Online Appendix for "Too Much Data: Prices and Inefficiencies in Data Markets" by Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, Asu Ozdaglar

This Appendix includes the remaining examples, proofs, and the detail of the statements in the text, with the order they appear in the main text.

Example B-1. Consider a setting related to Example 2, but with just a single community consisting of n users with homogeneous values of privacy v and homogeneous correlations given by $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho$ for all $i, j \in \mathcal{V}$, $i \neq j$. As $n \to \infty$, if $v \leq \frac{1}{(1-\rho)^2}$, the equilibrium involves $\mathbf{a} = \mathbf{1}$ and if $v > \frac{1}{(1-\rho)^2}$, the equilibrium has $\mathbf{a} = \mathbf{0}$. Take the former case: $v \leq \frac{1}{(1-\rho)^2}$, and suppose $v \in (1, \frac{1}{(1-\rho)^2}]$. Then equilibrium surplus is negative. In particular, each user's utility is $-v\rho < 0$. However, the (per-user) utility of the platform is $\frac{1}{2-\rho} - v\frac{(1-\rho)^2}{2-\rho} > 0$.

Proof of Lemma 4

The payoff of user i from joining platform $k \in \{1, 2\}$ can be lower bounded by $c_i(J_k) + p_i^{J_k,\mathbf{E}} - v_i\mathcal{I}_i(\mathbf{a}^{J_k,\mathrm{E}}) \geq c_i(\{i\}) - v_i\sigma_i^2$, which is positive given Assumption 1. ■

Definition and Existence of Mixed Strategy Equilibria when Platforms Compete over Data Prices

Definition B-1. Let $\mathcal{P}$ be the set of probability measures over $\mathbb{R}_+^n$. For any user $i \in \mathcal{V}$, let $\mathcal{A}_i$ be the set of probability measures over $\{0, 1\}$ and $\mathcal{B}_i$ be the set of probability measures over $\{1, 2\}$. For given price vectors $\mathbf{p}^1 \in \mathbb{R}^n$ and $\mathbf{p}^2 \in \mathbb{R}^n$, the joining and sharing profiles $\boldsymbol{\beta} \in \prod_{i\in\mathcal{V}} \mathcal{B}_i$ and $\boldsymbol{\alpha} \in \prod_{i\in\mathcal{V}} \mathcal{A}_i$ constitute a mixed user equilibrium if for any $i \in \mathcal{V}$, we have

$$u_i(\alpha_i, \beta_i, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i}, \mathbf{p}^1, \mathbf{p}^2) \geq u_i(\alpha_i', \beta_i', \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i}, \mathbf{p}^1, \mathbf{p}^2), \quad \text{for all } \alpha_i' \in \mathcal{A}_i \text{ and } \beta_i' \in \mathcal{B}_i,$$

where $u_i(\alpha_i, \beta_i, \boldsymbol{\alpha}_{-i}, \boldsymbol{\beta}_{-i}, \mathbf{p}^1, \mathbf{p}^2) = \mathbb{E}_{a_i\sim\alpha_i, \mathbf{a}_{-i}\sim\boldsymbol{\alpha}_{-i}, b_i\sim\beta_i, \mathbf{b}_{-i}\sim\boldsymbol{\beta}_{-i}}\left[u_i(a_i, b_i, \mathbf{a}_{-i}, \mathbf{b}_{-i}, \mathbf{p}^1, \mathbf{p}^2)\right]$. We denote by $\mathcal{A}(\boldsymbol{\pi}^1, \boldsymbol{\pi}^2)$ the set of mixed strategy user equilibria for given price strategies $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$. Strategy price profiles $\boldsymbol{\pi}^{k,\mathrm{E}} \in \mathcal{P}$, $k = 1, 2$, joining profile $\boldsymbol{\beta}^{\mathrm{E}}$, and sharing profile $\boldsymbol{\alpha}^{\mathrm{E}}$ constitute a mixed strategy equilibrium if $(\boldsymbol{\alpha}^{\mathrm{E}}, \boldsymbol{\beta}^{\mathrm{E}}) \in \mathcal{A}(\boldsymbol{\pi}^{1,\mathrm{E}}, \boldsymbol{\pi}^{2,\mathrm{E}})$ and for any $\boldsymbol{\pi} \in \mathcal{P}$ there exists $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in \mathcal{A}(\boldsymbol{\pi}, \boldsymbol{\pi}^{2,\mathrm{E}})$ such that

$$\mathbb{E}_{\mathbf{p}^1\sim\boldsymbol{\pi}^{1,\mathrm{E}}, \mathbf{p}^2\sim\boldsymbol{\pi}^{2,\mathrm{E}}, \mathbf{a}\sim\boldsymbol{\alpha}^{\mathrm{E}}, \mathbf{b}\sim\boldsymbol{\beta}^{\mathrm{E}}}\left[U^{(1)}(\mathbf{p}^1, \mathbf{p}^2, \mathbf{a}, \mathbf{b})\right] \geq \mathbb{E}_{\mathbf{p}^1\sim\boldsymbol{\pi}, \mathbf{p}^2\sim\boldsymbol{\pi}^{2,\mathrm{E}}, \mathbf{a}\sim\boldsymbol{\alpha}, \mathbf{b}\sim\boldsymbol{\beta}}\left[U^{(1)}(\mathbf{p}^1, \mathbf{p}^2, \mathbf{a}, \mathbf{b})\right],$$

and there exists $(\boldsymbol{\alpha}', \boldsymbol{\beta}') \in \mathcal{A}(\boldsymbol{\pi}^{1,\mathrm{E}}, \boldsymbol{\pi})$ such that

$$\mathbb{E}_{\mathbf{p}^1\sim\boldsymbol{\pi}^{1,\mathrm{E}}, \mathbf{p}^2\sim\boldsymbol{\pi}^{2,\mathrm{E}}, \mathbf{a}\sim\boldsymbol{\alpha}^{\mathrm{E}}, \mathbf{b}\sim\boldsymbol{\beta}^{\mathrm{E}}}\left[U^{(2)}(\mathbf{p}^1, \mathbf{p}^2, \mathbf{a}, \mathbf{b})\right] \geq \mathbb{E}_{\mathbf{p}^1\sim\boldsymbol{\pi}^{1,\mathrm{E}}, \mathbf{p}^2\sim\boldsymbol{\pi}, \mathbf{a}\sim\boldsymbol{\alpha}', \mathbf{b}\sim\boldsymbol{\beta}'}\left[U^{(2)}(\mathbf{p}^1, \mathbf{p}^2, \mathbf{a}, \mathbf{b})\right].$$

Our main existence result in the case of competition over data prices is:

Theorem B-1. *There exists a mixed strategy equilibrium strategy.*

This theorem follows because for any price vector $\mathbf{p}^1$ and $\mathbf{p}^2$, the second-stage game is a finite game and therefore has a mixed strategy equilibrium. If there are multiple equilibria, we select the one with the highest sum of platform's utilities. We next establish the existence of the mixed strategy equilibirium in the first-stage game by using Dasgupta-Maskin Theorem Dasgupta et al. [1986]. In particular, we show that the conditions of this theorem are satisfied, establishing a mixed strategy equilibrium exists. First, note that the price each platform offers to any user cannot exceed the highest overall leaked information, i.e., $\sum_{i\in\mathcal{V}} \mathcal{I}_i(\mathcal{V})$. Therefore, without loss of generality, we assume the action space of both platforms is $[0, \sum_{i\in\mathcal{V}} \mathcal{I}_i(\mathcal{V})]^n$.

