Data Evaluation: Mechanisms and Implications

Online Supplement

Hong Zheng, School of Management, Beijing Institute of Technology

Lin Tian, School of Management, Fudan University

Guo Li, School of Management, Beijing Institute of Technology

---

Appendix A: Proofs of lemmas and propositions

In this appendix, we present all the proofs of our results in the main paper. To increase readability, Table A1 summarizes the key comparative statistics and the corresponding brief implications. Table A2 is a profit diagram illustrating the sellers' profits under different scenarios (i.e., under all scenarios combining free samples and data evaluation).

Table A1. Comparative statics and the corresponding implications

Comparative statics	Implications
$\frac{\partial T(\sigma,x)}{\partial x} > 0$	Data value increases with data selling volume
$\frac{\partial T(\sigma,x)}{\partial \sigma} > 0$	Data value increases with demand variance
$\frac{\partial T(\sigma,x)}{\partial \gamma} < 0$	Data value decreases with competition intensity
$\frac{\partial^2 T(\sigma,x)}{\partial x^2} < 0$	Marginal data value decreases with data selling volume
$\frac{\partial^2 T(\sigma,x)}{\partial x \partial \sigma} > 0$	Marginal data value increases with demand variance
$\frac{\partial^2 T(\sigma,x)}{\partial x \partial \gamma} < 0$	Marginal data value decreases with competition intensity

Table A2. Payoff diagram of sellers

	With data evaluation	Without data evaluation
With samples	$\Pi_{sH}^{F,sep}, \Pi_{sL}^{F,sep}$	$\Pi_{sH}^{F,pool}, \Pi_{sL}^{F,pool}$
Without samples	$\Pi_{sH}^{N,sep}, \Pi_{sL}^{N,sep}$	$\Pi_{sH}^{N,pool}, \Pi_{sL}^{N,pool}$

Note that the pooling equilibrium outcomes are the same in both scenarios—with and without free samples; specifically, $\Pi_{sH}^{F,pool} = \Pi_{sH}^{N,pool}$ and $\Pi_{sL}^{F,pool} = \Pi_{sL}^{N,pool}$.

Proof of Lemma 1

We derive the properties of $T(\sigma,x)$. From the first-order derivative of $T(\sigma,x)$ with respect to x, we know that $\frac{\partial T(\sigma,x)}{\partial x}=\frac{\sigma^2(2+(2-\gamma)\sigma)^2(4+4(1+x)\sigma+x(4+\gamma^2)\sigma^2)}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}>0$. From the first-order derivative of $T(\sigma,x)$ with respect to σ, we have $\frac{\partial T(\sigma,x)}{\partial \sigma}=\frac{x\sigma(2+(2-\gamma)\sigma)Q_1}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}$, where

$$Q_1=16+8(4+3x-2\gamma)\sigma+4\big(4+2x^2-2\gamma+x(12-7\gamma+\gamma^2)\big)\sigma^2$$
$$+2x(6+x(4-\gamma))(2-\gamma)\sigma^3+x^2(2-\gamma)^2(2+\gamma)\sigma^4.$$

It is easy to verify that each part of function Q_1 is positive; hence, $Q_1>0$. Therefore, we have proven that $\frac{\partial T(\sigma,x)}{\partial \sigma}>0$. Moreover, $\frac{\partial T(\sigma,x)}{\partial \gamma}=-\frac{2x\sigma^3(1+x\sigma)(2+(2-\gamma)\sigma)\big(4+(4+4x(1-\gamma))\sigma+x(2-\gamma)^2\sigma^2\big)}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}<0$ can be easily verified.

Next, we show how the marginal value of the data changes with respect to x, σ and γ. $\frac{\partial^2 T(\sigma,x)}{\partial x^2}=-\frac{2\sigma^3(2+(2-\gamma)\sigma)^2Q_2}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^4}$, where

$$Q_2=16+8(4+2x-\gamma^2)\sigma+8(2+4x-\gamma^2)\sigma^2+x(16-\gamma^4)\sigma^3>0.$$

Hence, we have $\frac{\partial^2 T(\sigma,x)}{\partial x^2}<0$.

Moreover, $\frac{\partial^2 T(\sigma,x)}{\partial x\partial \sigma}=\frac{4\sigma(2+(2-\gamma)\sigma)Q_3}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^4}$, where

$$Q_3=16+16(3+x-\gamma)\sigma+8(6-3\gamma+x(6-3\gamma+2\gamma^2))\sigma^2+8(2-\gamma+x^2(-1+\gamma)\gamma-x(-6+$$
$$4\gamma-3\gamma^2+\gamma^3))\sigma^3+x(-2+\gamma)(-4(2+\gamma^2)+x\gamma(4-2\gamma+\gamma^2))\sigma^4.$$

Because $\frac{\partial^2 Q_3}{\partial x^2}=-2\sigma^3(8(1-\gamma)\gamma+(2-\gamma)\gamma(4-2\gamma+\gamma^2)\sigma)<0$, $\frac{\partial Q_3}{\partial x}$ decreases in x. The minimum of $\frac{\partial Q_3}{\partial x}$ is given by $\frac{\partial Q_3}{\partial x}|_{x=1}=\sigma(16+8(6-3\gamma+2\gamma^2)\sigma+8(6-6\gamma+\gamma^2(5-\gamma))\sigma^2+2(2-\gamma)(4-4\gamma+\gamma^2(4-\gamma)\sigma^3)$, which is positive. Therefore, Q_3 is increasing in x. The minimum of Q_3 is given by $Q_3|_{x=0}=8(2+2(3-\gamma)\sigma+3(2-\gamma)\sigma^2+(2-\gamma)\sigma^3)>0$. Hence, $Q_3>0$; thus, we obtain $\frac{\partial^2 T(\sigma,x)}{\partial x\partial \sigma}>0$.

Note that $\frac{\partial^2 T(\sigma,x)}{\partial x\partial \gamma}=-\frac{2\sigma^3(2+(2-\gamma)\sigma)Q_4}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^4}$, where

$$Q_4=16+32\big(1+x(1-\gamma)\big)\sigma+16\big(1+x(2-\gamma)^2+x^2(1-2\gamma)\big)\sigma^2+4x(4(2-2\gamma+\gamma^2)+x(8-$$
$$16\gamma+4\gamma^2-\gamma^3))\sigma^3+x^2(16-32\gamma+16\gamma^2-4\gamma^3+\gamma^4)\sigma^4.$$

Because $\frac{\partial^2 Q_4}{\partial \gamma^2}=4x\sigma^2\big(8+\big(8+x(8-6\gamma)\big)\sigma+x(8-6\gamma+3\gamma^2)\sigma^2\big)>0$, $\frac{\partial Q_4}{\partial \gamma}$ is increasing in γ. Note that we consider the scenario with less market competition; i.e., $\gamma\le\frac{2}{3}$ (see the Proof of Proposition 1). The maximum of $\frac{\partial Q_4}{\partial \gamma}$ is given by $\frac{\partial Q_4}{\partial \gamma}|_{\gamma=\frac{2}{3}}=-\frac{16}{27}x\sigma(54+18(4+3x)\sigma+9(2+9x)\sigma^2+25x\sigma^3)<0$. Hence, Q_4 is decreasing in γ. The minimum of Q_4 is given by $Q_4|_{\gamma=\frac{2}{3}}=\frac{16}{81}(81+54(3+x)\sigma+$

$9(9+16x-3x^2)\sigma^2+6(15-4x)x\sigma^3+4x^2\sigma^4)>0$. Therefore, $Q_4>0$; hence, $\frac{\partial^2 T(\sigma,x)}{\partial x \partial \gamma}<0$. This concludes the proof. □

Proof of Proposition 1

In this proof, we derive the seller's optimal data selling decision with complete information. If the seller does not sell the data, the corresponding expected profit is given by $\Pi_s(\sigma,0)$. If the seller sells the volume x of data, the corresponding profit maximization problem is as follows:

$$x^* = \text{argmax}\, \Pi_s(\sigma,x)+T(\sigma,x)$$
$$s.t. \qquad 0 \le x \le 1.$$

Note that $\frac{\partial[\Pi_s(\sigma,x)+T(\sigma,x)]}{\partial x}=\frac{\sigma^2(2+(2-\gamma)\sigma_i)Q_5}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}$, where $Q_5=8+4(4+2x-3\gamma)\sigma+2(4-6\gamma+x(8-6\gamma+3\gamma^2))\sigma^2-x(-2+\gamma)^3\sigma^3$. Because $\frac{\partial Q_5}{\partial x}=\sigma(8+2(8-6\gamma+3\gamma^2)\sigma+(2-\gamma)^3\sigma^2)>0$, we know that Q_5 is increasing in x. The minimum of Q_5 is given by $Q_5|_{x=0}=4(1+\sigma)(2+(2-3\gamma)\sigma)>0$ for any $0<\gamma<\frac{2}{3}$. Therefore, $Q_5>0$; hence, $\Pi_s(\sigma,x)+T(\sigma,x)$ is increasing in x. For this reason, under the option of selling data, the optimal selling volume of data is $x=1$, and the corresponding price is given by $T(\sigma,1)$. Next, we compare the seller's optimal profit under selling, i.e., $\Pi_s(\sigma,1)+T(\sigma,1)$, and that under the option without selling, i.e., $\Pi_s(\sigma,0)$.

$$\Pi_s(\sigma,1)+T(\sigma,1)-\Pi_s(\sigma,0)=\frac{\sigma^2(4+4(2-\gamma)\sigma+(4-4\gamma-\gamma^2)\sigma^2)}{4(1+\sigma)(2+(2+\gamma)\sigma)^2}>0 \text{ for any } 0<\gamma<\frac{2}{3}.$$

Therefore, we know that the seller will sell all the data to the buyer (i.e., $x=1$) at a price $p_d=T(\sigma,1)$. This concludes the proof. □

Proof of Proposition 2

In this proof, we derive the seller's optimal selling decision with incomplete information under the separating equilibrium. Because both types of sellers sell all the data to the buyer in a complete information setting, the price of the high-type seller is higher than that of the low-type seller. Hence, in an incomplete information setting, the low-type seller has an incentive to mimic the decision of the high-type seller, and the high-type seller has an incentive to separate herself.

We first prove that, for any selling volume, the seller will charge the data selling price at the buyer's maximum willingness to pay, and hence, the sellers will be separated by the decision on selling volume x (i.e., the data selling price has no signaling role).

To derive the separating equilibrium, we need to ensure that both sellers will not deviate. Suppose that the separating equilibrium is supported by the belief that if $\{x,p_d\}\in\Phi$, the buyer believes that the market is of a high type; otherwise, the buyer believes that the market is of a low type. Therefore, a separating equilibrium exists if the following constraints are satisfied:

$$\begin{cases} \underbrace{max_{\{x,p_d\}\notin\Phi}\Pi_s(\sigma_L,x)+p_d}_{Low-type\ seller's\ first\ best\ profit} \geq \underbrace{max_{\{x,p_d\}\in\Phi}\Pi_s(\sigma_L,x)+p_d}_{Low-type\ seller's\ profit\ under\ deviating}, & (\text{C1}) \\ \underbrace{max_{\{x,p_d\}\in\Phi}\Pi_s(\sigma_H,x)+p_d}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{\{x,p_d\}\notin\Phi}\Pi_s(\sigma_H,x)+p_d}_{High-type\ seller's\ profit\ under\ deviating}. & (\text{C2}) \end{cases}$$

For the low-type seller, the maximization of his first best profit is organized by $\max_{\{x,p_d\}\notin\Phi}\Pi_s(\sigma_L,x)+p_d$, with the constraint $p_d \leq T(\sigma_L,x)$ for any data selling volume x. Otherwise, if the price is higher than the buyer's willingness to pay, the buyer will not purchase the data, leaving the low-type seller's profit as $\Pi_s(\sigma_L,x)$, which is less attractive. It is obvious that when maximizing the first best profit, the low-type seller will price the data at $T(\sigma_L,x)$ for any data selling volume. That is, $\max \Pi_s(\sigma_L,x)+p_d$ with respect to x and p_d is equivalent to $\max \Pi_s(\sigma_L,x)+T(\sigma_L,x)$ with respect to x. In a similar vein, we know that, for the low-type seller's profit under deviating, $\max \Pi_s(\sigma_L,x)+p_d$ with respect to x and p_d is equivalent to $\max \Pi_s(\sigma_L,x)+T(\sigma_H,x)$ with respect to x. For the high-type seller, by following the same procedure, we can conclude that for any data selling volume, the price will be set at the buyer's maximum willingness to pay. Consequently, the two-dimensional profit maximization problem reduces to a one-dimensional maximization with respect to the data selling volume. Therefore, in the least-cost separating equilibrium, the data selling price plays no signaling role; instead, sellers are separated solely by their choice of data selling volume.

