimes related to national security. The Federal Trade Commission (FTC) can address privacy violations and inadequate data security as deceptive and unfair practice, following the 1914 FTC Act whereas the U.S. Constitution, in particular the First and Fourth Amendments, together with the Electronic Communications Privacy Act of 1986 (ECPA), the Stored Communications Act (1986), the Pen Register Act (1986), and the 2001 USA Patriot Act – stipulate when and how the government can collect and process electronic information of individuals. But in practice, it is still case by case. The debate on whether the United States should follow European-style regulation is still ongoing. See "Ad world flocks to Congress urging federal data privacy legislation", The Drum, 26 February 2019, and "Should Congress override state privacy rules? Not so fast," The Washington Post, February 26, 2019.

Dewald, Thursby, and Anderson, 1986) and policy (Commission et al., 2017). Advocates of open access to data argue that it facilitates subsequent research, including replication of existing works, and increases the diffusion of knowledge thereby enhancing the efficiency of the research system (Piwowar, Day, and Fridsma, 2007; Glenn and Ellis Lee, 2012; Piwowar and Vision, 2013). Empirical studies also document benefits of open data (Jetzek, Avital, and Bjorn-Andersen, 2012; Martens et al., 2020).[56] Open data initiatives take various forms. Centralized data commons suffer from privacy issues (Milles, 2019). Real world example includes UPI in India. The Indian authorities have the concept of a data fiduciary, which would be a body that would basically manage the data of individuals and help give them effective control over consent management. South Korean MyData and Brazil's open banking systems use similar setups. The New York Times conducted full-scale investigation in 2018 concerning Facebook (now Meta) forming ongoing partnerships with other firms, including Netflix, Apple, and Microsoft, and granting these companies access to different aspects of consumer data. See full news coverage at https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html. Blockchains and secure-MPC through ZKP etc., constitute an interesting route to explore. Overall, the challenge for data sharing is not only about the technology that enables privacy protection, but also about economic incentives.[57]

Recently, there have been many attempts to promote open data access, including Open Banking and Open Finance initiatives (He et al., 2022; Goldstein et al., 2022). For example, the International Data Spaces Association constitutes a private investment for secure data sharing (Richter and Slowinski, 2019). The EU proposed the Digital Market Act (DMA) in 2020, explicitly emphasizing data sharing for a fair competition.[58] China and South Korea have built open platforms for data sharing to aggregate scattered, isolated, and varied data to help integrate technology and business data to lower information barriers.[59]

Data companies such as Acxiom and Datalogix gather and sell personal information. Policy discussions often suggest that requiring large digital platforms to share data with smaller ones breaks their dominance and leads to more competition among platforms, which is beneficial for users. Such rationale, for instance, underlies the open banking regulation where banks have to share data with FinTech companies (e.g., lenders).

[56] Martens et al. (2020) empirically examine the effectiveness of mandatory data sharing. They find that user welfare is not maximized due to increases in product price, which corroborates our model prediction.

[57] Federated learning and privacy-preserving data sharing infrastrutures build on blockchains may enable decentralized sharing of data (e.g., Sockin and Xiong, 2022). Ocean Protocol, a nonprofit platform developed by a Singapore-based foundation, is a salient example of data marketplaces in which companies consumers, and other parties share or trade data.

[58] See the Digital Markets Act by the European Commission.

[59] See here.