For two vector of prices $\mathbf{p}^1$ and $\mathbf{p}^2$ and user $i \in \mathcal{V}$ we define functions $f_{12} : \mathbf{p}^1 \to \mathbf{p}^2$ such that

$$[f_{12}(\mathbf{p}^1)]_i = p_i^1 - v_i\mathcal{I}_i(S_1) + c_i(J_1) + v_i\mathcal{I}_i(S_2) - c_i(J_2), \forall S_1, S_2, J_1, J_2 \subseteq \mathcal{V}.$$

Note that there are finitely many such functions and in particular at most $n^{2^n\times 2^n\times 2^n}$ of them (this is because there are n components and for each of them J_1 has 2^n possibilities, $J_2 = \mathcal{V}\backslash J_1$, and each of S_1 and S_2 have 2^n possibilities). Also, note that the functions f_{12} are all linear and hence bijective and continuous. By changing the prices $\mathbf{p}^1$ and $\mathbf{p}^2$, as long as user equilibria of the second-stage game are the same, the payoff functions remain continuous. It becomes discontinuous when a user $i \in \mathcal{V}$ who is sharing on platform 1 changes her decision and starts sharing on platform 2. For this to happen we must have $p_i^1 - v_i\mathcal{I}_i(S_1) + c_i(J_1) = p_i^2 - v_i\mathcal{I}_i(S_2) + c_i(J_2)$, where S_1 is the set of users who are sharing on platform 1, S_2 is the set of users who are sharing on platform 2, and J_1 and J_2 are the sets of users who are joining platforms 1 and 2, respectively. Therefore, for any discontinuity point of $U^1(\mathbf{p}^1, \mathbf{p}^2)$, there exists f_{12} such that $\mathbf{p}^2 = f_{12}(\mathbf{p}^1)$. This establishes that the first condition of Dasgupta-Maskin Theorem holds.

The second condition of Dasgupta-Maskin Theorem holds because as long as user equilibria of the second-stage game remains the same, payoff functions are continuous in the first stage prices. When user equilibria changes, we select the one with the highest sum of the platforms' utilities. This implies that the sum of platforms' utilities is an upper semicontinuous function in prices.

The third condition of Dasgupta-Maskin Theorem holds because by changing $\mathbf{p}^1$, as long as the equilibrium of the second-stage game has not changed, the payoff of platform 1 is continuous. At the point that the equilibrium changes, we have multiplicity of equilibria and we have chosen the one that gives maximum payoff of platforms. Therefore, we have $\liminf_{\mathbf{p}'^1 \downarrow \mathbf{p}^1} U^1(\mathbf{p}'^1, \mathbf{p}^2) = U^1(\mathbf{p}^1, \mathbf{p}^2)$, which by definition is weakly lower semicontinuous with the choice of $\lambda = 0$. ■

Proof of Theorem 4

Part 1: In this case, there is no externality among users, and both the first best and the equilibrium involve all users joining platform 1 (or platform 2) and all low-value users sharing their data. In particular, we show that the following prices with a user equilibrium in which all users join platform 1 and all low-value users share on platform 1 is an equilibrium. For all $i \in \mathcal{V}^{(l)}$ we let

$$p_i^{1,\mathrm{E}} = \begin{cases} v_i \mathcal{I}_i(\{i\}), & c_i(\mathcal{V}) - c_i(\{i\}) \geq (1-v_i)\mathcal{I}_i(\{i\}), \\ \mathcal{I}_i(\{i\}) - (c_i(\mathcal{V}) - c_i(\{i\})), & c_i(\mathcal{V}) - c_i(\{i\}) < (1-v_i)\mathcal{I}_i(\{i\}), \end{cases}$$

and

$$p_i^{2,\mathrm{E}} = \begin{cases} v_i \mathcal{I}_i(\{i\}), & c_i(\mathcal{V}) - c_i(\{i\}) \geq (1-v_i)\mathcal{I}_i(\{i\}), \\ \mathcal{I}_i(\{i\}), & c_i(\mathcal{V}) - c_i(\{i\}) < (1-v_i)\mathcal{I}_i(\{i\}). \end{cases}$$

This is a user equilibrium because the payoff of a user on platform 1 is $c_i(\mathcal{V})$ that is larger than her payoff on platform 2 which is $c_i(\{i\})$. We next show that platform 1 does not have a profitable deviation. For any user i for which $c_i(\mathcal{V}) - c_i(\{i\}) \geq (1-v_i)\mathcal{I}_i(\{i\})$ platform 1 cannot increase its payoff by reducing its price offer because the user would then stop sharing her data. For any user with $c_i(\mathcal{V}) - c_i(\{i\}) < (1-v_i)\mathcal{I}_i(\{i\})$, a lower price would make the user join platform 2. This establishes that the first platform does not have a profitable deviation. We next show that platform 2 does not have a profitable deviation. The maximum price that platform 2 can offer to user i without making negative profits is $\mathcal{I}_i(\{i\})$ (this is because there exists no externality). Such a price offer is not sufficient to attract users from platform 1. In particular, if $c_i(\mathcal{V}) - c_i(\{i\}) < (1-v_i)\mathcal{I}_i(\{i\})$ then the price $\mathcal{I}_i(\{i\})$ offered to user i by the second platform would make her indifferent between the two platforms and if $c_i(\mathcal{V}) - c_i(\{i\}) \geq (1-v_i)\mathcal{I}_i(\{i\})$ then the price $\mathcal{I}_i(\{i\})$ offered to user i by the second platform would give her a lower payoff. Therefore, the second platform does not have a profitable deviation. This completes the proof of the first part.

Part 2: Note that the first best is to have all users join the same platform and low-value users share on it, which we assume is platform 1. Since there is no externality from high-value users, without loss of generality, we show the proof when they are removed from the market as they will join (and not share) the platform that has a higher joining value for them.

Part 2-1: We show that if $\delta \geq \underline{\delta}$ for

$$\underline{\delta} = \max_{i \in \mathcal{V}} \left(\sum_{j \in \mathcal{V}} \mathcal{I}_j(\mathcal{V}) \right) + v_i(\mathcal{I}_i(\mathcal{V} \setminus \{i\}) - \mathcal{I}_i(\{i\})), \tag{B-1}$$

the first best is an equilibrium, supported by the following prices:

$$p_i^{1,\mathrm{E}} = v_i(\mathcal{I}_i(\mathcal{V}) - \mathcal{I}_i(\mathcal{V} \setminus \{i\})), \quad p_i^{2,\mathrm{E}} = c_i(\mathcal{V}) - c_i(\{i\}) - v_i \mathcal{I}_i(\mathcal{V} \setminus \{i\}) + v_i \mathcal{I}_i(\{i\}), \quad i \in \mathcal{V}. \tag{B-2}$$

Note that all (low-value) users joining and sharing on platform 1 is a user equilibrium. This is because the payoff of each user i is she deviates and shares on the other platform is equal to her current payoff. Platform 1 does not have a profitable deviation. This is because, using Theorem 2, platform 1 is paying the minimum prices to get the data of all users. We next show that platform 2 does not have a profitable deviation, establishing the prices specified above and all users sharing on platform 1 is an equilibrium. The highest price that platform 2 can offer to get one of users, e.g., user i, share on it is bounded by the total information leakage $\sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V})$. We show that given the condition on δ all users joining and sharing on platform 1 is always a user equilibrium which gives the second platform zero payoff. Consider user $i\in\mathcal{V}$, given all other users are sharing on platform 1 it is best response for this user to share on platform 1 because $p_i^{1,\mathrm{E}}-v_i\mathcal{I}_i(\mathcal{V})+c_i(\mathcal{V})\overset{(a)}{=}-v_i\mathcal{I}_i(\mathcal{V}\setminus\{i\})+c_i(\mathcal{V})\overset{(b)}{\geq}-v_i\mathcal{I}_i(\mathcal{V}\setminus\{i\})+\delta+c_i(\{i\})\overset{(c)}{\geq}\sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V})-v_i\mathcal{I}_i(\{i\})+c_i(\{i\})\geq\tilde{p}_i^2-v_i\mathcal{I}_i(\{i\})+c_i(\{i\})$, where (a) follows from plugging in the prices given in (B-2), (b) follows from the definition of δ, and (c) follows from (B-1).