From the above discussion, we know that the separating equilibrium characterized by conditions (C1) and (C2) is equivalent to that characterized by the following conditions $(\text{C1}')$ and $(\text{C2}')$:

$$\begin{cases} \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_L,x)+T(\sigma_L,x)}_{Low-type\ seller's\ first\ best\ profit} \geq \underbrace{max_{x\in\Phi}\Pi_s(\sigma_L,x)+T(\sigma_H,x)}_{Low-type\ seller's\ profit\ under\ deviating}, & (\text{C1}') \\ \underbrace{max_{x\in\Phi}\Pi_s(\sigma_H,x)+T(\sigma_H,x)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_H,x)+T(\sigma_L,x)}_{High-type\ seller's\ profit\ under\ deviating}. & (\text{C2}') \end{cases}$$

The corresponding supported belief is as follows: if $x \in \Phi$, the buyer believes that the market is of a high type; otherwise, the buyer believes that the market is of a low type. A similar belief structure has been widely considered in previous studies (i.e., Guo and Jiang 2016, Chen and Jiang 2021, Li et al. 2021). In a separating equilibrium, the low-type seller will select his first-best profit, selling all the data $x_L^{sep}=1$ at a price $T(\sigma_L,1)$. His first-best profit is $\Pi_s(\sigma_L,1)+T(\sigma_L,1)$. To ensure the condition $(\text{C1}')$, we know that $\Phi \equiv \{x: \Pi_s(\sigma_L,x)+T(\sigma_H,x) \leq \Pi_s(\sigma_L,1)+T(\sigma_L,1)\}$. It is easy to verify that $\frac{\partial[\Pi_s(\sigma_L,x)+T(\sigma_H,x)]}{\partial x} > \frac{\partial[\Pi_s(\sigma_L,x)+T(\sigma_L,x)]}{\partial x} > 0$. The first inequality holds because $\frac{\partial^2 T(\sigma_H,x)}{\partial x \partial \sigma_H} > 0$ (see Lemma 1). The second inequality applies to the results in Proposition 1. The minimum of $\Pi_s(\sigma_L,x)+T(\sigma_H,x)$ is given by $\Pi_s(\sigma_L,0)+T(\sigma_H,0)$, which is lower than the low-type seller's first-best profit. The maximum of $\Pi_s(\sigma_L,x)+T(\sigma_H,x)$ is given by $\Pi_s(\sigma_L,1)+T(\sigma_H,1)$, which is greater than the low-type seller's first-best profit. The reason is that $T(\sigma_H,1) > T(\sigma_L,1)$. Therefore, we know that there exists a threshold $x_H^{sep} \in$

(0,1) such that if $x \le x_H^{sep}$, we have $\Pi_s(\sigma_L, x) + T(\sigma_H, x) \le \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. In this sense, Φ actually captures the region of $x \le x_H^{sep}$. Next, we look at the condition (C2$'$). Because we consider the least-cost separating equilibrium for the high-type seller, the high-type seller will separate herself at x_H^{sep}. The reason is that $\Pi_s(\sigma_H, x) + T(\sigma_H, x)$ is increasing in x. To sustain the above separating equilibrium, we need to guarantee that the high-type seller has no incentive to deviate (i.e., selecting a selling volume with $x_H^{sep} < x < 1$). If the type of seller is σ_i but the buyer's belief is σ_j, the profit of the seller is given by $\Pi_s(\sigma_i, x) + T(\sigma_j, x)$. It is easy to verify that $\frac{\partial^2[\Pi_s(\sigma_i,x)+T(\sigma_j,x)]}{\partial\sigma_i\partial x} = -\frac{4\gamma\sigma_i^2 Q_6}{(4+4(1+x)\sigma_i+x(4-\gamma^2)\sigma_i^2)^4}$, where

$$Q_6 = 48 - 32(-4 + x(-2+\gamma) + \gamma)\sigma_i - 8(-14 + x^2(-2+\gamma) + 6\gamma + x(-20 + 17\gamma - 4\gamma^2))\sigma_i^2 - 4(-2+\gamma)(4 + x(16 - 10\gamma + \gamma^2) + x^2(4 - 2\gamma + \gamma^2))\sigma_i^3 + x(-2+\gamma)^2(8 - 2\gamma + x(4 - 2\gamma + \gamma^2))\sigma_i^4.$$

Because $\frac{\partial^2 Q_6}{\partial x^2} = 2(2-\gamma)\sigma_i^2\left(8 + 4(4 - 2\gamma + \gamma^2)\sigma_i + (2-\gamma)(4 - 2\gamma + \gamma^2)\sigma_i^2\right) > 0$, we know that $\frac{\partial Q_6}{\partial x}$ is increasing in x. The minimum of $\frac{\partial Q_6}{\partial x}$ is given by $\frac{\partial Q_6}{\partial x}|_{x=0} = \sigma_i\left(32(2-\gamma) + 8(20 - 17\gamma + 4\gamma^2)\sigma_i + 4(8-\gamma)(2-\gamma)^2\sigma_i^2 + 2(4-\gamma)(2-\gamma)^2\sigma_i^3\right) > 0$. Therefore, we have $\frac{\partial Q_6}{\partial x} > 0$; hence, Q_6 is increasing in x. Because $Q_6|_{x=0} = 16(1+\sigma_i)\left(3 + (5 - 2\gamma)\sigma_i + (2-\gamma)\sigma_i^2\right) > 0$; thus, we know that $Q_6 > 0$. Therefore, $\frac{\partial^2[\Pi_s(\sigma_i,x)+T(\sigma_j,x)]}{\partial\sigma_i\partial x} < 0$. That is, $\frac{\partial[\Pi_s(\sigma_i,x)+T(\sigma_j,x)]}{\partial\sigma_i}$ is decreasing in x. Suppose that the seller deviates by setting the data selling volume at $\dot{x}$ ($\dot{x} > x_H^{sep}$) and that the buyer believes that the seller is of a low type. By applying the above property, we obtain that $\Pi_s(\sigma_i, \dot{x}) + T(\sigma_j, \dot{x}) - [\Pi_s(\sigma_i, x_H^{sep}) + T(\sigma_j, x_H^{sep})]$ is decreasing in σ_i. Hence, $\Pi_s(\sigma_H, \dot{x}) + T(\sigma_j, \dot{x}) - [\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_j, x_H^{sep})] < \Pi_s(\sigma_L, \dot{x}) + T(\sigma_j, \dot{x}) - [\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_j, x_H^{sep})]$. Let $j = L$; $\Pi_s(\sigma_H, \dot{x}) + T(\sigma_L, \dot{x}) - [\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_L, x_H^{sep})] < \Pi_s(\sigma_L, \dot{x}) + T(\sigma_L, \dot{x}) - [\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_L, x_H^{sep})]$. We simultaneously add the expression $-T(\sigma_H, x_H^{sep}) + T(\sigma_L, x_H^{sep})$ on each side. We then obtain $\Pi_s(\sigma_H, \dot{x}) + T(\sigma_L, \dot{x}) - [\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep})] < \Pi_s(\sigma_L, \dot{x}) + T(\sigma_L, \dot{x}) - [\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_H, x_H^{sep})]$. From condition (C1$'$), we know that $\Pi_s(\sigma_L, \dot{x}) + T(\sigma_L, \dot{x}) - [\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_H, x_H^{sep})] < 0$. Hence, we obtain $\Pi_s(\sigma_H, \dot{x}) + T(\sigma_L, \dot{x}) < [\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep})]$. Thus, the high-type seller does not deviate from the separating equilibrium, and the above separating equilibrium can be sustained. Specifically, the high-type seller sells data with volume $x_H^{sep} \in (0,1)$ at a price $T(\sigma_H, x_H^{sep})$, whereas the low-type seller still follows his first-best solution with selling volume $x_L^{sep} = 1$ at a price $T(\sigma_L, 1)$. This concludes the proof. □

Proof of Proposition 3

In this proof, we derive the seller's optimal selling decision in the complete information setting, given the decision of free samples provision (x_1). If the seller does not sell data, the expected profit is $\Pi_s(\sigma, x_1)$. Otherwise, if the seller decides to sell the data with volume x_2, the corresponding expected profit is

$\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2) - T(\sigma, x_1)$. It is easy to show that $\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2)$ is increasing in x_2 for any fixed x_1. Hence, given the option of selling data, the seller will sell all the remaining data; that is, $x_2 = 1 - x_1$. The corresponding price is given by $T(\sigma, 1) - T(\sigma, x_1)$. To determine the seller's optimal decision on data selling, we need to compare $\Pi_s(\sigma, x_1)$ and $\Pi_s(\sigma, 1) + T(\sigma, 1) - T(\sigma, x_1)$. Because $\Pi_s(\sigma, x) + T(\sigma, x)$ is increasing in x, we have $\Pi_s(\sigma, 1) + T(\sigma, 1) - T(\sigma, x_1) - \Pi_s(\sigma, x_1) > 0$. Therefore, the seller sells all the remaining data. This concludes the proof. □

Proof of Proposition 4

In this proof, we derive the seller's optimal free samples provision in the complete information setting. The seller's profit is given by $\Pi_s(\sigma, 1) + T(\sigma, 1) - T(\sigma, x_1)$. Because $T(\sigma, x_1)$ is increasing in x_1, the seller will not offer free samples in the complete information setting. This concludes the proof. □

Proof of Proposition 5

In this proof, with free samples provision, we derive the seller's optimal data selling decision in an incomplete information setting under the separating equilibrium. Nevertheless, the low-type seller has an incentive to mimic the high-type seller's decision. Note that we consider the case in which the sellers are separated by the decision on offering free samples rather than that on the selling volume, which is due to two reasons. First, given any positive volume of sample data, if the sellers are separated by the decision on the data selling volume, the sellers should select the same volume of samples to sustain the buyer's prior belief at the stage of offering free samples. By Proposition 4, offering free samples hurts the sellers. Therefore, in the context of the separating equilibrium, the low-type seller would not prefer separation by the decision to sell data. Second, if we do not restrict the positive volume of sample data (especially for the high-type seller), the equilibrium outcome will be the same as that under the scenario without free samples, which makes our analysis in this section trivial. Note also that if the sellers are separated by the decision on free samples, they will sell all the remaining data (see Proposition 3).

We consider the following buyer's belief supporting the corresponding separating equilibrium. If $x_1 \in \Phi$, the buyer will believe that the market is of a high type. Otherwise, the buyer will believe that the market is of a low type. A similar belief structure has been widely considered in previous studies (i.e., Guo and Jiang 2016, Chen and Jiang 2021, Li et al. 2021). Therefore, a separating equilibrium exists if the following constraints are satisfied:

$$\begin{cases} \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_L, 1) + T(\sigma_L, 1) - T(\sigma_L, x_1)}_{Low-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_L, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C3)} \\ \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_H, 1) + T(\sigma_L, 1) - T(\sigma_L, x_1)}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C4)} \end{cases}$$

For the low-type seller, his profit without mimicking is given by $\Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$, which is equal to his first-best profit. To ensure the condition (C3), we need $\Pi_s(\sigma_L, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1) \leq \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$; hence, $T(\sigma_H, x_1) \geq T(\sigma_H, 1) - T(\sigma_L, 1)$. Because $T(\sigma_H, x_1)$ is increasing in x_1, there exists a threshold x_{1H}^{sep} such that Φ captures a region of $x_{1H}^{sep} \leq x_1 \leq 1$. Because $\Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)$ is decreasing in x_1, the high-type seller will select the volume of free samples at $x_1 = x_{1H}^{sep}$, where condition (C4) can be satisfied. An approach similar to that adopted in the Proof of Proposition 2 leads to the result that the above separating equilibrium can be sustained. This concludes the proof. □

Proof of Proposition 6

In this proof, we derive the strategic impacts of free samples. We first prove that $T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right) > T(\sigma_H, x_H^{sep})$. We know that x_H^{sep} is derived by letting $\Pi_s\left(\sigma_L, x_H^{sep}\right) + T\left(\sigma_H, x_H^{sep}\right) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Because $\Pi_s\left(\sigma_L, x_H^{sep}\right) > \Pi_s(\sigma_L, 1)$ for any $0 < x_H^{sep} < 1$, we have $T\left(\sigma_H, x_H^{sep}\right) < T(\sigma_L, 1)$. Proposition 5 yields the monetary payment for data in the case with free samples being $T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right) = T(\sigma_L, 1)$. Therefore, we show that the monetary payment for data in the case without free samples is lower, i.e., $T\left(\sigma_H, x_H^{sep}\right) < T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right)$. For the data selling volumes, the above result implies that $T\left(\sigma_H, x_H^{sep}\right) - T(\sigma_H, 0) < T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right)$. Suppose that $x_H^{sep} - 0 > 1 - x_{1H}^{sep}$; because $\frac{\partial^2 T(\sigma,x)}{\partial x^2} < 0$ and $\frac{\partial T(\sigma,x)}{\partial x} > 0$, we should have $T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right) < T\left(\sigma_H, x_H^{sep}\right) - T(\sigma_H, 0)$, which contradicts $T\left(\sigma_H, x_H^{sep}\right) - T(\sigma_H, 0) < T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right)$. Hence, in equilibrium, $x_H^{sep} < 1 - x_{1H}^{sep}$. Figure A1 provides an illustration.

For the high-type seller's profit, $\Pi_{sH}^{N,sep} - \Pi_{sH}^{F,sep} = \Pi_s\left(\sigma_H, x_H^{sep}\right) + T(\sigma_H, x_H^{sep}) - \left[\Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T\left(\sigma_H, x_{1H}^{sep}\right)\right] = \Pi_s\left(\sigma_H, x_H^{sep}\right) - \Pi_s(\sigma_H, 1) - \left[\Pi_s\left(\sigma_L, x_H^{sep}\right) - \Pi_s(\sigma_L, 1)\right]$. This result can be obtained by using $T\left(\sigma_H, x_{1H}^{sep}\right) = T(\sigma_H, 1) - T(\sigma_L, 1)$ and $\Pi_s\left(\sigma_L, x_H^{sep}\right) + T\left(\sigma_H, x_H^{sep}\right) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Because $\frac{\partial[\Pi_s(\sigma, x_H^{sep}) - \Pi_s(\sigma,1)]}{\partial \sigma} > 0$, we have $\Pi_s\left(\sigma_H, x_H^{sep}\right) - \Pi_s(\sigma_H, 1) - \left[\Pi_s\left(\sigma_L, x_H^{sep}\right) - \Pi_s(\sigma_L, 1)\right] > 0$. Therefore, $\Pi_{sH}^{N,sep} > \Pi_{sH}^{F,sep}$.