Part 2-2: We show that given $\Delta\leq\bar{\Delta}$ and $v_i\leq\bar{v}$, where

$$\bar{\Delta}=\max\Bigg\{\max_{i,S:\,i\notin S}(1-v_i)\mathcal{I}_i(\mathcal{V})+v_i(2\mathcal{I}_i(\{i\})-\mathcal{I}_i(S^c)),\\ \max_{S\subseteq\mathcal{V}}\frac{1}{2|S|}\sum_{i\in S}\mathcal{I}_i(\mathcal{V})-(1-v_i)\mathcal{I}_i(S)-v_i\mathcal{I}_i(S^c\cup\{i\})\Bigg\}. \tag{B-3}$$

and

$$\bar{v}=\min\left\{\frac{1}{2},\min_{i,S:\,i\in S,\mathcal{I}_i(\mathcal{V})-\mathcal{I}_i(S)\neq 0}\frac{\mathcal{I}_i(\mathcal{V})-\mathcal{I}_i(S)}{\mathcal{I}_i(\mathcal{V})-\mathcal{I}_i(S)+\mathcal{I}_i(S^c\cup\{i\})-\mathcal{I}_i(\{i\})}\right\}. \tag{B-4}$$

the first best is an equilibrium.

Before we proceed with the proof of this case, note that both $\bar{\Delta}$ and $\bar{v}$ are non-zero. The latter is by hypothesis. Consider the former, (B-3). The first term on the right-hand side of this expression is non-zero if $v_i\leq\frac{1}{2}$. The second term on the right-hand side of this expression is non-zero, because $v_i\leq 1$ and leaked information is monotonically increasing in the set of users who share. If $\mathcal{I}_i(\mathcal{V})=\mathcal{I}_i(S)$ for some S, then user i is uncorrelated with users and the complement of S, S^c. If $\mathcal{I}_i(\mathcal{V})=\mathcal{I}_i(S^c)$ for some S, then user i is uncorrelated with users in S. Therefore, for the right-hand side of the above expression to be zero all users must be uncorrelated with all other users. But this contradicts the assumption that at least the data of two low-value users are correlated.

We next show that the following prices form an equilibrium:

$$p_i^{1,\mathrm{E}}=\mathcal{I}_i(\mathcal{V})-(c_i(\mathcal{V})-c_i(\{i\})),\quad p_i^{2,\mathrm{E}}=(1-v_i)\mathcal{I}_i(\mathcal{V})+\mathcal{I}_i(\{i\}),\quad i\in\mathcal{V}. \tag{B-5}$$

Note that with these prices all users sharing on platform 1 is a user equilibrium. This is because if a user deviates and shares on the second platform, she receives the same payoff. We next show that the second platform does not have a profitable deviation. Suppose the second platform deviates to get users in a set S share on it by offering prices $\tilde{p}_i^2$. We show that the payoff of platform 2 becomes strictly negative. Consider one of the user equilibria after this deviation and suppose that the set $J_1\subseteq S^c$ of users join platform 1 and a subset of J_1, S_1, shares on platform 1. Users in S must prefer to share on platform 2, which leads to $\tilde{p}_i^2-v_i\mathcal{I}_i(S)+c_i(S)\geq p_1^i-v_i\mathcal{I}_i(S_1\cup\{i\})+c_i(J_1\cup\{i\})\geq$

$\mathcal{I}_i(\mathcal{V}) - (c_i(\mathcal{V}) - c_i(\{i\})) - v_i\mathcal{I}_i(S^c \cup \{i\}) + c_i(\{i\})$. Hence

$$\begin{aligned}\tilde{p}_i^2 &\geq v_i\mathcal{I}_i(S) - c_i(S) + \mathcal{I}_i(\mathcal{V}) - (c_i(\mathcal{V}) - c_i(\{i\})) - v_i\mathcal{I}_i(S^c \cup \{i\}) + c_i(\{i\}) \\ &\geq -2\Delta + \mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(S) - v_i\mathcal{I}_i(S^c \cup \{i\}). \end{aligned} \tag{B-6}$$

Therefore, the payoff of platform 2 becomes $\sum_{i \in S} \mathcal{I}_i(S) - \tilde{p}_i^2 \overset{(a)}{\leq} \sum_{i \in S} \mathcal{I}_i(S) + 2\Delta - \mathcal{I}_i(\mathcal{V}) - v_i\mathcal{I}_i(S) + v_i\mathcal{I}_i(S^c \cup \{i\}) \overset{(b)}{<} 0$, where (a) follows by using (B-6) and (b) follows from the choice of $\bar{\Delta}$ in (B-3).

We next show that platform 1 does not have a profitable deviation. Suppose that platform 1 deviates to get users in the set S to share on it with prices $\tilde{p}_i^1$. We first claim that if $\Delta \leq \bar{\Delta}$ (where $\bar{\Delta}$ is given in (B-3)), in one of the user equilibria all users prefer to share on platform 2. In particular, for a user $i \in S^c$, her payoff if she shares on platform 2 is higher than her payoff if she joins platform 2 and does not share because $p_i^2 - v_i\mathcal{I}_i(S^c) + c_i(S^c) \overset{(a)}{=} (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(\{i\}) - v_i\mathcal{I}_i(S^c) + c_i(S^c) \overset{(b)}{\geq} -v_i\mathcal{I}_i(S^c \setminus \{i\}) + c_i(S^c)$, where (a) follows by using the prices given in (B-5) and (b) follows the submodularity of leaked information and in particular $v_i\mathcal{I}_i(\{i\}) \geq v_i(\mathcal{I}_i(S^c) - \mathcal{I}_i(S^c \setminus \{i\}))$. Also, for a user $i \in S^c$, her payoff if she shares on platform 2 is higher than her payoff if she joins platform 1 because (she does not share on platform 1 as the price offered to her is 0)

$$\begin{aligned} p_i^2 - v_i\mathcal{I}_i(S^c) + c_i(S^c) &\overset{(a)}{=} (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(\{i\}) - v_i\mathcal{I}_i(S^c) + c_i(S^c) \\ &\overset{(b)}{\geq} -v_i\mathcal{I}_i(\{i\}) + c_i(S \cup \{i\}) \geq -v_i\mathcal{I}_i(S \cup \{i\}) + c_i(S \cup \{i\}) \end{aligned}$$

where (a) follows by using the prices given in (B-5) and (b) follows from the definition of Δ and the choice of $\bar{\Delta}$ given in (B-3). To have users in the set S share on platform 1 the new prices must satisfy $\tilde{p}_i^1 - v_i\mathcal{I}_i(S) + c_i(S) \geq p_i^2 - v_i\mathcal{I}_i(S^c \cup \{i\}) + c_i(S^c \cup \{i\}) \overset{(a)}{=} (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(\{i\}) - v_i\mathcal{I}_i(S^c \cup \{i\}) + c_i(S^c \cup \{i\})$, where (a) follows from substituting for prices from (B-5). Therefore,

$$\begin{aligned} \tilde{p}_i^1 &\geq (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(\{i\}) - v_i\mathcal{I}_i(S^c \cup \{i\}) + v_i\mathcal{I}_i(S) + c_i(S^c \cup \{i\}) - c_i(S) \\ &\geq (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(\{i\}) - v_i\mathcal{I}_i(S^c \cup \{i\}) + v_i\mathcal{I}_i(S) - \Delta. \end{aligned} \tag{B-7}$$

We next show that this user equilibrium generates no higher payoff for platform 1 (compared to its equilibrium payoff of $\Delta|\mathcal{V}|$). In particular, the payoff of platform 1 can be written as $\sum_{i \in S} \mathcal{I}_i(S) - \tilde{p}_i^1 \overset{(a)}{\leq} \sum_{i \in S} (1 - v_i)\mathcal{I}_i(S) - (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(S^c \cup \{i\}) - v_i\mathcal{I}_i(\{i\}) + \Delta \overset{(b)}{\leq} \Delta|S| \leq \Delta|\mathcal{V}|$ where in (a) we used the inequality (B-7), and (b) follows from the choice of $\bar{v}$ given in (B-4). In particular, using $v_i \leq \bar{v}$, for any i and S such that $\mathcal{I}_i(\mathcal{V}) - \mathcal{I}_i(S) \neq 0$, we have $(1 - v_i)\mathcal{I}_i(S) - (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(S^c \cup \{i\}) - v_i\mathcal{I}_i(\{i\}) \leq 0$. For any S and $i \in S$ for which $\mathcal{I}_i(\mathcal{V}) - \mathcal{I}_i(S) = 0$, user i's data must be independent of the data of users in S^c. Therefore, we have $\mathcal{I}_i(S^c \cup \{i\}) = \mathcal{I}_i(\{i\})$. This leads to $(1 - v_i)\mathcal{I}_i(S) - (1 - v_i)\mathcal{I}_i(\mathcal{V}) + v_i\mathcal{I}_i(S^c \cup \{i\}) - v_i\mathcal{I}_i(\{i\}) = 0$ that completes the proof of this case.

Part 2-3: We show that if $\Delta \leq \tilde{\Delta}$ where

$$\tilde{\Delta} = \min\left\{\min_{i\in\mathcal{V}} \frac{1}{2}(1-v_i)\mathcal{I}_i(\{i\}), \max_{i\in\mathcal{V}} \frac{\sum_{i\neq j}\mathcal{I}_i(\{i,j\}) - \mathcal{I}_i(\{i\})}{2(3|\mathcal{V}|-2)}\right\}, \tag{B-8}$$

then there exist $\tilde{\mathbf{v}}$ such that for $\mathbf{v}^{(l)} \geq \tilde{\mathbf{v}}$ the equilibrium is inefficient.

Before, we proceed with the proof, note that the second argument of maximum is non-zero since there exits at least two low-value users whose data are correlated.

We show that there exists no prices for both platforms to sustain all (low-value) users share on platform 1 as an equilibrium. We suppose the contrary and then reach a contradiction. In particular, we let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be the equilibrium prices offered by platform 1 and 2. Since all users sharing on platform 1 is a user equilibrium we must have $x_i - v_i\mathcal{I}_i(\mathcal{V}) + c_i(\mathcal{V}) \geq y_i - v_i\mathcal{I}_i(\{i\}) + c_i(\{i\})$. Also, note that we must have $x_i - v_i\mathcal{I}_i(\mathcal{V}) \leq c_i(\mathcal{V}) - c_i(\{i\}) + y_i - v_i\mathcal{I}_i(\{i\})$. This is because, otherwise platform 1 can deviate by decreasing its prices and increase its payoff. Therefore, we have

$$x_i - v_i\mathcal{I}_i(\mathcal{V}) \in \left[-(c_i(\mathcal{V}) - c_i(\{i\})) + y_i - \mathcal{I}_i(\{i\}), (c_i(\mathcal{V}) - c_i(\{i\})) + y_i - \mathcal{I}_i(\{i\})\right], \quad i \in \mathcal{V}. \tag{B-9}$$

Moreover, we also have

$$x_i - v_i\mathcal{I}_i(\mathcal{V}) + c_i(\mathcal{V}) \geq (1-v_i)\mathcal{I}_i(\{i\}) + c_i(\{i\}), \quad i \in \mathcal{V}. \tag{B-10}$$

This is because if this inequality does not hold for some $i \in \mathcal{V}$, i.e., $\epsilon = ((1-v_i)\mathcal{I}_i(\{i\}) + c_i(\{i\})) - (x_i - v_i\mathcal{I}_i(\mathcal{V}) + c_i(\mathcal{V})) > 0$, then platform 2 will have a profitable deviation by letting $y_i' = \mathcal{I}_i(\{i\}) - \frac{\epsilon}{2}$ for all $i \in \mathcal{V}$. This is because in any user equilibria after this deviation at least one user shares on this platform, guaranteeing a positive profit.

We now consider the deviation of platform 1. For any $j \in \mathcal{V}$, the following prices guarantee that the only user equilibria is to have all users i in $\mathcal{V} \setminus \{j\}$ to share on platform 1 and user j share on platform 2: $x_i' = c_i(\mathcal{V}) - c_i(\{i\}) + y_i - v_i\mathcal{I}_i(\{i,j\}) + v_i\mathcal{I}_i(\mathcal{V} \setminus \{j\}) + \Delta, \quad i \in \mathcal{V} \setminus \{j\}$. This is a user equilibrium because for any user $i \in \mathcal{V} \setminus \{j\}$ the price offered to her with the maximum leaked information on platform 1 and minimum joining value is larger than her payoff with the price offered on platform 2 with minimum leaked information and maximum joining value, i.e., we have $x_i' - v_i\mathcal{I}_i(\mathcal{V} \setminus \{j\}) + c_i(\{i\}) > y_i - v_i\mathcal{I}_i(\{i,j\}) + c_i(\mathcal{V})$. Also, user j shares on platform 2 because we have

$$\begin{aligned}
&y_j - v_j\mathcal{I}_j(\{j\}) + c_j(\{j\}) \overset{(a)}{\geq} x_j - v_j\mathcal{I}_j(\mathcal{V}) - c_j(\mathcal{V}) + c_j(\{j\}) + c_j(\{j\}) \\
&\overset{(b)}{\geq} (1-v_j)\mathcal{I}_j(\{j\}) + 2(c_j(\{j\}) - c_j(\mathcal{V})) + c_j(\{j\}) \overset{(c)}{\geq} (1-v_j)\mathcal{I}_j(\{j\}) - 2\Delta + c_j(\{j\}) \overset{(d)}{\geq} c_j(\{j\})
\end{aligned}$$

where (a) follows from (B-9), (b) follows from (B-10), (c) follows from the definition of Δ, and (d) follows from condition (B-8).

This should not be a profitable deviation for platform 1, which leads to

$$\begin{aligned}\sum_{i\neq j}\mathcal{I}_i(\mathcal{V}\setminus\{j\})-x_i' &= \sum_{i\neq j}\mathcal{I}_i(\mathcal{V}\setminus\{j\})-c_i(\mathcal{V})+c_i(\{i\})-y_i+v_i\mathcal{I}_i(\{i,j\})-v_i\mathcal{I}_i(\mathcal{V}\setminus\{j\})\\ &\overset{(a)}{=}\sum_{i\neq j}\mathcal{I}_i(\mathcal{V}\setminus\{j\})-2c_i(\mathcal{V})+2c_i(\{i\})-x_i-\Delta+v_i\left(\mathcal{I}_i(\mathcal{V})-\mathcal{I}_i(\{i\})+\mathcal{I}_i(\{i,j\})-\mathcal{I}_i(\mathcal{V}\setminus\{j\})\right)\\ &\overset{(b)}{\leq}\sum_{i\in\mathcal{V}}\mathcal{I}_i(\mathcal{V})-x_i,\end{aligned}$$

where in (a) we used (B-9), and in (b) we used the fact that in the only user equilibrium after this deviation the payoff of platform 1 cannot increase. Rearranging the previous inequality and using the definition of Δ, for any $j\in\mathcal{V}$ we obtain

$$x_j\leq\mathcal{I}_j(\mathcal{V})+\sum_{i\neq j}(1-v_i)\left(\mathcal{I}_i(\mathcal{V})-\mathcal{I}_i(\mathcal{V}\setminus\{j\})\right)+v_i\left(\mathcal{I}_i(\{i\})-\mathcal{I}_i(\{i,j\})\right)+3\Delta,\quad j\in\mathcal{V}. \tag{B-11}$$