For the consumer surplus, the detailed expressions are given by

$$CS^{F,sep} = \mathbb{E}_{\{S_s,S_b\}}\begin{bmatrix} \frac{(1-\beta)}{2}\left[q_s^2(\sigma_H,1) + 2\gamma q_s(\sigma_H,1) q_b(\sigma_H,1) + q_b^2(\sigma_H,1)\right] \\ +\frac{\beta}{2}\left[q_s^2(\sigma_L,1) + 2\gamma q_s(\sigma_L,1) q_b(\sigma_L,1) + q_b^2(\sigma_L,1)\right] \end{bmatrix},$$

$$CS^{N,sep} = \mathbb{E}_{\{S_s,S_b\}}\begin{bmatrix} \frac{(1-\beta)}{2}\left[q_s^2\left(\sigma_H, x_H^{sep}\right) + 2\gamma q_s\left(\sigma_H, x_H^{sep}\right) q_b\left(\sigma_H, x_H^{sep}\right) + q_b^2\left(\sigma_H, x_H^{sep}\right)\right] \\ +\frac{\beta}{2}\left[q_s^2(\sigma_L,1) + 2\gamma q_s(\sigma_L,1) q_b(\sigma_L,1) + q_b^2(\sigma_L,1)\right] \end{bmatrix}.$$

We have $CS^{F,sep} - CS^{N,sep} = \mathbb{E}_{\{S_s,S_b\}}\left[\frac{(1-\beta)}{2}\left[q_s^2(\sigma_H,1) + 2\gamma q_s(\sigma_H,1) q_b(\sigma_H,1) + q_b^2(\sigma_H,1)\right] - \frac{(1-\beta)}{2}\left[q_s^2\left(\sigma_H, x_H^{sep}\right) + 2\gamma q_s\left(\sigma_H, x_H^{sep}\right) q_b\left(\sigma_H, x_H^{sep}\right) + q_b^2\left(\sigma_H, x_H^{sep}\right)\right]\right]$.

Hence, the function $\mathbb{E}_{\{S_s,S_b\}}[q_s^2(\sigma,x)+2\gamma q_s(\sigma,x)q_b(\sigma,x)+q_b^2(\sigma,x)]$ is the key factor. Note that

$$\frac{\partial \mathbb{E}_{\{S_s,S_b\}}[q_s^2(\sigma,x)+2\gamma q_s(\sigma,x)q_b(\sigma,x)+q_b^2(\sigma,x)]}{\partial x}=\frac{\sigma^2(2+(2-\gamma)\sigma)\begin{bmatrix}\left(4(2+\gamma)+x(16+4\gamma-10\gamma^2)\right)\sigma^2\\+x(2-\gamma)^2(2+3\gamma)\sigma^3+8+4(4+2x+\gamma)\sigma\end{bmatrix}}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}>0.$$

Hence, the function $\mathbb{E}_{\{S_s,S_b\}}[q_s^2(\sigma,x)+2\gamma q_s(\sigma,x)q_b(\sigma,x)+q_b^2(\sigma,x)]$ is increasing in x. With free samples, the buyer will obtain all of the seller's data from both free samples and sales; therefore, $CS^{F,sep}>CS^{N,sep}$.

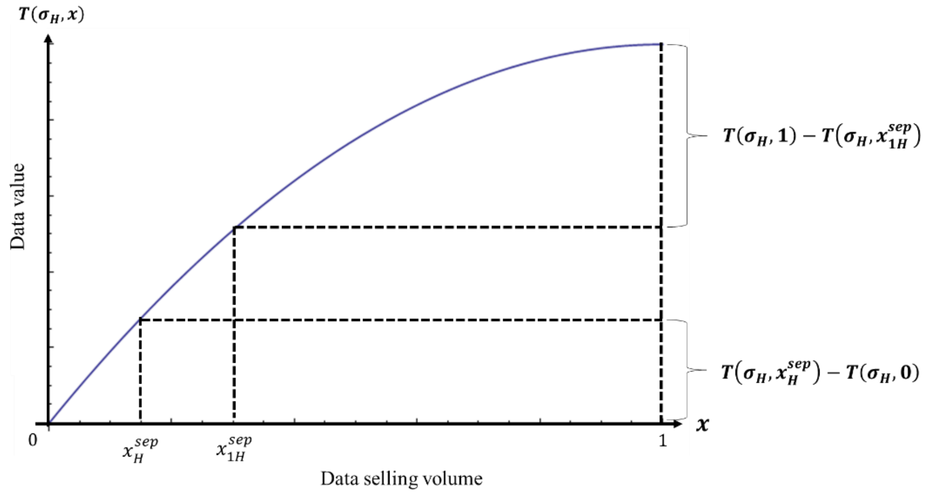

Figure A1. Illustration of the results.

This concludes the proof. □

Proof of Proposition 7

We derive the pooling equilibrium in the scenario without free samples. In such an equilibrium, the buyer cannot learn the seller type. Therefore, given any volume of data x, the buyer's maximum willingness to pay is $\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)$. We consider the following belief supporting the pooling equilibrium: if $(x,p_d)=(x^{pool},p_d^{pool})$, the buyer's belief is the same as the prior belief; otherwise, the buyer believes that the market is of a low type. A similar belief structure has been widely considered in previous studies (i.e., Guo and Jiang 2016, Chen and Jiang 2021, Li et al. 2021). Anticipating the buyer's purchasing decision, the high-type seller selects the optimal x^{pool} and p_d^{pool} by solving the following maximization problem:

$$(x^{pool},p_d^{pool})=\underset{x,p_d}{\operatorname{argmax}}\ \Pi_s(\sigma_H,x)+p_d$$

$$s.t.\qquad p_d\le \beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x).$$

The above problem can be reduced to the following: $x^{pool}=\operatorname{argmax}\Pi_s(\sigma_H,x)+\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)$, and $p_d^{pool}=\beta T(\sigma_L,x^{pool})+(1-\beta)T(\sigma_H,x^{pool})$. To sustain the above pooling equilibrium, we need the following conditions to hold:

$$\begin{cases} \underbrace{\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta) T(\sigma_H, x^{pool})}_{Low-type\ seller's\ profit\ under\ pooling} \geq \underbrace{max_{x \neq x^{pool}} \Pi_s(\sigma_L, x) + T(\sigma_L, x)}_{Low-type\ seller's\ profit\ under\ deviating}, & (C5) \\ \underbrace{\Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta) T(\sigma_H, x^{pool})}_{High-type\ seller's\ profit\ under\ pooling} \geq \underbrace{max_{x \neq x^{pool}} \Pi_s(\sigma_H, x) + T(\sigma_L, x)}_{High-type\ seller's\ profit\ under\ deviating}. & (C6) \end{cases}$$

For the high-type seller, because $T(\sigma_H, x) > T(\sigma_L, x)$, the maximum of $\Pi_s(\sigma_H, x) + \beta T(\sigma_L, x) + (1 - \beta) T(\sigma_H, x)$ is always greater than the maximum of $\Pi_s(\sigma_H, x) + T(\sigma_L, x)$. Hence, condition (C6) holds. For the low-type seller, we show later (in the Proof of Proposition 8) that if the pooling equilibrium outperforms the separating equilibrium, the low-type seller will not deviate (i.e., condition (C5) will hold). This concludes the proof. □

Proof of Proposition 8

In this proof, we compare the outcomes of the separating equilibrium in Proposition 2 and the pooling equilibrium in Proposition 7. Applying the refinement of the LMSE, we first select the unique equilibrium outcome. We then prove that if the pooling equilibrium outperforms the separating equilibrium, the data selling volume under the pooling equilibrium is greater than that under the separating equilibrium; i.e., $x_H^{sep} < x^{pool}$. The LMSE selects the unique equilibrium outcome from the high-type seller's perspective. The high-type seller's profit under the separating equilibrium is given by $\Pi_{sH}^{N,sep} = \pi_s\left(\sigma_H, x_H^{sep}\right) + T(\sigma_H, x_H^{sep})$, and her profit under the pooling equilibrium is given by

$$\Pi_{sH}^{N,pool} = \max \Pi_s(\sigma_H, x) + \beta T(\sigma_L, x) + (1-\beta) T(\sigma_H, x)$$
$$= \Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta) T(\sigma_H, x^{pool}).$$

Applying the envelope theorem, we obtain that $\frac{d\Pi_{sH}^{N,pool}}{d\beta} = \frac{\partial[\Pi_s(\sigma_H,x)+\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)]}{\partial\beta}|_{x=x^{pool}} = T(\sigma_L, x^{pool}) - T(\sigma_H, x^{pool}) < 0$. Hence, the high-type seller's profit under the pooling equilibrium decreases in β. However, her profit under the separating equilibrium is irrelevant with respect to β. $\Pi_{sH}^{N,pool}|_{\beta=0} = \max_{0<x<1} \Pi_s(\sigma_H, x) + T(\sigma_H, x)$, while her profit under the separating equilibrium is derived as $\Pi_{sH}^{N,sep} = \max_{x\in\Phi} \Pi_s(\sigma_H, x) + T(\sigma_H, x)$. Hence, we obtain that in this case, $\Pi_{sH}^{N,pool} > \Pi_{sH}^{N,sep}$, which occurs because region Φ belongs to $0 < x < 1$. Moreover, $\Pi_{sH}^{N,pool}|_{\beta=1} = \max_{0<x<1} \Pi_s(\sigma_H, x) + T(\sigma_L, x)$, while the high-type seller's profit under the separating equilibrium is given by $\Pi_{sH}^{N,sep} = \Pi_s\left(\sigma_H, x_H^{sep}\right) + T(\sigma_H, x_H^{sep})$. From condition (C2′) in the Proof of Proposition 2, we have $\Pi_{sH}^{N,pool} < \Pi_{sH}^{N,sep}$ in this case. Therefore, there exists a threshold $\bar{\beta}$; if $\beta \leq \bar{\beta}$, we have $\Pi_{sH}^{N,sep} \leq \Pi_{sH}^{N,pool}$; otherwise, we have $\Pi_{sH}^{N,sep} > \Pi_{sH}^{N,pool}$. Next, we prove that if the pooling equilibrium outperforms the separating equilibrium, $x_H^{sep} < x^{pool}$. If the pooling equilibrium outperforms the separating equilibrium, we have that $\Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta) T(\sigma_H, x^{pool}) \geq \Pi_s\left(\sigma_H, x_H^{sep}\right) + T(\sigma_H, x_H^{sep})$. Note that the term $\pi_s(\sigma_H, x) + T(\sigma_H, x)$ is increasing in x; hence, the optimal data selling volume x_H^{sep} is actually the upper bound of region Φ. Applying the approach of proof by contradiction, if $x_H^{sep} > x^{pool}$, we have $x^{pool} \in \Phi$. For any given selling volume x, the high-type seller's profit function under the pooling equilibrium, i.e., $\Pi_s(\sigma_H, x) + \beta T(\sigma_L, x) + (1-\beta)T(\sigma_H, x)$, is always lower than that under the separating equilibrium, i.e., $\Pi_s(\sigma_H, x) + T(\sigma_H, x)$. Hence, $\max_{x\in\Phi}\Pi_s(\sigma_H, x) + \beta T(\sigma_L, x) + (1-\beta)T(\sigma_H, x)$ cannot be higher than $\max_{x\in\Phi}\Pi_s(\sigma_H, x) + T(\sigma_H, x)$, which contradicts $\Pi_{sH}^{N,sep} \le \Pi_{sH}^{N,pool}$. Therefore, we have proven that $x_H^{sep} < x^{pool}$. Because $x_L^{sep} = 1$, thus $x_H^{sep} < x^{pool} \le x_L^{sep}$.

Finally, we prove that if the pooling equilibrium outperforms the separating equilibrium, the low-type seller will not deviate; that is, condition (C5) in the Proof of Proposition 7 holds. The low-type seller's profit that stays with the pooling equilibrium is given by $\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$, and the profit that deviates is given by $\pi_s(\sigma_L, 1) + T(\sigma_L, 1)$.

Because $\Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool}) \ge \Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep})$, the low-type seller's profit under the pooling equilibrium has the following property: $\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool}) \ge \Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_H, x^{pool}) + \Pi_s(\sigma_L, x^{pool})$.

Moreover, from the Proof of Proposition 2, we have that $\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_H, x_H^{sep}) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. We then obtain $\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_H, x^{pool}) + \Pi_s(\sigma_L, x^{pool}) = \Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) - [\Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})] + \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. It is easy to show that $\frac{\partial^2\Pi_s(\sigma_i, x)}{\partial\sigma_i\partial x} < 0$; hence, $\frac{\partial[\Pi_s(\sigma_H, x) - \Pi_s(\sigma_L, x)]}{\partial x} < 0$. Because $x_H^{sep} < x^{pool}$, we have $\Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) > \Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})$, which indicates that

$$\underbrace{\Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) - [\Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})]}_{>0} + \Pi_s(\sigma_L, 1) + T(\sigma_L, 1) > \Pi_s(\sigma_L, 1) + T(\sigma_L, 1).$$

Therefore, we have $\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool}) > \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Hence, the low-type seller does not deviate from the pooling equilibrium. Thus, the pooling equilibrium characterized in Proposition 7 can be sustained. This concludes the proof. □

Appendix B: Extended analyses

In Appendix B, we present extended analyses designed to assess robustness and delineate the boundaries of our main results. Specifically, Part I examines the scenario in which the demand forecasting signals generated by the seller and buyer are correlated. Part II incorporates a data platform into our model, which earns a profit by charging a commission. Part III extends the analyses to the case where the products of the data seller and the data buyer are complements rather than substitutes. Part IV examines an extended model in which the volume of free samples, the volume of data sold, and the data selling price jointly serve as signaling instruments (i.e., multidimensional signaling). Part V examines the scenario where free samples have functional impacts and directly convey information to the buyer. Part VI explores an alternative equilibrium refinement via the intuitive criterion.