We now consider the deviation of platform 2. In particular, we show that for any $j\in\mathcal{V}$, with price $y_j'=x_j-v_j\mathcal{I}_j(\mathcal{V})+c_j(\mathcal{V})+v_j\mathcal{I}_j(\{j\})-c_j(\{j\})+\Delta$ and zero for all other users the only user equilibrium is to have user j share on platform 2 and all other users share on platform 1. First note that all other users will still share on platform 1 as their information leakage has weakly decreased (since j is not sharing on platform 1) and they receive the same payment. now consider user j. She shares her data on platform 2, because with the choice of the price y_j' we have $y_j'-v_j\mathcal{I}_j(\{j\})+c_j(\{j\})>x_j-v_j\mathcal{I}_j(\mathcal{V})+c_j(\mathcal{V})$. This cannot be a profitable deviation for platform 2 leading to $\mathcal{I}_j(\{j\})-y_j'=\mathcal{I}_j(\{j\})-x_j+v_j\mathcal{I}_j(\mathcal{V})-v_j\mathcal{I}_i(\{j\})\leq 0$. Therefore, we have

$$x_j\geq(1-v_j)\mathcal{I}_j(\{j\})+v_j\mathcal{I}_j(\mathcal{V})-\Delta. \tag{B-12}$$

We next show that for sufficiently large $\mathbf{v}^{(l)}$ given the choice of $\tilde{\Delta}$, the inequalities (B-12), and (B-11) cannot simultaneously hold. In particular, consider $j\in\mathcal{V}$ who is with at least one other low-value user. For boundary low-values, i.e., $v_i=1$, we must have $\Delta(3|\mathcal{V}|-2)+\mathcal{I}_j(\mathcal{V})+\sum_{i\neq j}\left(\mathcal{I}_i(\{i\})-\mathcal{I}_i(\{i,j\})\right)\geq\mathcal{I}_j(\mathcal{V})$, which does not hold provided that $\Delta<\frac{1}{3|\mathcal{V}|-2}\sum_{i\neq j}\mathcal{I}_i(\{i,j\})-\mathcal{I}_i(\{i\})$. Since $\Delta\leq\tilde{\Delta}$ for sufficiently large values $\mathbf{v}^{(l)}$, there exists no prices for which the first best can be sustained as an equilibrium, establishing inefficiency in this case.

Part 3-1: We first show that there exists $\bar{\mathbf{v}}$ and $\underline{\mathbf{v}}$ such that if $\mathbf{v}^{(h)}\geq\bar{\mathbf{v}}$, $\mathbf{v}^{(l)}\geq\underline{\mathbf{v}}$, and

$$\delta>\frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}^{(l)}}(1-v_i)(\mathcal{I}_i(\mathcal{V}^{(l)})-\mathcal{I}_i(\mathcal{V}_1^{(l)})), \tag{B-13}$$

the first best is to have all users joining the same platform, say platform 1, and only low-value users who are not correlated with high-value users share their data. Let $\mathcal{V}_1^{(l)}=\{i\in\mathcal{V}^{(l)}:\ \forall j\in\mathcal{V}^{(h)},\Sigma_{ij}=0\}$, be those low-value users uncorrelated with all high-value users and $\mathcal{V}_2^{(l)}=\mathcal{V}\setminus\mathcal{V}_2^{(l)}$ to denote the rest of the low-value users (i.e., low-value users correlated with at least one high-value user). The first best is to have all users join one of the platforms, say platform 1, because we can upper bound the social surplus when set J_k ($\neq \emptyset$ and $\neq \mathcal{V}$) joins platform $k \in \{1,2\}$ and set S_k share on it as

$$\begin{aligned}
&\sum_{i\in J_1}(1-v_i)\mathcal{I}_i(S_1)+c_i(J_1)+\sum_{i\in J_2}(1-v_i)\mathcal{I}_i(S_2)+c_i(J_2)\\
&\overset{(a)}{\leq} -|\mathcal{V}|\delta+\sum_{i\in\mathcal{V}}c_i(\mathcal{V})+\sum_{i\in J_1\cap\mathcal{V}^{(l)}}(1-v_i)\mathcal{I}_i(S_1\cap\mathcal{V}^{(l)})+\sum_{i\in J_2\cap\mathcal{V}^{(l)}}(1-v_i)\mathcal{I}_i(S_2\cap\mathcal{V}^{(l)})\\
&\leq -|\mathcal{V}|\delta+\sum_{i\in\mathcal{V}}c_i(\mathcal{V})+\sum_{i\in\mathcal{V}^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}^{(l)}) \overset{(b)}{\leq} \sum_{i\in\mathcal{V}}c_i(\mathcal{V})+\sum_{i\in\mathcal{V}_1^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_1^{(l)}),
\end{aligned}$$

where (a) follows from the definition of δ and the fact that for sufficiently large $\bar{v}$ only low-value users share and the information leakage of high-value users is zero and (b) follows from condition (B-13). We next consider two possible cases and show that in both of them platform 2 has a profitable deviation.

Case 1: There exists no non-zero correlation between users in $\mathcal{V}_1^{(l)}$ and users in $\mathcal{V}_2^{(l)}$. We show that in this case platform 2 can induce all users in $\mathcal{V}_2^{(l)}$ to join and share their data. In particular, the following prices form a profitable deviation for platform 2: $\tilde{p}_i^2 = \Delta + v_i\mathcal{I}_i(\mathcal{V}_2^{(l)}), \quad i \in \mathcal{V}_2^{(l)}$, and zero price to all other users. We first show that with these prices in any user equilibrium all users in $\mathcal{V}_2^{(l)}$ will join and share on platform 2. This is because the payoff of a user $i \in \mathcal{V}_2^{(l)}$ after deviating to share on platform 2 is $\tilde{p}_i^2 - v_i\mathcal{I}_i(S_2\cup\{i\})+c_i(J_2\cup\{i\}) \overset{(a)}{\geq} \tilde{p}_i^2 - v_i\mathcal{I}_i(\mathcal{V}_2^{(l)})+c_i(J_2\cup\{i\}) \overset{(b)}{=} \Delta + c_i(J_2\cup\{i\}) \overset{(c)}{\geq} c_i(J_1\cup\{i\})$ where (a) follows from the fact that the price offered by platform 2 to users outside of $\mathcal{V}_2^{(l)}$ is zero and hence they never share on platform 2, (b) follows from the choice of price offered by platform 2, and (c) follows from definition of Δ. We next show that if

$$\sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)}) > \Delta|\mathcal{V}_2^{(l)}|, \tag{B-14}$$

then the payoff of platform 2 after this deviation becomes positive. We can also lower bound the payoff of platform 2 by $\sum_{i\in\mathcal{V}_2^{(l)}}\mathcal{I}_i(\mathcal{V}_2^{(l)}) - \tilde{p}_i^2 = \sum_{i\in\mathcal{V}_2^{(l)}}\mathcal{I}_i(\mathcal{V}_2^{(l)}) - \Delta - v_i\mathcal{I}_i(\mathcal{V}_2^{(l)}) = \sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)}) - \Delta|\mathcal{V}_2^{(l)}| > 0$. Finally, we show that there exist $\bar{\delta}$ and $\bar{\Delta}$ such that for $\delta \geq \bar{\delta}$ and $\Delta \leq \bar{\Delta}$ both (B-13) and (B-14) hold. Since in this case users in $\mathcal{V}_1^{(l)}$ and those in $\mathcal{V}_2^{(l)}$ are uncorrelated, condition (B-13) becomes $\delta > \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)})$. Therefore, letting $\Delta < \bar{\Delta} = \frac{1}{|\mathcal{V}_2^{(l)}|}\sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)})$ and $\delta > \bar{\delta} = \frac{1}{|\mathcal{V}^{(l)}|}\sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)})$, completes the proof.