1. Part I: Correlated demand forecasting signals

In our main paper, to clearly capture the fundamental mechanism underlying data evaluation and sample provision, we assume that the signals of the seller and buyer are uncorrelated. In this section, we extend the analysis to explore the results if the signals are correlated. Specifically, the signals are given by $S_s = \theta + \varepsilon_s$ and $S_b = \theta + \varepsilon_b$, respectively. We assume that ε_s and ε_b are positively correlated and define $\rho = Cov(\varepsilon_s, \varepsilon_b)$ to characterize the correlation. By definition, $0 \le Cov(\varepsilon_s, \varepsilon_b) \le \sqrt{Var[\varepsilon_s]Var[\varepsilon_b]}$. Accordingly, without free samples, the feasible range of ρ should be $[0, \frac{1}{\sqrt{x}}]$, whereas with free samples, it becomes $\rho \in [0, \frac{1}{\sqrt{x_1+x_2}}]$. If $\rho = 0$, the analysis reduces to our main model. Even if ρ reaches its maximum, ε_s and ε_b are not necessarily identical; thus, the seller and the buyer do not always generate the same signals. This is due to the asymmetry in the data held by each party. The extreme case in which they indeed generate identical signals occurs only if the seller sells all the data ($x = 1$ without free samples or $x_2 = 1 - x_1$ with free samples) and $\rho = 1$ (fully correlated).

Accordingly, we have the following conditional expectations: $\mathbb{E}[\theta|S_s] = \frac{\sigma}{\sigma+1}S_s$, $\mathbb{E}[S_b|S_s] = \frac{\sigma+\rho}{\sigma+1}S_s$, $\mathbb{E}[\theta|S_b] = \frac{\sigma}{\sigma+\frac{1}{x}}S_b$ and $\mathbb{E}[S_s|S_b] = \frac{\sigma+\rho}{\sigma+\frac{1}{x}}S_b$ in the case without free samples and $\mathbb{E}[\theta|S_s] = \frac{\sigma}{\sigma+1}S_s$, $\mathbb{E}[S_b|S_s] = \frac{\sigma+\rho}{\sigma+1}S_s$, $\mathbb{E}[\theta|S_b] = \frac{\sigma}{\sigma+\frac{1}{x_1+x_2}}S_b$ and $\mathbb{E}[S_s|S_b] = \frac{\sigma+\rho}{\sigma+\frac{1}{x_1+x_2}}S_b$ in the case with free samples. The other settings are unchanged.

Note that introducing the correlation of signals complicates our analysis significantly, making it intractable in most cases. In the following part, we first provide an analytical treatment of the case of low competition intensity; then, we numerically verify the robustness of the results when otherwise.

Without free samples. Following an approach similar to that in the main paper, we can obtain the optimal sales quantities of the seller and the buyer:

$$\begin{aligned} q_s &= \frac{a}{2+\gamma} + \frac{\sigma(2-x\gamma\rho-x(-2+\gamma)\sigma)}{4-x\gamma^2\rho^2+\left(4+x(4-2\gamma^2\rho)\right)\sigma-x(-4+\gamma^2)\sigma^2} S_s, \\ q_b &= \frac{a}{2+\gamma} + \frac{x\sigma(2-\gamma\rho-(-2+\gamma)\sigma)}{4-x\gamma^2\rho^2+(4+x(4-2\gamma^2\rho))\sigma-x(-4+\gamma^2)\sigma^2} S_b. \end{aligned} \tag{BI-1}$$

Accordingly, by applying the properties of $\mathbb{E}[\theta^2] = \mathbb{E}[\theta S_s] = \mathbb{E}[\theta S_b] = \sigma$, $\mathbb{E}[S_s S_b] = \sigma + \rho$, $\mathbb{E}\left[{S_s}^2\right] = \sigma + 1$ and $\mathbb{E}\left[{S_b}^2\right] = \sigma + \frac{1}{x}$, we can obtain the expected profits of the seller and the buyer as follows:

$$\begin{aligned} \Pi_s(\sigma, x) &= \frac{4a^2+4a^2\sigma+(2+\gamma)^2\sigma^2}{4(2+\gamma)^2(1+\sigma)} + \frac{x\gamma\sigma^2(\rho+\sigma)(-2(1+\sigma)+\gamma(\rho+\sigma))\begin{bmatrix} 8(1+\sigma)-x(-8\sigma(1+\sigma) \\ +2\gamma(1+\sigma)(\rho+\sigma)+\gamma^2(\rho+\sigma)^2) \end{bmatrix}}{4(1+\sigma)(4(1+\sigma)+4x\sigma(1+\sigma)-x\gamma^2(\rho+\sigma)^2)^2}, \\ \Pi_b(\sigma, x) &= \frac{a^2}{(2+\gamma)^2} + \frac{x\sigma^2(1+x\sigma)(-2(1+\sigma)+\gamma(\rho+\sigma))^2}{(4(1+\sigma)+4x\sigma(1+\sigma)-x\gamma^2(\rho+\sigma)^2)^2}. \end{aligned} \tag{BI-2}$$

Therefore, the buyer's willingness to pay for the data is $\Pi_b(\sigma, x) - \Pi_b(\sigma, 0) \equiv T(\sigma, x)$.

In the complete information setting, the seller charges for the data at the buyer's willingness to pay, thereby maximizing $\Pi_s(\sigma, x) + T(\sigma, x)$ with respect to $x \in [0,1]$.

Proposition BI-1. *With complete information, the seller sells all the data to the buyer (i.e.,* $x^* = 1$*) at the price* $p_d = T(\sigma, 1)$.

Proof of Proposition BI-1. We first investigate how the buyer's willingness to pay changes with data selling volume, $\frac{\partial T(\sigma,x)}{\partial x} = \frac{\sigma^2(-2+\gamma\rho+(-2+\gamma)\sigma)^2[4+x\gamma^2\rho^2+2(2+x(2+\gamma^2\rho))\sigma+x(4+\gamma^2)\sigma^2]}{(4-x\gamma^2\rho^2+(4+x(4-2\gamma^2\rho))\sigma-x(-4+\gamma^2)\sigma^2)^3} > 0$. We know that it is increasing in the data selling volume. Adopting a similar approach, one can easily prove that $\frac{\partial T(\sigma,x)}{\partial \sigma} > 0$. By taking the first-order derivative of the seller's profit with respect to the data selling volume, we have $\frac{\partial[\Pi_s(\sigma,x)+T(\sigma,x)]}{\partial x} = \frac{\sigma^2(2-\gamma\rho+(2-\gamma)\sigma)B1}{(4-x\gamma^2\rho^2+2(2+x(2-\gamma^2\rho))\sigma+x(4-\gamma^2)\sigma^2)^3}$, where $B1 = 8 - 12\gamma\rho + 6x\gamma^2\rho^2 - x\gamma^3\rho^3 + (16 - 12\gamma(1+\rho) + x(8 - 12\gamma\rho - 3\gamma^3\rho^2 + 6\gamma^2\rho(2+\rho)))\sigma + (8 - 12\gamma + x(16 - 3\gamma^3\rho - 12\gamma(1+\rho) + 6\gamma^2(1+2\rho)))\sigma^2 - x(-2+\gamma)^3\sigma^3$. Moreover, $\frac{\partial B1}{\partial x} = \gamma^2\rho^2(6-\gamma\rho) + \left(8 - 12\gamma\rho - 3\gamma^3\rho^2 + 6\gamma^2\rho(2+\rho)\right)\sigma + \left(16 - 3\gamma^3\rho - 12\gamma(1+\rho) + 6\gamma^2(1+2\rho)\right)\sigma^2 - (-2+\gamma)^3\sigma^3$, which is increasing in σ. If $\sigma = 0$, $\frac{\partial B1}{\partial x} = \gamma^2\rho^2(6-\gamma\rho) > 0$, implying that $B1$ increases in x. $B1|_{x=0} = -4(1+\sigma)(-2 + 3\gamma\rho + (-2+3\gamma)\sigma) > 0$ if γ is limited to zero. Therefore, for a relatively small γ, $\frac{\partial[\Pi_s(\sigma,x)+T(\sigma,x)]}{\partial x} > 0$; thus, the seller will sell all the data. □

In an incomplete information setting, because $T(\sigma, x)$ increases in σ, the low-type seller still has an incentive to mimic the high-type seller, leading to a signaling game.

Proposition BI-2. *In the separating equilibrium without free samples, the high-type seller sells a portion of the data with volume* $x_H^{sep} \in (0,1)$ *at the price* $T(\sigma_H, x_H^{sep})$*, whereas the low-type seller adheres to his* *first-best solution by selling all the data (i.e.,* $x_L^{sep} = 1$*) at the price* $T(\sigma_L, 1)$.

Proof of Proposition BI-2. The corresponding separating equilibrium can still be derived from the following conditions:

$$\begin{cases} \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_L,x)+T(\sigma_L,x)}_{Low-type\ seller's\ first-best\ profit} \geq \underbrace{max_{x\in\Phi}\Pi_s(\sigma_L,x)+T(\sigma_H,x)}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C1)} \\ \underbrace{max_{x\in\Phi}\Pi_s(\sigma_H,x)+T(\sigma_H,x)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_H,x)+T(\sigma_L,x)}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C2)} \end{cases}$$

For condition C1, the low-type seller's first-best profit is given by $\Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. To capture the profit in the case of deviation and the composition of Φ, we first prove that if the signals are correlated, $\frac{\partial^2 T(\sigma,x)}{\partial x \partial \sigma} > 0$. $\frac{\partial^2 T(\sigma,x)}{\partial x \partial \sigma} = B2 * B3$, where $B2 = \frac{2\sigma(2-\gamma\rho+(2-\gamma)\sigma)}{(-4+x\gamma^2\rho^2+2(-2+x(-2+\gamma^2\rho))\sigma+x(-4+\gamma^2)\sigma^2)^4} > 0$ and $\frac{\partial^2 B3}{\partial x^2} = 2\gamma\Big(16(-1+\rho)\sigma^3(1+\sigma)+\gamma^4\rho(\rho+\sigma)^4+8\gamma^2\sigma(\rho+\sigma)^2(\rho-\sigma+2\rho\sigma)-2\gamma^3(\rho+\sigma)^3(\rho-\sigma+2\rho\sigma)-16\gamma\sigma(1+\sigma)\big(-\sigma^2+2\rho\sigma^2+\rho^2(1+2\sigma)\big)\Big) < 0$ for all $0<\gamma<1$, $\sigma>0$ and $0<\rho<1$. Hence, $\frac{\partial B3}{\partial x}$ decreases in x. Moreover, $\frac{\partial B3}{\partial x}|_{x=1} = 2\Big(16\sigma(1+\sigma)^3+\gamma^5\rho(\rho+\sigma)^4-2\gamma^4(\rho+\sigma)^3(\rho-\sigma+2\rho\sigma)+8\gamma\sigma(1+\sigma)^2\big(-3\sigma+\rho(-1+2\sigma)\big)+4\gamma^3\sigma(\rho+\sigma)^2\big(-2-3\sigma+\rho(3+4\sigma)\big)-8\gamma^2\sigma(1+\sigma)\big(-\sigma(2+3\sigma)+\rho^2(3+4\sigma)+\rho(-2+4\sigma^2)\big)\Big)$. Clearly, this equals $32\sigma+96\sigma^2+96\sigma^3+32\sigma^4>0$ when γ is limited to zero. Therefore, if γ is relatively small, $B3$ increases in x. $B3|_{x=0} = 16(1+\sigma)(2-\gamma\rho+2(2-\gamma)\sigma+(2-\gamma)\sigma^2) > 0$ if γ is limited to zero. Hence, $B3>0$, and thus, $\frac{\partial^2 T(\sigma,x)}{\partial x \partial \sigma} > 0$ if γ is relatively small. Following a similar reasoning to that in the main paper, we know that there exists a threshold x_H^{sep} such that Φ actually captures the region of $x \leq x_H^{sep}$. We then consider the high-type seller's problem. It is easy to verify that the high-type seller will select the optimal data selling volume at x_H^{sep}. Finally, we prove that the high-type seller has no incentive to deviate; therefore, condition C2 holds. To this end, we first prove that $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} < 0$.

Note that $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} = \frac{4\gamma\sigma B4}{(-4+x\gamma^2\rho^2+2(-2+x(-2+\gamma^2\rho))\sigma+x(-4+\gamma^2)\sigma^2)^4}$, where $\frac{\partial^4 B4}{\partial \rho^4} = 24x\gamma^3\Big(-4-6\sigma+x\big(5\gamma\rho(2+3\sigma)+\gamma\sigma(7+12\sigma)-2(2+9\sigma+8\sigma^2)\big)\Big)$. One can easily verify that $-4-6\sigma+x\big(5\gamma\rho(2+3\sigma)+\gamma\sigma(7+12\sigma)-2(2+9\sigma+8\sigma^2)\big)<0$ if γ is limited to zero. That is, for a relatively small γ, $\frac{\partial^3 B4}{\partial \rho^3}$ decreases in ρ. Similarly, we can prove that for a relatively small γ, $\frac{\partial^3 B4}{\partial \rho^3}|_{\rho=1} > 0$. Therefore, $\frac{\partial^2 B4}{\partial \rho^2}$ increases in ρ. $\frac{\partial^2 B4}{\partial \rho^2}|_{\rho=0} = 4\gamma\big(8+8x-4(-2+x^2+x(-7+4\gamma))\sigma+8(-1+x)x(-2+\gamma)\sigma^2+x(-4+8\gamma-6\gamma^2+x(-28+40\gamma-18\gamma^2+\gamma^3))\sigma^3+2x^2(-2+\gamma)(4-6\gamma+3\gamma^2)\sigma^4\big)$, which is positive if γ is limited to zero and $0 < \sigma < 1$. Thus, $\frac{\partial B4}{\partial \rho}$ increases in ρ. $\frac{\partial B4}{\partial \rho}|_{\rho=1} = (1+\sigma)^2 \left(x^2\gamma^4(1+\sigma)^2(10+3\sigma) + 16(-2-(1+2x)\sigma + x^2\sigma^3) + 16x\gamma^2\sigma(1+\sigma)\big(-1+x(3+2\sigma)\big) - 16x\gamma^3(1+\sigma)\big(1+x(1+3\sigma+\sigma^2)\big) - 16\gamma\big(-2-\sigma-2x(1+3\sigma+\sigma^2)+x^2\sigma(1+2\sigma+2\sigma^2)\big)\right)$. Nevertheless, if γ is limited to zero and the demand variance is low, i.e., $0 < \sigma < 1$, $\frac{\partial B4}{\partial \rho}|_{\rho=1} < 0$. Therefore, $B4$ decreases in ρ if both competition intensity and demand variance are low. Similarly, in this case, we can prove that $B4|_{\rho=0} < 0$ holds. Therefore, $B4 < 0$ implies that $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} < 0$ when both competition intensity and demand variance are low. Similarly, one can easily verify that if γ is limited to zero and $\sigma > 1$, $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} < 0$ still holds. That is, $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} < 0$ always holds if the intensity of competition is low.