Case 2: There exists $j \in \mathcal{V}_2^{(l)}$ who is correlated with at least one other user in $\mathcal{V}_1^{(l)}$. We show that platform 2 has a profitable deviation to take user j join and share on it. In particular, the following prices constitute a profitable deviation for platform 2: $\tilde{p}_j^2 = c_j(\mathcal{V}) - c_j(\{j\}) + v_j\mathcal{I}_j(\{j\}) - v_j\mathcal{I}_j(\mathcal{V}_1^{(l)})$, and zero price to all other user. First note that in any user equilibrium user j shares on platform 2. This is because $\tilde{p}_j^2 - v_j\mathcal{I}_j(\{j\}) + c_j(J_2\cup\{j\}) = c_j(\mathcal{V}) - c_j(\{j\}) - v_j\mathcal{I}_j(\mathcal{V}_1^{(l)}) + \epsilon + c_j(J_2\cup\{j\}) \geq$

$c_j(\mathcal{V}) - v_j\mathcal{I}_j(\mathcal{V}_1^{(l)})$. This deviation is profitable for platform 2 provided that

$$(1 - v_j)\mathcal{I}_j(\{j\}) + v_j\mathcal{I}_j(\mathcal{V}_1^{(l)}) > \Delta. \tag{B-15}$$

Note that in this case $\mathcal{I}_j(\mathcal{V}_1^{(l)}) > 0$. Letting $\bar{\Delta} = \mathcal{I}_j(\mathcal{V}_1^{(l)}) \geq \Delta$ and $\bar{\delta} = \frac{1}{|\mathcal{V}^{(l)}|}\sum_{i\in\mathcal{V}_2^{(l)}}(1-v_i)\mathcal{I}_i(\mathcal{V}_2^{(l)}) < \delta$, guarantees that the first best is not an equilibrium. The proof is completed by observing that for $\mathbf{v}^{(l)}$ sufficiently close to 1, we have $\bar{\Delta} > \bar{\delta}$.

Part 3-2: We prove that for

$$\delta > \tilde{\delta} = \max_{i\in\mathcal{V}} v_i\mathcal{I}_i(\mathcal{V}) + \sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V}), \tag{B-16}$$

the first best is an equilibrium.

We first show that when δ is large, the first best involves all users joining the same platform, say platform 1. This is proved by showing that if users split and join different platforms social surplus decreases. Let $J_1 \neq \emptyset, \mathcal{V}$ be the set of users who join platform 1 and J_2 be the set of users who join platform 2. Also, let S_1 and S_2 be the sets of users who share on platforms 1 and 2, respectively. We can upper bound the surplus as

$$\left(\sum_{i\in J_1} c_i(J_1) + (1-v_i)\mathcal{I}_i(S_1)\right) + \left(\sum_{i\in J_2} c_i(J_2) + (1-v_i)\mathcal{I}_i(S_2)\right)$$
$$\overset{(a)}{\leq} \left(\sum_{i\in\mathcal{V}} c_i(\mathcal{V}) - \delta\right) + \sum_{i\in J_1}(1-v_i)\mathcal{I}_i(S_1) + \sum_{i\in J_2}(1-v_i)\mathcal{I}_i(S_2) \overset{(b)}{\leq} -\delta|\mathcal{V}| + \sum_{i\in\mathcal{V}} c_i(\mathcal{V}) + \mathcal{I}_i(\mathcal{V}) \overset{(c)}{<} \sum_{i\in\mathcal{V}} c_i(\mathcal{V}),$$

where (a) follows from the definition of δ and $J_1, J_2 \neq \emptyset$, (b) follows from replacing $(1-v_i)\mathcal{I}_i(S_k)$ for $k = 1, 2$ by its upper bound $\mathcal{I}_i(\mathcal{V})$, and (c) follows from the condition given in (B-16). We next show that the first best, $\mathbf{a}^{\mathrm{W}}$, can be supported as an equilibrium. In particular, the prices $p_i^1 = v_i(\mathcal{I}_i(\mathbf{a}^{\mathrm{W}}) - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}^{\mathrm{W}}))$ for all $i \in \mathcal{V}$ and $p_i^2 = 0$ for all $i \in \mathcal{V}$ makes $\mathbf{a}^{\mathrm{W}}$ an equilibrium. We next verify that at these prices $\mathbf{a}^{\mathrm{W}}$ is indeed a user equilibrium and than that none of the platforms have a profitable deviation. It is a user equilibrium because the payoff of any user such as user i on platform 1 is larger than or equal to $c_i(\mathcal{V}) - v_i\mathcal{I}_i(\mathcal{V})$. If user i deviates and joins platform 2, then her payoff is smaller than or equal to $c_i(\{i\}) + \sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V})$ (because the highest price offered to users on platform 2 is the total leaked information). User i does not have a profitable deviation because $c_i(\mathcal{V}) - v_i\mathcal{I}_i(\mathcal{V}) > c_i(\{i\}) + \delta - v_i\mathcal{I}_i(\mathcal{V}) \geq c_i(\{i\}) + \sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V})$. We next show that platforms do not have a profitable deviation. Next suppose that platform 1 deviates and offers price vector $\tilde{\mathbf{p}}^1$. Since $c_i(\mathcal{V}) - v_i\mathcal{I}_i(\mathcal{V}) > c_i(\{i\}) + \delta - v_i\mathcal{I}_i(\mathcal{V}) \geq c_i(\{i\}) + \sum_{j\in\mathcal{V}}\mathcal{I}_j(\mathcal{V})$, all users joining platform 2 is a user equilibrium of the price vectors $(\tilde{\mathbf{p}}^1, \mathbf{p}^2)$. This user equilibrium gives platform 1 zero payoff and therefore deviating and offering price vector $\tilde{\mathbf{p}}^1$ is not profitable for platform 1. The same argument also establishes that platform 2 does not have a profitable deviation. ■

Unknown Valuations

We first characterize the "second best" which takes into account that the value of privacy of each user is their private information, and then show that the second best coincides with the first best.

Proposition B-1. *Let* $\mathbf{v}$ *be the reported vector of values of privacy. Then the pricing scheme*

$$p_i(\mathbf{v}) = \left(\mathcal{I}_i(\mathbf{a}(\mathbf{v})) + \sum_{j\neq i}(1-v_j)\mathcal{I}_j(\mathbf{a}(\mathbf{v}))\right) - \min_{\mathbf{a}\in\{0,1\}^n}\left(\mathcal{I}_i(\mathbf{a}) + \sum_{j\neq i}(1-v_j)\mathcal{I}_j(\mathbf{a})\right),$$

where $\mathbf{a}(\mathbf{v}) = \mathit{argmax}_{\mathbf{a}\in\{0,1\}^n}\sum_{i\in\mathcal{V}}(1-v_i)\mathcal{I}_i(\mathbf{a})$ *incentivizes users to report their value of privacy truthfully, and thus the second best coincide with the first best.*

This mechanism is a variation of Vickery-Clarke-Grove mechanism Vickrey [1961], Clarke [1971], Groves [1973] (the proof of this proposition is similar to a typical VCG mechanism and therefore is. omitted). In particular, for any $i\in\mathcal{V}$ the price offered to user i is equal to the surplus of all other users on the platform when user i is present minus by the surplus when user i is absent. The second term in the price $p_i(\mathbf{v})$ can be any function of the values $\mathbf{v}_{-i}$, and the choice specified in Proposition B-1 guarantees that the prices are nonnegative.

We impose the following standard assumption on the (reversed) hazard rate and maintain it throughout this subsection without explicitly mentioning it.