Adopting the same approach as in the Proof of Proposition 2, one can determine that the high-type seller will not deviate from the above separating equilibrium; therefore, the separating equilibrium characterized in Proposition BI-2 can be sustained. □

With free samples. In this case, the total volume of data the buyer can use is $x_1 + x_2$ rather than x. In addition, the expected profits of the seller and the buyer are $\Pi_s(\sigma, x_1 + x_2)$ and $\Pi_b(\sigma, x_1 + x_2)$, respectively. Consequently, the buyer's willingness to pay for data is $T(\sigma, x_1 + x_2) - T(\sigma, x_1)$.

In the complete information setting, the seller charges for the data at the buyer's willingness to pay, thereby maximizing $\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2) - T(\sigma, x_1)$ with respect to $x_2 \in [0,1]$ when selling the data.

Proposition BI-3. *With complete information, given the sample data volume x_1, the seller sells all the remaining data (i.e., $x_2 = 1 - x_1$) at price $p_d = T(\sigma, 1) - T(\sigma, x_1)$.*

Proof of Proposition BI-3. It is easy to show that $\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2)$ is increasing in x_2 for any fixed x_1. Hence, the seller will sell all the remaining data, that is, $x_2 = 1 - x_1$. The corresponding price is given by $T(\sigma, 1) - T(\sigma, x_1)$. □

Accordingly, if the seller offers free samples with volume x_1, then the expected profit is given by $\Pi_s(\sigma, 1) + T(\sigma, 1) - T(\sigma, x_1)$, which decreases in the sample volume x_1. Thus, we obtain the following result.

Proposition BI-4. *With complete information, the seller does not offer free samples (i.e., $x_1{}^* = 0$).*

Proof of Proposition BI-4. For any given samples' volume, the seller's profit is given by $\Pi_s(\sigma, 1) + T(\sigma, 1) - T(\sigma, x_1)$. Because $T(\sigma, x_1)$ is increasing in x_1, the seller will not offer free samples in the complete information setting. □

In the incomplete information setting, the low-type seller has an incentive to mimic the high-type seller's decision. The corresponding path to achieve data evaluation is characterized as follows:

Proposition BI-5. *In the separating equilibrium with free samples, the high-type seller offers a volume* $x_{1H}^{sep} \in (0,1)$ *of free samples to the buyer, whereas the low-type seller does not offer any free samples (i.e.,* $x_{1L}^{sep} = 0$*).*

Proof of Proposition BI-5. The above separating equilibrium can be obtained from the following conditions:

$$\begin{cases} \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_L, 1) + T(\sigma_L, 1) - T(\sigma_L, x_1)}_{Low-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_L, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)}_{Low-type\ seller's\ profit\ under\ deviating}, (\text{C3}) \\ \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_H, 1) + T(\sigma_L, 1) - T(\sigma_L, x_1)}_{High-type\ seller's\ profit\ under\ deviating}. (\text{C4}) \end{cases}$$

Because considering the correlation of signals does not fundamentally alter the fact that $\frac{\partial T(\sigma,x)}{\partial x} > 0$, one can easily derive the results by adopting the same approach as in the Proof of Proposition 5. □

Impact of free samples on data evaluation. Next, we compare the results under these scenarios to highlight the strategic impacts of free samples.

Proposition BI-6. *Comparing the separating equilibria with and without free samples yields the following strategic effects of free samples when achieving data evaluation:*

(i) free samples incentivize a higher volume of data selling (i.e., $1 - x_{1H}^{sep} > x_H^{sep}$*),*

(ii) free samples increase the monetary payment for data (i.e., $T(\sigma_H, 1) - T(\sigma_H, x_{1H}^{sep}) > T(\sigma_H, x_H^{sep})$*), and*

(iii) free samples hurt the high-type seller (i.e., $\Pi_{sH}^{F,sep} < \Pi_{sH}^{N,sep}$*).*

Proof of Proposition BI-6. The result regarding the monetary payment for data can be easily proven by adopting the same procedure as in the Proof of Proposition 6. For the result for the data selling volume, we need to first ensure that $\frac{\partial^2 T(\sigma,x)}{\partial x^2} < 0$, and then we can adopt the same procedure in the Proof of Proposition 6 to finish the proof. $\frac{\partial^2 T(\sigma,x)}{\partial x^2} = \frac{2\sigma^2(-2+\gamma\rho+(-2+\gamma)\sigma)^2 B5}{(-4+x\gamma^2\rho^2+2(-2+x(-2+\gamma^2\rho))\sigma+x(-4+\gamma^2)\sigma^2)^4}$, where $B5 = \gamma^2\rho^2(8 + x\gamma^2\rho^2) + 4(-4 + x\gamma^4\rho^3 + 2\gamma^2\rho(2+\rho))\sigma + 2(-16 + 4\gamma^2(1+2\rho) + x(-8+3\gamma^4\rho^2))\sigma^2 + 4(-4 - 8x + 2\gamma^2 + x\gamma^4\rho)\sigma^3 + x(-16+\gamma^4)\sigma^4$. Taking the first-order derivative with respect to x yields $\frac{\partial B5}{\partial x} = \gamma^4\rho^4 + 4\gamma^4\rho^3\sigma + 2(-8+3\gamma^4\rho^2)\sigma^2 + 4(-8+\gamma^4\rho)\sigma^3 + (-16+\gamma^4)\sigma^4 < 0$ if γ is limited to zero. That is, for a low competition intensity, $B5$ decreases in x. $B5|_{x=0} = 8(1+\sigma)(\gamma^2\rho^2 + 2(-1+\gamma^2\rho)\sigma + (-2+\gamma^2)\sigma^2) < 0$ always holds for a low competition intensity; hence, $B5 < 0$, and $\frac{\partial^2 T(\sigma,x)}{\partial x^2} < 0$. Given this, we can prove that $1 - x_{1H}^{sep} > x_H^{sep}$ by adopting the same procedure as in the Proof of Proposition 6.

For the high-type seller's profits, we have proven that $\frac{\partial^2 \Pi_s(\sigma,x)}{\partial x \partial \sigma} < 0$ if the signals are correlated. The same procedure can be used to prove that $\Pi_{sH}^{F,sep} < \Pi_{sH}^{N,sep}$. □

Strategic impact of data evaluation. Next, we examine the scenario in which the buyer is unable to evaluate the data during the purchase (i.e., a pooling equilibrium arises).

Proposition BI-7. *In the pooling equilibrium (i.e., without data evaluation), both types of sellers set the same data selling volume* $x^{pool} \in [0,1]$ *at the price* $\beta T(\sigma_L, x^{pool}) + (1-\beta) T(\sigma_H, x^{pool})$.

Proof of Proposition BI-7. The above pooling equilibrium can be obtained by adopting the same procedure as in the Proof of Proposition 7. □

Proposition BI-8. *There exists a threshold* $\bar{\beta} \in (0,1)$ *such that if* $\beta \leq \bar{\beta}$, *the pooling equilibrium is selected, and data evaluation hurts both types of sellers, i.e.,* $\Pi_{sH}^{N,sep} \leq \Pi_{sH}^{N,pool}$ *and* $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$; *otherwise, the separating equilibrium is selected, and data evaluation benefits the high-type seller but hurts the low-type seller, i.e.,* $\Pi_{sH}^{N,sep} > \Pi_{sH}^{N,pool}$ *and* $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$.

Proof of Proposition BI-8. The equilibrium outcome selected by LMSE can be proven by adopting the same procedure as in the Proof of Proposition 8. □

The above analyses imply that even if the signals generated by the seller and the buyer are correlated, we can analytically prove that our main results remain qualitatively unchanged for a low level of competition intensity.

In the following section, owing to the analysis being intractable, we numerically verify our results for a high competition intensity. To this end, we set $a = 1$, $\sigma_L = 0.5$, $\sigma_H = 1$ and $\gamma = 0.6$.

Figure B1 compares the data selling volume in the scenarios with and without free samples. The numerical results support our main finding: offering free samples leads to a higher volume of data sales. That is, this conclusion holds qualitatively the same for a high level of competition intensity.

Figure B2 compares the high-type seller's profits under the scenarios with and without free samples if data evaluation is achieved. Regardless of the correlation of signals, our numerical results show that the high-type seller prefers not providing free samples, which is consistent with our main findings. Regarding the equilibrium outcome, Figure B3 illustrates that there still exists a threshold with respect to the probability assigned to the low-type seller. If the seller is more likely to be a high-type seller, the equilibrium outcome is more likely to be the pooling equilibrium; otherwise, the separating equilibrium appears to be more attractive to the high-type seller. The above numerical analysis reveals that our main findings remain qualitatively unchanged if market competition is intense.

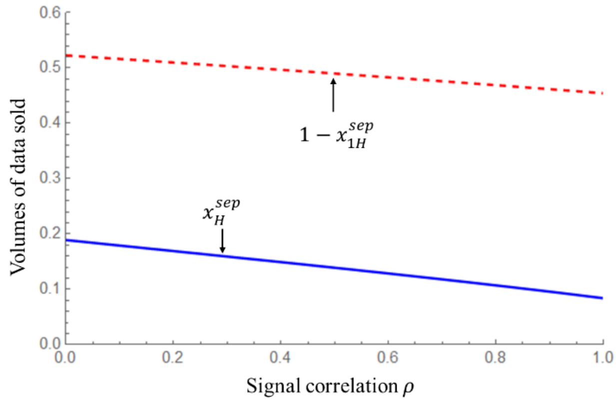

Figure B1. Volume of data sold with ρ $(a=1,\ \sigma_L=0.5,\ \sigma_H=1,\ \gamma=0.6)$.

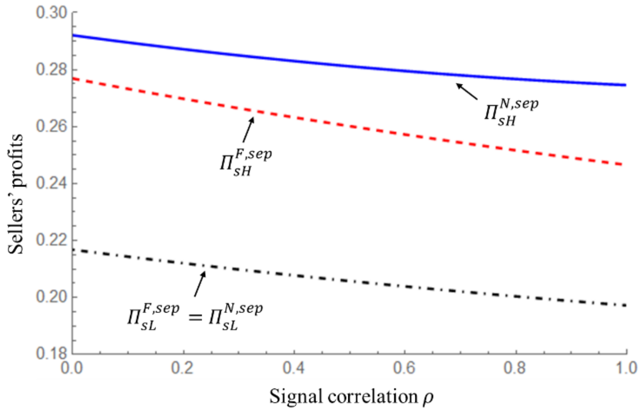

Figure B2. Sellers' profits with ρ $(a=1,\ \sigma_L=0.5,\ \sigma_H=1,\ \gamma=0.6)$.

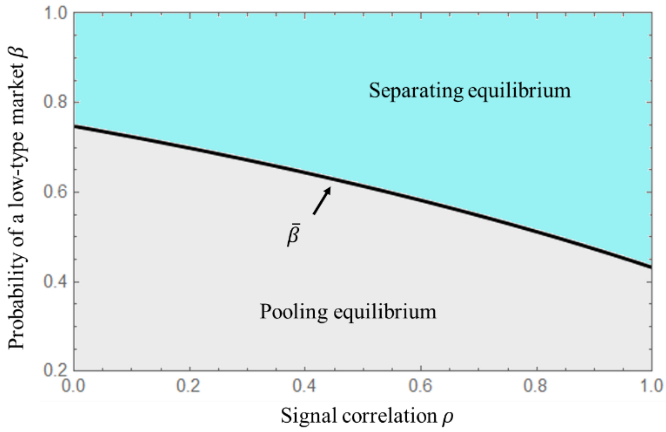

Figure B3. Equilibrium outcome with ρ $(a=1,\ \sigma_L=0.5,\ \sigma_H=1,\ \gamma=0.6)$.

In summary, while we assume uncorrelated signals in the main model for clarity and analytical tractability, our results are not driven by this assumption.

Furthermore, we obtain new insights into the impact of signal correlation (ρ) on equilibrium outcomes in data evaluation. First, under the separating equilibrium (i.e., data evaluation), an increase in ρ reduces the high-type seller's optimal volume of data sold in both scenarios, with and without free samples (see Figure

B1 for an illustration).[1] The intuition is as follows. A higher signal correlation reduces the buyer's willingness to pay for the data (i.e., $\frac{\partial T(\sigma,x)}{\partial\rho} < 0$) and the marginal value of the data (i.e., $\frac{\partial^2 T(\sigma,x)}{\partial x\partial\rho} < 0$). This effect strengthens the low-type seller's incentive to mimic the high-type seller, as a high-variance market preserves the value of the data (i.e., $\frac{\partial T(\sigma,x)}{\partial\sigma} > 0$ and $\frac{\partial^2 T(\sigma,x)}{\partial x\partial\sigma} > 0$). Consequently, without free samples, the high-type seller must sell less data to credibly signal her type. If free samples are offered, she must provide more of them to signal her type, thereby reducing the amount of data sold. Second, as shown in Figure B2, under the separating equilibrium, the profits of both seller types decrease in ρ in both scenarios—with and without free samples—due to the reduced value of the data.[2] Third, as shown in Figure B3, an increase in ρ makes the separating equilibrium (i.e., data evaluation) more likely to emerge as the final equilibrium outcome. The intuition again is that while a higher signal correlation lowers the data value, a high-variance market enhances it. These opposing forces strengthen the high-type seller's incentive to distinguish herself from the low-type seller.

[1] Under the separating equilibrium, the low-type seller sells all the data in both scenarios, with and without free samples.

[2] Under the separating equilibrium, the low-type seller adopts his first-best strategy—namely, refraining from providing free samples and selling all available data—and earns the same profit in both scenarios.

2. Part II: Incorporating a data platform

In our main model, we capture the mechanism underlying data trading while omitting the role of the platform. In practice, however, data trading typically occurs on data platforms (e.g., Datarade) that earn profits by charging commissions. In this section, we verify the robustness of our results in the presence of a data platform and examine the platform's preferences regarding free samples and data evaluation. Specifically, we assume that the platform earns a profit by charging a commission to the seller for each data transaction at a rate α. The other settings are unchanged. Note that the seller's incentive to sell data hinges critically on the commission rate. To avoid a trivial case in which the seller chooses not to sell data, we focus on the scenario with a low commission rate, i.e., $\alpha < \frac{2+(2-3\gamma)\sigma_H}{2+(2-\gamma)\sigma_H}$.

Without free samples. We first consider the scenario with complete information and then consider the scenario with incomplete information to explore the mechanism of data evaluation.

In the complete information setting, the buyer's willingness to pay for data is $T(\sigma, x)$. The seller will charge for the data exactly at the buyer's willingness to pay, thereby maximizing $\Pi_s(\sigma, x) + (1-\alpha)T(\sigma, x)$ with respect to $0 < x < 1$.

Proposition BII-1. *With complete information, the seller sells all the data to the buyer (i.e., $x^* = 1$) at the price $p_d = T(\sigma, 1)$.*

Proof of Proposition BII-1. The seller's problem can be formulated as $\max[\Pi_s(\sigma, x) + (1-\alpha)T(\sigma, x)]$ with respect to $0 < x < 1$. $\frac{\partial[\Pi_s(\sigma,x)+(1-\alpha)T(\sigma,x)]}{\partial x} = \frac{\sigma^2(2+(2-\gamma)\sigma)B6}{(4+4(1+x)\sigma+x(4-\gamma^2)\sigma^2)^3}$, where $B6 = 8 + 4(4 + 2x - 3\gamma)\sigma + 2(4 - 6\gamma + x(8 - 6\gamma + 3\gamma^2))\sigma^2 - x(-2+\gamma)^3\sigma^3 + \alpha(-2 + (-2+\gamma)\sigma)(4 + 4(1+x)\sigma + x(4+\gamma^2)\sigma^2)$. Moreover, $\frac{\partial B6}{\partial x} = \sigma(8 + 16\sigma - 12\gamma\sigma + 6\gamma^2\sigma - (-2+\gamma)^3\sigma^2) + \alpha\sigma(-8 - 2(8 - 2\gamma + \gamma^2)\sigma + (-2+\gamma)(4+\gamma^2)\sigma^2)$. It is easy to verify that if $0 < \gamma < \frac{2}{3}$ and $0 < \alpha < \frac{2+(2-3\gamma)\sigma}{2+(2-\gamma)\sigma}$, $\frac{\partial B6}{\partial x} > 0$ and $B6|_{x=0} > 0$; therefore, $B6 > 0$ and $\Pi_s(\sigma, x) + (1-\alpha)T(\sigma, x)$ is increasing in x. Therefore, the seller will sell all the data to the buyer. Because $\frac{2+(2-3\gamma)\sigma}{2+(2-\gamma)\sigma}$ decreases in σ, in the following analysis, we focus on the case of $0 < \alpha < \frac{2+(2-3\gamma)\sigma_H}{2+(2-\gamma)\sigma_H}$. □

In the incomplete information setting, because the low-type seller has an incentive to mimic the high-type seller, a signaling game needs to be considered.

Proposition BII-2. *In the separating equilibrium without free samples, the high-type seller sells a portion of the data with volume $x_H^{sep} \in (0,1)$ at the price $T(\sigma_H, x_H^{sep})$, whereas the low-type seller adheres to his first-best solution by selling all the data (i.e., $x_L^{sep} = 1$) at the price $T(\sigma_L, 1)$.*

Proof of Proposition BII-2. The above separating equilibrium can be obtained by the following conditions:

$$\begin{cases} \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_L,x)}_{Low-type\ seller's\ first-best\ profit} \geq \underbrace{max_{x\in\Phi}\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_H,x)}_{Low-type\ seller's\ profit\ under\ deviating}, & (C7)\\ \underbrace{max_{x\in\Phi}\Pi_s(\sigma_H,x)+(1-\alpha)T(\sigma_H,x)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x\notin\Phi}\Pi_s(\sigma_H,x)+(1-\alpha)T(\sigma_L,x)}_{High-type\ seller's\ profit\ under\ deviating}. & (C8)\end{cases}$$

For the low-type seller, his profit without mimicking is exactly the first-best profit, given by $\Pi_s(\sigma_L,1)+(1-\alpha)T(\sigma_L,1)$. Since $\frac{\partial^2T(\sigma,x)}{\partial x\partial\sigma}>0$, it is easy to verify that $\frac{\partial[\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_H,x)]}{\partial x}>\frac{\partial[\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_L,x)]}{\partial x}>0$. The minimum of $\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_H,x)$ is given by $\Pi_s(\sigma_L,0)+(1-\alpha)T(\sigma_H,0)$, which is lower than the low-type seller's first-best profit. The maximum of $\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_H,x)$ is given by $\Pi_s(\sigma_L,1)+(1-\alpha)T(\sigma_H,1)$, which is greater than the low-type seller's first-best profit. The reason is that $T(\sigma_H,1)>T(\sigma_L,1)$. Therefore, we know that there exists a threshold $x_H^{sep}\in(0,1)$ such that if $x\leq x_H^{sep}$, we have $\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_H,x)\leq\Pi_s(\sigma_L,1)+(1-\alpha)T(\sigma_L,1)$. In this sense, Φ actually captures the region of $x\leq x_H^{sep}$. Intuitively, in the separating equilibrium, the high-type seller will select x_H^{sep} as the optimal solution. To meet condition (C8), we need to ensure that the high-type seller has no incentive to deviate. Because $\frac{\partial^2[\Pi_s(\sigma_i,x)+(1-\alpha)T(\sigma_j,x)]}{\partial\sigma_i\partial x}<0$, suppose that $\dot{x}>x_H^{sep}$ is the off-equilibrium solution for the high-type seller; then, $\Pi_s(\sigma_i,\dot{x})+(1-\alpha)T(\sigma_j,\dot{x})-\left[\Pi_s(\sigma_i,x_H^{sep})+(1-\alpha)T(\sigma_j,x_H^{sep})\right]$ is decreasing in σ_i. Hence, $\Pi_s(\sigma_H,\dot{x})+(1-\alpha)T(\sigma_j,\dot{x})-\left[\Pi_s(\sigma_H,x_H^{sep})+(1-\alpha)T(\sigma_j,x_H^{sep})\right]<\Pi_s(\sigma_L,\dot{x})+(1-\alpha)T(\sigma_j,\dot{x})-\left[\Pi_s(\sigma_L,x_H^{sep})+(1-\alpha)T(\sigma_j,x_H^{sep})\right]$. We let $j=L$, $\Pi_s(\sigma_H,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})-\left[\Pi_s(\sigma_H,x_H^{sep})+(1-\alpha)T(\sigma_L,x_H^{sep})\right]<\Pi_s(\sigma_L,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})-\left[\Pi_s(\sigma_L,x_H^{sep})+(1-\alpha)T(\sigma_L,x_H^{sep})\right]$. We simultaneously add the expression $-(1-\alpha)T(\sigma_H,x_H^{sep})+(1-\alpha)T(\sigma_L,x_H^{sep})$ on each side. We then obtain $\Pi_s(\sigma_H,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})-\left[\Pi_s(\sigma_H,x_H^{sep})+(1-\alpha)T(\sigma_H,x_H^{sep})\right]<\Pi_s(\sigma_L,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})-\left[\Pi_s(\sigma_L,x_H^{sep})+(1-\alpha)T(\sigma_H,x_H^{sep})\right]$. We already know that $\Pi_s(\sigma_L,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})-\left[\Pi_s(\sigma_L,x_H^{sep})+(1-\alpha)T(\sigma_H,x_H^{sep})\right]<0$. Hence, $\Pi_s(\sigma_H,\dot{x})+(1-\alpha)T(\sigma_L,\dot{x})<\left[\Pi_s(\sigma_H,x_H^{sep})+(1-\alpha)T(\sigma_H,x_H^{sep})\right]$ implies that the high-type seller does not deviate from the separating equilibrium. □

With free samples. In this case, the seller will price the data at the buyer's willingness to pay, which is $T(\sigma,x_1+x_2)-T(\sigma,x_1)$.

In the complete information setting, the seller's optimal data selling strategy is given by the following proposition.

Proposition BII-3. *With complete information, given the sample data volume* x_1*, the seller sells all the remaining data (i.e.,* $x_2=1-x_1$*) at the price* $p_d=T(\sigma,1)-T(\sigma,x_1)$.

Proof of Proposition BII-3. When the data are sold, the seller's profit is given by $\Pi_s(\sigma, x_1 + x_2) + (1 - \alpha)T(\sigma, x_1 + x_2) - (1 - \alpha)T(\sigma, x_1)$. For any x_1, because we know that $\Pi_s(\sigma, x_1 + x_2) + (1 - \alpha)T(\sigma, x_1 + x_2)$ increases in x_2, the seller will sell all the remaining data. □

Proposition BII-4. *With complete information, the seller does not offer free samples (i.e., $x_1^* = 0$).*

Proof of Proposition BII-4. From the above results, we know that the seller maximizes $\Pi_s(\sigma, 1) + (1 - \alpha)T(\sigma, 1) - (1 - \alpha)T(\sigma, x_1)$ with respect to $0 < x_1 < 1$. Because $T(\sigma, x_1)$ increases in x_1, the seller will not provide free samples. □

In the incomplete information setting, the low-type seller has an incentive to mimic the high-type seller, leading to a signaling game.

Proposition BII-5. *In the separating equilibrium with free samples, the high-type seller offers a volume $x_{1H}^{sep} \in (0,1)$ of free samples to the buyer, whereas the low-type seller does not offer any free samples (i.e., $x_{1L}^{sep} = 0$).*

Proof of Proposition BII-5. The above separating equilibrium can be derived from the following conditions:

$$\begin{cases} \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_L, 1) + (1 - \alpha)T(\sigma_L, 1) - (1 - \alpha)T(\sigma_L, x_1)}_{Low-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_L, 1) + (1 - \alpha)T(\sigma_H, 1) - (1 - \alpha)T(\sigma_H, x_1)}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C9)} \\ \underbrace{max_{x_1 \in \Phi} \Pi_s(\sigma_H, 1) + (1 - \alpha)T(\sigma_H, 1) - (1 - \alpha)T(\sigma_H, x_1)}_{High-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1 \notin \Phi} \Pi_s(\sigma_H, 1) + (1 - \alpha)T(\sigma_L, 1) - (1 - \alpha)T(\sigma_L, x_1)}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C10)} \end{cases}$$

For the low-type seller, his profit without mimicking is given by $\Pi_s(\sigma_L, 1) + (1 - \alpha)T(\sigma_L, 1)$, which is equal to his first-best profit. To ensure the condition (C9), we need $\Pi_s(\sigma_L, 1) + (1 - \alpha)T(\sigma_H, 1) - (1 - \alpha)T(\sigma_H, x_1) \leq \Pi_s(\sigma_L, 1) + (1 - \alpha)T(\sigma_L, 1)$ so that $T(\sigma_H, x_1) \geq T(\sigma_H, 1) - T(\sigma_L, 1)$. Because $T(\sigma_H, x_1)$ is increasing in x_1, there exists a threshold x_{1H}^{sep} such that Φ captures a region of $x_{1H}^{sep} \leq x_1 \leq 1$. Because $\Pi_s(\sigma_H, 1) + (1 - \alpha)T(\sigma_H, 1) - (1 - \alpha)T(\sigma_H, x_1)$ is decreasing in x_1, the high-type seller will select the volume of free samples at $x_1 = x_{1H}^{sep}$, where condition (C10) can be satisfied. Through a similar approach as we adopt in the Proof of Proposition BII-2, the above separating equilibrium can be sustained. This concludes the proof. □

Impact of free samples on data evaluation. Next, we compare the results regarding data evaluation with and without free samples to uncover the strategic impacts of free samples.