Assumption 2. *For all* $i\in\mathcal{V}$*, the function* $\Phi_i(v) = v + \frac{F_i(v)}{f_i(v)}$ *is nondecreasing.*

Here $\Phi_i(v)$ is the well-known "virtual value" in incomplete information models, representing the additional rent that the agent will capture in incentive-compatible mechanisms. In our setting, it will enable users to obtain more of the surplus the platform infers from their data.

A sufficient condition for Assumption 2 to hold is for the reversed hazard rate $f_i(x)/F_i(x)$ to be nonincreasing. This requirement is satisfied for a variety of distributions such as uniform and exponential (see e.g., Burkschat and Torrado [2014]).

Theorem B-2. *For any reported vector of values* $\mathbf{v}$*, the equilibrium is given by*

$$\mathbf{a}^{\mathrm{E}}(\mathbf{v}) = \mathit{argmax}_{\mathbf{a}\in\{0,1\}^n}\sum_{i=1}^n(1-\Phi_i(v_i))\mathcal{I}_i(\mathbf{a}) + \Phi_i(v_i)\mathcal{I}_i(\mathbf{a}_{-i}, a_i=0),$$

and

$$p_i^{\mathrm{E}}(v_i) = \int_v^{v_{\max}}\left(\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}(x,\mathbf{v}_{-i})) - \mathcal{I}_i(\mathbf{a}^{\mathrm{E}}_{-i}(x,\mathbf{v}_{-i}), a_i=0)\right)dx + v_i\left(\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}(v_i,\mathbf{v}_{-i})) - \mathcal{I}_i(\mathbf{a}^{\mathrm{E}}_{-i}(v_i,\mathbf{v}_{-i}), a_i=0)\right).$$

Moreover, all users report truthfully and thus the expected payoff of the platform is

$$\mathbb{E}_{\mathbf{v}}\left[\max_{\mathbf{a}\in\{0,1\}^n}\sum_{i=1}^n(1-\Phi_i(v_i))\mathcal{I}_i(\mathbf{a}) + \Phi_i(v_i)\mathcal{I}_i(\mathbf{a}_{-i}, a_i=0)\right].$$

Proof of Theorem B-2: Using the revelation principle, we can focus on direct mechanisms and then find the optimal direct incentive compatible mechanism. For the given prices, users in the set $\mathbf{a}(\mathbf{v})$ share their data. Letting $A_i(v, \mathbf{v}_{-i}) = \mathcal{I}_i(\mathbf{a}(v, \mathbf{v}_{-i})) - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}(v, \mathbf{v}_{-i}))$, the incentive compatibility constraint can be written as

$$p_i(v, \mathbf{v}_{-i}) - vA_i(v, \mathbf{v}_{-i}) \geq p_i(v', \mathbf{v}_{-i}) - vA_i(v', \mathbf{v}_{-i}). \tag{B-17}$$

Writing the first order condition for inequality (B-17) yields $p_i'(v, \mathbf{v}_{-i}) = vA_i'(v, \mathbf{v}_{-i})$ for all v. Taking integral of both sides yields

$$p_i(v, \mathbf{v}_{-i}) = -\int_v^{v_{\max}} xA_i'(x, \mathbf{v}_{-i})dx \overset{(a)}{=} \int_v^{v_{\max}} A_i(x, \mathbf{v}_{-i})dx + vA_i(v, \mathbf{v}_{-i}), \tag{B-18}$$

where we used integration by part in (a). Taking expectation of both sides, results in the expected payment to user i equal to $\mathbb{E}_{\mathbf{v}}\left[p_i(\mathbf{v})\right] = \mathbb{E}_{\mathbf{v}}\left[\Phi_i(v_i)A_i(\mathbf{v})\right]$. Therefore, the expected payoff of the platform is $\mathbb{E}_{\mathbf{v}}\left[\sum_{i=1}^n \mathcal{I}_i(\mathbf{a}(\mathbf{v})) - \Phi_i(v_i)(\mathcal{I}_i(\mathbf{a}(\mathbf{v})) - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}(\mathbf{v})))\right]$. The equilibrium sharing profile maximizes $\sum_{i=1}^n (1 - \Phi_i(v_i))\mathcal{I}_i(\mathbf{a}(\mathbf{v})) + \Phi_i(v_i)\mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}(\mathbf{v}))$ for any reported vector $\mathbf{v}$. Finally, note that maximizing this expression yield the following property: if $\mathbf{a}(v, \mathbf{v}_{-i}) = 1$, then for all $v' \leq v$ we have $\mathbf{a}(v', \mathbf{v}_{-i}) = 1$. This follows from the fact that by Assumption 2 we have $\Phi_i(v') \leq \Phi_i(v)$ for all $v' \leq v$. Since the leaked information is monotone, we obtain that $A_i(v, \mathbf{v}_i)$ is decreasing in v because as we increase v, $\Phi_i(v)$ increases, which means $a_i(v, \mathbf{v}_{-i})$ decreases, which in turn means that $A_i(v, \mathbf{v}_i) = \mathcal{I}_i(\mathbf{a}(v, \mathbf{v}_{-i})) - \mathcal{I}_i(a_i = 0, \mathbf{a}_{-i}(v, \mathbf{v}_{-i}))$ decreases. This monotonicity property together with the payment identity (B-18) guarantees that the incentive compatibility constraint holds as we show next. Using the payment identity (B-18), the incentive compatibility constraint $p_i(v, \mathbf{v}_{-i}) - vA_i(v, \mathbf{v}_{-i}) \geq p_i(v', \mathbf{v}_{-i}) - vA_i(v', \mathbf{v}_{-i})$, is equivalent to $\left(\int_v^{v_{\max}} A_i(x, \mathbf{v}_{-i})dx\right) + vA_i(v, \mathbf{v}_{-i}) - vA_i(v, \mathbf{v}_{-i}) \geq \left(\int_{v'}^{v_{\max}} A_i(x, \mathbf{v}_{-i})dx\right) + v'A_i(v', \mathbf{v}_{-i}) - vA_i(v', \mathbf{v}_{-i})$. After canceling out the term $vA_i(v, \mathbf{v}_{-i})$ and rearranging the other terms, this inequality becomes equivalent to $\int_v^{v'} A_i(x, \mathbf{v}_{-i})dx \geq (v' - v)A_i(v', \mathbf{v}_{-i})$. To show this inequality we consider the following two possible cases:

- $v' \geq v$: we have $\int_v^{v'} A_i(x, \mathbf{v}_{-i})dx \overset{(a)}{\geq} \int_v^{v'} A_i(v', \mathbf{v}_{-i})dx = (v' - v)A_i(v', \mathbf{v}_{-i})$, where (a) follows from the fact that $A_i(x, \mathbf{v}_{-i})$ is decreasing in x and hence for all $x \in [v, v']$ we have $A_i(x, \mathbf{v}_{-i}) \geq A_i(v', \mathbf{v}_{-i})$.

- $v' < v$: we have $\int_v^{v'} A_i(x, \mathbf{v}_{-i})dx = \int_{v'}^{v} -A_i(x, \mathbf{v}_{-i})dx \overset{(a)}{\geq} \int_{v'}^{v} -A_i(v', \mathbf{v}_{-i})dx = (v' - v)A_i(v', \mathbf{v}_{-i})$, where (a) follows from the fact that $A_i(x, \mathbf{v}_{-i})$ is decreasing in x and hence for all $x \in [v', v]$ we have $-A_i(x, \mathbf{v}_{-i}) \geq -A_i(v', \mathbf{v}_{-i})$. ■

We next establish that the equilibrium is inefficient under fairly plausible conditions in this incomplete information setting as well. The main difference from our analysis so far is that another relevant set is the subset of low-value users with virtual value of privacy less than one, i.e.,

$\Phi_i(v_i) \le 1$. For the next theorem, we use the notation $\mathcal{V}_\Phi^{(l)} = \{i \in \mathcal{V} : \ \Phi_i(v_i) \le 1\}$ to denote this set of users.