Proposition BII-6. *Comparing the separating equilibria with and without free samples yields the following strategic effects of free samples if data evaluation is achieved:*

(i) free samples incentivize a higher volume of data selling (i.e., $1-x_{1H}^{sep}>x_{H}^{sep}$*),*

(ii) free samples increase the monetary payment for data (i.e., $(1-\alpha)T(\sigma_H,1)-(1-\alpha)T\left(\sigma_H,x_{1H}^{sep}\right)>(1-\alpha)T(\sigma_H,x_H^{sep})$*), and*

(iii) free samples hurt the high-type seller (i.e., $\Pi_{sH}^{F,sep}<\Pi_{sH}^{N,sep}$*).*

Proof of Proposition BII-6. We first prove that $(1-\alpha)T(\sigma_H,1)-(1-\alpha)T\left(\sigma_H,x_{1H}^{sep}\right)>(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)$. We know that x_H^{sep} is derived by letting $\Pi_s\left(\sigma_L,x_H^{sep}\right)+(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)=\Pi_s(\sigma_L,1)+(1-\alpha)T(\sigma_L,1)$. Because $\Pi_s\left(\sigma_L,x_H^{sep}\right)>\Pi_s(\sigma_L,1)$ for any $0<x_H^{sep}<1$, $(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)<(1-\alpha)T(\sigma_L,1)$. Proposition BII-5 yields that the monetary payment for data in the case with free samples is $T(\sigma_H,1)-T\left(\sigma_H,x_{1H}^{sep}\right)=T(\sigma_L,1)$. Therefore, we show that the monetary payment for data in the case without free samples is lower, i.e., $(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)<(1-\alpha)T(\sigma_H,1)-(1-\alpha)T\left(\sigma_H,x_{1H}^{sep}\right)$. For the data selling volumes, the above result implies that $T\left(\sigma_H,x_H^{sep}\right)-T(\sigma_H,0)<T(\sigma_H,1)-T\left(\sigma_H,x_{1H}^{sep}\right)$. Following the same approach as in the Proof of Proposition 6, we can prove that $x_H^{sep}<1-x_{1H}^{sep}$. For the high-type seller's profit, $\Pi_{sH}^{N,sep}-\Pi_{sH}^{F,sep}=\Pi_s\left(\sigma_H,x_H^{sep}\right)+(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)-\left[\Pi_s(\sigma_H,1)+(1-\alpha)T(\sigma_H,1)-(1-\alpha)T\left(\sigma_H,x_{1H}^{sep}\right)\right]=\Pi_s\left(\sigma_H,x_H^{sep}\right)-\Pi_s(\sigma_H,1)-(1-\alpha)[\Pi_s\left(\sigma_L,x_H^{sep}\right)-\Pi_s(\sigma_L,1)]$. This result can be obtained by using $T\left(\sigma_H,x_{1H}^{sep}\right)=T(\sigma_H,1)-T(\sigma_L,1)$ and $\Pi_s\left(\sigma_L,x_H^{sep}\right)+(1-\alpha)T\left(\sigma_H,x_H^{sep}\right)=\Pi_s(\sigma_L,1)+(1-\alpha)T(\sigma_L,1)$. Because $\frac{\partial[\Pi_s(\sigma,x_H^{sep})-\Pi_s(\sigma,1)]}{\partial\sigma}>0$, we have $\Pi_s\left(\sigma_H,x_H^{sep}\right)-\Pi_s(\sigma_H,1)-\left[\Pi_s\left(\sigma_L,x_H^{sep}\right)-\Pi_s(\sigma_L,1)\right]>0$. Hence, $\Pi_{sH}^{N,sep}>\Pi_{sH}^{F,sep}$. This concludes the proof. □

Strategic impact of data evaluation. In this section, we examine the scenario in which the buyer is unable to evaluate the data during the purchase (i.e., a pooling equilibrium arises).

Proposition BII-7. *In the pooling equilibrium (i.e., without data evaluation), both types of sellers set the same volume of data,* $x^{pool}\in[0,1]$*, at the price* $\beta T(\sigma_L,x^{pool})+(1-\beta)T(\sigma_H,x^{pool})$*.*

Proof of Proposition BII-7. To sustain the above pooling equilibrium, we need the following conditions to hold:

$$\begin{cases}\underbrace{\Pi_s(\sigma_L,x^{pool})+(1-\alpha)[\beta T(\sigma_L,x^{pool})+(1-\beta)T(\sigma_H,x^{pool})]}_{Low-type\ seller's\ profit\ under\ pooling}\geq\\ \underbrace{max_{x\neq x^{pool}}\Pi_s(\sigma_L,x)+(1-\alpha)T(\sigma_L,x)}_{Low-type\ seller's\ profit\ under\ deviating},\text{(C11)}\\ \underbrace{\Pi_s(\sigma_H,x^{pool})+(1-\alpha)[\beta T(\sigma_L,x^{pool})+(1-\beta)T(\sigma_H,x^{pool})]}_{High-type\ seller's\ profit\ under\ pooling}\geq\\ \underbrace{max_{x\neq x^{pool}}\Pi_s(\sigma_H,x)+(1-\alpha)T(\sigma_L,x)}_{High-type\ seller's\ profit\ under\ deviating}.\text{(C12)}\end{cases}$$

For the high-type seller, because $T(\sigma_H,x)>T(\sigma_L,x)$, the maximum of $\Pi_s(\sigma_H,x)+(1-\alpha)[\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)]$ is always greater than the maximum of $\Pi_s(\sigma_H,x)+(1-\alpha)T(\sigma_L,x)$.

Hence, condition (C11) holds. For the low-type seller, we will show later (in the Proof of Proposition BII-8) that if the pooling equilibrium outperforms the separating one, the low-type seller will not deviate (i.e., condition (C12) will hold). This concludes the proof. □

Proposition BII-8. *There exists a threshold, $\bar{\beta} \in (0,1)$, such that if $\beta \leq \bar{\beta}$, the pooling equilibrium is selected, and data evaluation hurts both types of sellers, i.e., $\Pi_{sH}^{N,sep} \leq \Pi_{sH}^{N,pool}$ and $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$; otherwise, the separating equilibrium is selected, and data evaluation benefits the high-type seller but hurts the low-type seller, i.e., $\Pi_{sH}^{N,sep} > \Pi_{sH}^{N,pool}$ and $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$.*

Proof of Proposition BII-8. Applying the envelope theorem, we obtain that $\frac{d\Pi_{sH}^{N,pool}}{d\beta} = \frac{\partial \Pi_s(\sigma_H,x)+(1-\alpha)[\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)]}{\partial \beta}|_{x=x^{pool}} = (1-\alpha)[T(\sigma_L, x^{pool}) - T(\sigma_H, x^{pool})] < 0$. Hence, the high-type seller's profit under the pooling equilibrium decreases in β. However, her profit under the separating equilibrium is irrelevant with respect to β . $\Pi_{sH}^{N,pool}|_{\beta=0} = \max_{0<x<1} \Pi_s(\sigma_H, x) + (1-\alpha)T(\sigma_H, x)$, while her profit under the separating equilibrium is derived as $\Pi_{sH}^{N,sep} = \max_{x\in\Phi} \Pi_s(\sigma_H, x) + (1-\alpha)T(\sigma_H, x)$. Hence, in this case, $\Pi_{sH}^{N,pool} > \Pi_{sH}^{N,sep}$ because region Φ belongs to $0 < x < 1$. Moreover, $\Pi_{sH}^{N,pool}|_{\beta=1} = \max_{0<x<1} \Pi_s(\sigma_H, x) + (1-\alpha)T(\sigma_L, x)$, while the high-type seller's profit under the separating equilibrium is given by $\Pi_{sH}^{N,sep} = \Pi_s(\sigma_H, x_H^{sep}) + (1-\alpha)T(\sigma_H, x_H^{sep})$. From condition (C8) in the Proof of Proposition BII-2, we have $\Pi_{sH}^{N,pool} < \Pi_{sH}^{N,sep}$ in this case. Therefore, there exists a threshold $\bar{\beta}$ such that if $\beta \leq \bar{\beta}$, we have $\Pi_{sH}^{N,sep} \leq \Pi_{sH}^{N,pool}$; otherwise, we have $\Pi_{sH}^{N,sep} > \Pi_{sH}^{N,pool}$. To sustain the above pooling equilibrium, we need to prove that the low-type seller will not deviate from the pooling equilibrium at least if it outperforms the separating equilibrium. The low-type seller's profit that stays with the pooling equilibrium is given by $\Pi_s(\sigma_L, x^{pool}) + (1-\alpha)[\beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})]$, and the profit that deviates is given by $\pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1)$. Because $\Pi_s(\sigma_H, x^{pool}) + (1-\alpha)[\beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})] \geq \Pi_s(\sigma_H, x_H^{sep}) + (1-\alpha)T(\sigma_H, x_H^{sep})$, the low-type seller's profit under the pooling equilibrium has the following property: $\Pi_s(\sigma_L, x^{pool}) + (1-\alpha)[\beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})] \geq \Pi_s(\sigma_H, x_H^{sep}) + (1-\alpha)T(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_H, x^{pool}) + \Pi_s(\sigma_L, x^{pool})$. Moreover, from the Proof of Proposition BII-2, we have that $\Pi_s(\sigma_L, x_H^{sep}) + (1-\alpha)T(\sigma_H, x_H^{sep}) = \Pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1)$. We then obtain $\Pi_s(\sigma_H, x_H^{sep}) + (1-\alpha)T(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_H, x^{pool}) + \Pi_s(\sigma_L, x^{pool}) = \Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) - [\Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})] + \Pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1)$. We have already shown that $\frac{\partial^2 \Pi_s(\sigma_i,x)}{\partial \sigma_i \partial x} < 0$; hence, $\frac{\partial[\Pi_s(\sigma_H,x)-\Pi_s(\sigma_L,x)]}{\partial x} < 0$. Because $x_H^{sep} < x^{pool}$, we have $\Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) > \Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})$, which indicates that $\Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) - [\Pi_s(\sigma_H, x^{pool}) - \Pi_s(\sigma_L, x^{pool})] + \Pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1) > \Pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1)$. Therefore, we have $\Pi_s(\sigma_L, x^{pool}) + (1-\alpha)[\beta T(\sigma_L, x^{pool}) + (1-$

$\beta)T(\sigma_H, x^{pool})]$>$\Pi_s(\sigma_L, 1) + (1-\alpha)T(\sigma_L, 1)$. Hence, the low-type seller does not deviate from the pooling equilibrium. This concludes the proof. □

In the following analysis, we aim to elucidate the platform's preference for free samples as well as data evaluation.

Proposition BII-9. *The data platform always prefers the option of free samples when data evaluation is achieved.*

Proof of Proposition BII-9. Because of the commission-based mechanism, the platform's profit without free samples is given by $\alpha[(1-\beta)T(\sigma_H, x_H^{sep}) + \beta T(\sigma_L, 1)]$, while its profit with free samples is given by $\alpha[(1-\beta)[T(\sigma_H, 1) - T(\sigma_H, x_{1H}^{sep})] + \beta T(\sigma_L, 1)]$. The results in Proposition BII-6 imply that the platform prefers the scenario with samples. □

Proposition BII-10. *There exists a threshold,* $\hat{\beta} \in (0,1)$, *such that if* $\beta \le \hat{\beta}$, *the platform prefers the scenario without data evaluation.*

Proof of Proposition BII-10. The platform's profit with data evaluation is given by $\Pi_P^{N,sep} = \alpha[(1-\beta)T(\sigma_H, x_H^{sep}) + \beta T(\sigma_L, 1)]$. In addition, its profit without data evaluation is given by $\Pi_P^{N,pool} = \alpha[\beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})]$. $\Delta \equiv \Pi_P^{N,pool} - \Pi_P^{N,sep} = \alpha[\beta[T(\sigma_L, x^{pool}) - T(\sigma_L, 1)] + (1-\beta)[T(\sigma_H, x^{pool}) - T(\sigma_H, x_H^{sep})]]$. $\frac{\partial \Delta}{\partial \beta} = T(\sigma_L, x^{pool}) - T(\sigma_H, x^{pool}) + T(\sigma_H, x_H^{sep}) - T(\sigma_L, 1) + \beta * B7$, where $B7 = \frac{\partial x^{pool}}{\partial \beta}[\frac{\partial T(\sigma_L, x^{pool})}{\partial x^{pool}} - \frac{\partial T(\sigma_H, x^{pool})}{\partial x^{pool}}]$. It is easy to show that $T(\sigma_L, x^{pool}) - T(\sigma_H, x^{pool}) + T(\sigma_H, x_H^{sep}) - T(\sigma_L, 1) < 0$ and $\frac{\partial T(\sigma_L, x^{pool})}{\partial x^{pool}} < \frac{\partial T(\sigma_H, x^{pool})}{\partial x^{pool}}$. The reason is that $\frac{\partial T(\sigma, x)}{\partial x}$ increases in σ. Although we cannot verify the sign of $\frac{\partial x^{pool}}{\partial \beta}$, we know that if it is positive, $\frac{\partial \Delta}{\partial \beta} < 0$. If $\frac{\partial x^{pool}}{\partial \beta} < 0$, we conclude that there at least exists a threshold $\bar{\bar{\beta}}$ such that if $\beta < \bar{\bar{\beta}}$, $\frac{\partial \Delta}{\partial \beta} < 0$.

If $\beta = 0$, $\Delta = \alpha\left[T(\sigma_H, x^{pool}) - T(\sigma_H, x_H^{sep})\right] > 0$ because if $\beta < \bar{\beta}$, the pooling equilibrium is selected and $x^{pool} > x_H^{sep}$. Therefore, if $\beta < \hat{\beta} \equiv \min\{\bar{\beta}, \bar{\bar{\beta}}\}$, we can have $\Delta > 0$, implying that the platform prefers the scenario without data evaluation. □

In summary, our results indicate that the main findings remain qualitatively the same. Moreover, we derive new insights into the platform's preference regarding the offer of free samples and data evaluation. The analysis reveals that the platform consistently favors the scenario with free samples because of the higher transfer payment if achieving data evaluation. This finding is consistent with empirical evidence that platforms often encourage the provision of free samples, while sellers tend to resist doing so. Furthermore, similar to the high-type seller, the platform prefers the scenario without data evaluation if the probability of encountering a low-type seller is relatively low. This finding suggests that improving the efficiency of data trading through data evaluation may not always benefit the platform, thereby offering practical managerial implications.

3. Part III: Complementary products

In this part, we extend our analysis to examine the case in which the products of the data seller and the data buyer are complements. The corresponding inverse demand function is given by $p_i = a + \theta - q_i - \gamma q_j$, $i \neq j$, i, $j = \{s, b\}$, where $\gamma \in (-1,0)$ characterizes the extent of complementarity between products. The other settings remain unchanged.

Without free samples. The buyer's willingness to pay for the data is given by $T(\sigma, x)$, and this is the level at which the seller will price the data.

In the complete information setting, the seller maximizes $\Pi_s(\sigma, x) + T(\sigma, x)$ with respect to $0 < x < 1$, leading to the following result.

Proposition BIII-1. *With complete information, the seller sells all the data to the buyer (i.e., $x^* = 1$) at the price $p_d = T(\sigma, 1)$.*

Proof of Proposition BIII-1. It is easy to verify that both terms $\Pi_s(\sigma, x)$ and $T(\sigma, x)$ are increasing in x. Therefore, the seller's profit increases in x, leading to the result that the seller sells all the data. □

In the incomplete information setting, because $T(\sigma, x)$ increases in σ, the low-type seller still has an incentive to mimic the high-type seller.