Theorem B-3. 1. *Suppose high-value users are uncorrelated with all other users and $\mathcal{V}^{(l)} = \mathcal{V}_\Phi^{(l)}$. Then the equilibrium is efficient.*

2. *Suppose some high-value users (those in $\mathcal{V}^{(h)}$) are correlated with users in $\mathcal{V}_\Phi^{(l)}$. Then there exists $\bar{\mathbf{v}} \in \mathbb{R}^{|\mathcal{V}^{(h)}|}$ such that for $\mathbf{v}^{(h)} \ge \bar{\mathbf{v}}$ the equilibrium is inefficient.*

3. *Suppose every high-value user is uncorrelated with all users in $\mathcal{V}_\Phi^{(l)}$, but users in a nonempty subset $\hat{\mathcal{V}}^{(l)}$ of $\mathcal{V}^{(l)} \setminus \mathcal{V}_\Phi^{(l)}$ are correlated with at least one high-value user. Then there exist $\bar{\mathbf{v}}$ and $\tilde{v}$ such that if $\mathbf{v}^{(h)} \ge \bar{\mathbf{v}}$ and $v_i < \tilde{v}$ for some $i \in \hat{\mathcal{V}}^{(l)}$, the equilibrium is inefficient.*

4. *Suppose every high-value user is uncorrelated with all low-value users and at least one high-value user is correlated with another high-value user. Let $\tilde{\mathcal{V}}^{(h)} \subseteq \mathcal{V}^{(h)}$ be the subset of high-value users correlated with at least one other high-value user. Then for each $i \in \tilde{\mathcal{V}}^{(h)}$ there exists $\bar{v}_i > 0$ such that if for any $i \in \tilde{\mathcal{V}}^{(h)}$ $v_i < \bar{v}_i$, the equilibrium is inefficient.*

The inefficiency results in this theorem again have clear parallels to those in Theorem 3, but with some notable differences. First, efficiency now requires all low-value users to also have virtual valuations that are less than one, since otherwise user incentive compatibility constraints prevent the efficient allocation. Second, the conditions for inefficiency are slightly different depending on whether high-value users are correlated with low-value users whose virtual valuations are less than one or greater than one.

The proof of Theorem B-3 is omitted as it is similar to the proof of Theorem 3 and uses the analogous of Lemma 3, establishing that all users with $\Phi_i(v_i) \le 1$ share in equilibrium. ■

Finally, we present the analogue of Proposition 3 in this setting.

Proposition B-2. *Consider a setting with unknown valuations. For a given $\mathbf{v}$, we have*

$$\text{Social surplus}(\mathbf{a}^{\mathrm{E}}) \le \sum_{i\,:\,v_i \in \mathcal{V}^{(l)}} (1 - v_i)\mathcal{I}_i(\mathcal{V}) - \sum_{i\,:\,v_i \in \mathcal{V}^{(h)}} (v_i - 1)\mathcal{I}_i(\mathcal{V}_\Phi^{(l)}).$$

Proposition B-2 is similar to Proposition 3 and provides a sufficient condition for equilibrium surplus to be negative. The only difference is that the lower bound on the negative (second) term is evaluated for information leaked by users in $\mathcal{V}_\Phi^{(l)}$ (rather than those in $\mathcal{V}^{(l)}$). This is because, with incomplete information, the platform has to compensate users according to their virtual value of privacy and may find it too expensive to purchase the data of low-value users with $\Phi_i(v_i) > 1$.

Proof of Lemma 5

Without loss of generality, suppose $a_2 = \cdots = a_n = 1$. We have $\mathbb{E}[\mathbf{X}\tilde{\mathbf{S}}^T] = \mathbb{E}[\mathbf{X}\mathbf{S}^T]\Sigma^{-1} = I$. We also have $\mathbb{E}[\tilde{\mathbf{S}}\tilde{\mathbf{S}}^T] = \Sigma^{-1}\mathbb{E}[\mathbf{S}\mathbf{S}^T]\Sigma^{-1} = \Sigma^{-1}(I + \Sigma)\Sigma^{-1}$. We first find the leaked information of user 1 if she does not share. Since, the correlation between user 1's type and the shared data

$\tilde{S}_2, \ldots, \tilde{S}_n$ is zero, this leaked information is zero. We next find the leaked information of user 1 if she shares her information. Note that $\tilde{\mathbf{S}}$ and X_1 are jointly normal. Using the characterization of Theorem 2, this leaked information is equal to $(1, 0, \ldots, 0)\left(\Sigma^{-1}(I+\Sigma)\Sigma^{-1}\right)^{-1}(1, 0, \ldots, 0)^T = (1, 0, \ldots, 0)\Sigma(I+\Sigma)^{-1}\Sigma(1, 0, \ldots, 0)^T = (\sigma_1^2, \Sigma_{12}, \ldots, \Sigma_{1n})(I+\Sigma)^{-1}(\sigma_1^2, \Sigma_{12}, \ldots, \Sigma_{1n})^T = \mathcal{I}_1(a_1 = 1, \mathbf{a}_{-1})$, where the last equality follows from Theorem 2. This completes the proof of Lemma. ■

Proof of Theorem 6

Part 1. First note that the minimum price offered to user i to share her information with action profile $\mathbf{a}_{-i}$ must make her indifferent between her payoff if she shares which is given by $p_i - v_i\mathcal{I}_i(a_i = 1, \mathbf{a}_{-1})$ (where we used Lemma 5) and her payoff if she does not share which is zero. Therefore, the minimum price offered to users i to share is $v_i\mathcal{I}_i(a_i = 1, \mathbf{a}_{-i})$. Since the leaked information of users who do not share is zero, for a given action profile $\mathbf{a} \in \{0,1\}^n$, the payoff of the platform with minimum prices becomes $\sum_{i:\ a_i=1}\mathcal{I}_i(\mathbf{a}) - \sum_{i:\ a_i=1} p_i = \sum_{i:\ a_i=1}(1-v_i)\mathcal{I}_i(\mathbf{a})$. Choosing the action profile that maximizes this payoff, completes the proof.
Part 2. Suppose $\mathbf{a}^{\mathrm{E}}$ is the equilibrium action profile before de-correlation. We have $\text{Social surplus}(\mathbf{a}^{\mathrm{E}}) = \sum_{i\in\mathcal{V}}(1-v_i)\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}) = \sum_{i:\ a_i^{\mathrm{E}}=1}(1-v_i)\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}) + \sum_{i:\ a_i^{\mathrm{E}}=0}(1-v_i)\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}) \overset{(a)}{\leq} \sum_{i:\ a_i^{\mathrm{E}}=1}(1-v_i)\mathcal{I}_i(\mathbf{a}^{\mathrm{E}}) \overset{(b)}{\leq} \text{Social surplus}(\tilde{\mathbf{a}}^{\mathrm{E}})$, where (a) follows from the fact that all low-value users share in equilibrium (Lemma 3) and (b) follows because Part 1 shows that the action profile $\tilde{\mathbf{a}}^{\mathrm{E}}$ is the maximizer of $\sum_{i:\ a_i=1}(1-v_i)\mathcal{I}_i(\mathbf{a})$. Finally, note that the equilibrium social surplus after de-correlation cannot be negative because it is equal to $\sum_{i\in\mathcal{V}}(1-v_i)\tilde{\mathcal{I}}_i(\mathbf{a}^{\mathrm{E}}) \geq \sum_{i\in\mathcal{V}}(1-v_i)\tilde{\mathcal{I}}_i(\mathbf{a}=\mathbf{0}) = 0$. ■