Proposition BIII-2. *If free samples are not offered, separating equilibrium referring to data evaluation cannot be sustained.*

Proof of Proposition BIII-2. The feasible separating equilibrium can be derived from the following conditions:

$$\begin{cases} \underbrace{max_{x \notin \Phi} \Pi_s(\sigma_L, x) + T(\sigma_L, x)}_{Low-type\ seller's\ first\ best\ profit} \geq \underbrace{max_{x \in \Phi} \Pi_s(\sigma_L, x) + T(\sigma_H, x)}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C1)} \\ \underbrace{max_{x \in \Phi} \Pi_s(\sigma_H, x) + T(\sigma_H, x)}_{High-type\ seller's\ profit\ under\ separating} \geq \underbrace{max_{x \notin \Phi} \Pi_s(\sigma_H, x) + T(\sigma_L, x)}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C2)} \end{cases}$$

It is easy to verify that the low-type seller's profit under mimicking is increasing in x. Additionally, if $x = 0$, the low-type seller's profit under the first-best solution is higher, whereas if $x = 1$, his profit under mimicking is higher. Therefore, there exists x_H^{sep} below which the low-type seller does not mimic the high-type seller. Intuitively, the high-type seller will select x_H^{sep} as the optimal solution under the separating equilibrium by maximizing profit. Finally, to sustain the separating equilibrium, we need to ensure that the high-type seller has no incentive to deviate. The high-type seller's profit with deviation is given by $max_{x \notin \Phi} \Pi_s(\sigma_H, x) + T(\sigma_L, x) = \Pi_s(\sigma_H, 1) + T(\sigma_L, 1)$, and the high-type seller's profit under the separating equilibrium is $\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep})$. From condition C1, we know that $\Pi_s(\sigma_L, x_H^{sep}) + T(\sigma_H, x_H^{sep}) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Hence, $T(\sigma_H, x_H^{sep}) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1) - \Pi_s(\sigma_L, x_H^{sep})$. Therefore, $\Pi_s(\sigma_H, x_H^{sep}) + T(\sigma_H, x_H^{sep}) - [\Pi_s(\sigma_H, 1) + T(\sigma_L, 1)] = \Pi_s(\sigma_H, x_H^{sep}) - \Pi_s(\sigma_L, x_H^{sep}) -$

$[\Pi_s(\sigma_H,1)-\Pi_s(\sigma_L,1)]$. Because $\gamma\in(-1,0)$, one can easily verify that $\frac{\partial^2[\Pi_s(\sigma_i,x)]}{\partial\sigma_i\partial x}>0$, which implies that $\Pi_s(\sigma_H,x)-\Pi_s(\sigma_L,x)$ increases in x. Since $x_H^{sep}<1$, we obtain $\Pi_s(\sigma_H,x_H^{sep})-\Pi_s(\sigma_L,x_H^{sep})<\Pi_s(\sigma_H,1)-\Pi_s(\sigma_L,1)$. Therefore, the high-type seller's profit in the case of deviation is even greater than the profit under the separating equilibrium; hence, the separating equilibrium cannot be sustained. □

The above analysis reveals that in the scenario with complementary products, data evaluation cannot be achieved without free samples.

With free samples. For any given volume of free samples, the buyer's willingness to pay for data is $T(\sigma,x_1+x_2)-T(\sigma,x_1)$.

In the complete information setting, the seller's optimal data selling strategy is given by the following proposition.

Proposition BIII-3. *With complete information, given the sample data volume* x_1*, the seller sells all the remaining data (i.e.,* $x_2=1-x_1$*) at the price* $p_d=T(\sigma,1)-T(\sigma,x_1)$*.*

Proof of Proposition BIII-3. Because $\Pi_s(\sigma,x_1+x_2)+T(\sigma,x_1+x_2)-T(\sigma,x_1)$ is increasing in x_2 for any given x_1, the seller will sell all the remaining data. □

Proposition BIII-4. *With complete information, the seller does not offer free samples (i.e.,* ${x_1}^*=0$*).*

Proof of Proposition BIII-4. Because $T(\sigma,x_1)$ increases in x_1, the seller will not provide free samples when maximizing $\Pi_s(\sigma,1)+T(\sigma,1)-T(\sigma,x_1)$. □

In the incomplete information setting, the separating equilibrium outcome is characterized as follows:

Proposition BIII-5. *In the separating equilibrium with free samples, the high-type seller offers a volume* $x_{1H}^{sep}\in(0,1)$ *of free samples to the buyer, whereas the low-type seller does not offer any free samples (i.e.,* $x_{1L}^{sep}=0$*).*

Proof of Proposition BIII-5. The corresponding results can be proven by adopting the same approach as in the Proof of Proposition 5. Thus, we omit this proof. □

Strategic impact of data evaluation. Next, we consider the results for the pooling equilibrium and then find the final equilibrium outcome.

Proposition BIII-6. *In the pooling equilibrium (i.e., without data evaluation), both types of sellers set the same volume of data,* $x^{pool}=1$*, at the price* $\beta T(\sigma_L,1)+(1-\beta)T(\sigma_H,1)$*.*

Proof of Proposition BIII-6. Note that in any pooling equilibrium, sellers of both types will not provide free samples (see part V in Online Appendix B). The corresponding pooling equilibrium outcome is given by $x^{pool}=\text{argmax}\,\Pi_s(\sigma_H,x)+\beta T(\sigma_L,x)+(1-\beta)T(\sigma_H,x)=1$. The reason is that the terms $\Pi_s(\sigma_H,x)$,

$T(\sigma_L, x)$ and $T(\sigma_H, x)$ increase with x. To sustain this pooling equilibrium, we need the following conditions:

$$\begin{cases} \underbrace{\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})}_{Low-type\ seller's\ profit\ under\ pooling} \geq \underbrace{max_{x \neq x^{pool}} \Pi_s(\sigma_L, x) + T(\sigma_L, x)}_{Low-type\ seller's\ profit\ under\ deviating}, & (C5) \\ \underbrace{\Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})}_{High-type\ seller's\ profit\ under\ pooling} \geq \underbrace{max_{x \neq x^{pool}} \Pi_s(\sigma_H, x) + T(\sigma_L, x)}_{High-type\ seller's\ profit\ under\ deviating}. & (C6) \end{cases}$$

For the low-type seller, the profit under the pooling equilibrium is $\Pi_s(\sigma_L, 1) + \beta T(\sigma_L, 1) + (1 - \beta)T(\sigma_H, 1)$, which is greater than his profit under the first-best solution. Hence, the low-type seller does not deviate. For the high-type seller, the profit under the pooling equilibrium is $\Pi_s(\sigma_H, 1) + \beta T(\sigma_L, 1) + (1 - \beta)T(\sigma_H, 1)$, which is always greater than her profit in the case of deviating because the high-type seller can sell more data and earn a higher price under the pooling equilibrium than in the case of deviating. Therefore, the above pooling equilibrium can be sustained. □

Although Proposition BIII-5 shows that data evaluation can be achieved with free samples, it is dominated by the pooling equilibrium.

Proposition BIII-7. *The pooling equilibrium always dominates the separating equilibrium so that the pooling equilibrium is the unique outcome.*

Proof of Proposition BIII-7. We compare the high-type seller's profits under the separating equilibrium characterized in Proposition BIII-5 and the pooling equilibrium characterized in Proposition BIII-6. Her profit under the separating equilibrium is $\Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T(\sigma_H, x_{1H}^{sep})$, while her profit under the pooling equilibrium is $\Pi_s(\sigma_H, 1) + \beta T(\sigma_L, 1) + (1 - \beta)T(\sigma_H, 1)$, which is decreasing in β. If $\beta = 0$, her profit under the pooling equilibrium is greater. If $\beta = 1$, her profit under the pooling equilibrium reduces to $\Pi_s(\sigma_H, 1) + T(\sigma_L, 1)$, which is equal to her profit under the separating equilibrium $\Pi_s(\sigma_H, 1) + T(\sigma_H, 1) - T(\sigma_H, x_{1H}^{sep})$. The reason is that $T(\sigma_H, 1) - T(\sigma_H, x_{1H}^{sep}) = T(\sigma_L, 1)$. Therefore, the pooling equilibrium will always be selected by the LMSE. □

In summary, our analysis reveals that the results differ from those in the main model, where the products are substitutes. First, in the scenario without free samples, the separating equilibrium referring to data evaluation can no longer be sustained. In the absence of market competition, the benefit of the seller's information advantage over the buyer is nullified, giving the seller an incentive to provide the buyer with more data. Consequently, separating by selling less data will lead to higher signaling costs for the high-type seller in facilitating data evaluation, making the separating equilibrium unsustainable. Second, in the scenario with free samples, although the separating equilibrium exists, it is dominated by the pooling equilibrium. The intuition is that selling more data in the pooling equilibrium enhances the value of data monetization, which yields a higher payoff for the high-type seller than offering free samples in the separating equilibrium does. Therefore, if the products are complements, providing free samples becomes the only effective means of achieving data evaluation. However, data evaluation through signaling (i.e., the separating equilibrium) cannot be the final equilibrium outcome regardless of whether free samples are provided. This extended analysis underscores that competition in the product market is the key to exploring the trade-off underlying free sample provision and data evaluation with signaling, which constitutes the core focus of this paper.

4. Part IV: Multiple signaling instruments

In this section, we analyze the scenario with multiple signaling instruments in which the volume of free samples, the volume of data sold and the data selling price jointly serve as the signaling instruments. We first investigate the case without free samples and then move to the case with free samples.

Without free samples. Note that the findings under both complete and incomplete information settings remain unchanged as the volume of free samples becomes irrelevant. These findings are listed below as Propositions IV-1 and IV-2, which correspond to Propositions 1 and 2 in the main model, respectively.

Proposition IV-1. *With complete information, the seller sells all the data to the buyer (i.e.,* $x^* = 1$*) at the price* $p_d = T(\sigma, 1)$.

Proof of Proposition IV-1. The proof is identical to that of Proposition 1 and is therefore omitted here. □

In the incomplete information setting, we find that data evaluation can still be achieved with the high-type seller selling less data than the low-type seller does.

Proposition IV-2. *Without free samples, in the separating equilibrium, the high-type seller sells a portion with volume* $x_H^{sep} \in (0,1)$ *of the data at the price* $T(\sigma_H, x_H^{sep})$*, whereas the low-type seller adheres to his first-best solution by selling all the data (i.e.,* $x_L^{sep} = 1$*) at the price* $T(\sigma_L, 1)$.

Proof of Proposition IV-2. The proof is identical to that of Proposition 2 and is therefore omitted here. □

With free samples. Note that in the main model with free samples, only the volume of free samples x_1 serves as the signaling instrument. The reason is that the seller makes decisions on sample volume and sales volume sequentially. This allows the buyer to update belief immediately after observing the sample volume but before the final sales volume is revealed. Consequently, once the buyer's belief is updated, the seller's optimal strategy is to sell all remaining data, i.e., $x_2 = 1 - x_1$; therefore, only the volume of the free samples can help signaling. In this section, we relax this assumption of a single-dimensional signaling instrument and examine the case where the volume of free samples (x_1), the volume of data sold (x_2), and the data selling price (p_d) jointly serve as signaling instruments (i.e., multidimensional signaling).

Note that in this case, the seller will price the data at the buyer's maximum willingness to pay, which is determined by the volumes of free samples (x_1) and data sold (x_2); therefore, p_d has no signaling role (see Proof of Proposition IV-4), as in our main paper. One notable difference from our main analysis is that the seller maximizes the profit in a two-dimensional parameter region rather than a one-dimensional parameter range. The other settings are unchanged.

First, we examine the case with complete information. Proposition IV-3 corresponds to Propositions 3 and 4 in the main model, which shows that the multiple signaling instruments do not change the sellers' first best profits in a complete information setting.

Proposition IV-3. *With complete information, the seller does not offer free samples (i.e.,* ${x_1}^* = 0$*) and sells all the data (i.e.,* ${x_2}^* = 1$*).*

Proof of Proposition IV-3. The seller maximizes $\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2) - T(\sigma, x_1)$. It is easy to show that $\Pi_s(\sigma, x_1 + x_2) + T(\sigma, x_1 + x_2)$ is increasing in x_2 for any fixed x_1. Hence, the seller will sell all the remaining data; that is, $x_2 = 1 - x_1$. Accordingly, because $T(\sigma, x_1)$ increases in x_1, the seller will not provide free samples. □

Second, we focus on the case with incomplete information. For data evaluation, the corresponding separating equilibrium outcome is characterized as follows:

Proposition IV-4. *In the separating equilibrium, both types of sellers will not provide free samples. Accordingly, the high-type seller sells a portion of the data with volume* $x_H^{sep} \in (0,1)$ *at the price* $T(\sigma_H, x_H^{sep})$*, whereas the low-type seller sells all the data (i.e.,* $x_L^{sep} = 1$*) at the price* $T(\sigma_L, 1)$.

Proof of Proposition IV-4. When the volume of free samples (x_1), the volume of data sold (x_2), and the data selling price (p_d) jointly serve as signaling instruments, the corresponding conditions leading to the separating equilibrium are as follows:

$$\begin{cases} \underbrace{max_{x_1,x_2,p_d \notin \Phi} \Pi_s(\sigma_L, x_1 + x_2) + p_d}_{Low-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1,x_2,p_d \in \Phi} \Pi_s(\sigma_L, x_1 + x_2) + p_d}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C13)} \\ \underbrace{max_{x_1,x_2,p_d \in \Phi} \Pi_s(\sigma_H, x_1 + x_2) + p_d}_{High-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1,x_2,p_d \notin \Phi} \Pi_s(\sigma_H, x_1 + x_2) + p_d}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C14)} \end{cases}$$

Suppose that the separating equilibrium is supported by the belief that if $\{x_1, x_2, p_d\} \in \Phi$, the buyer believes that the market is of a high type; otherwise, the buyer believes that the market is of a low type. In each profit maximization problem (staying with separating equilibrium or deviating), the sellers' profits increase with the data selling price. Consequently, the seller will price the data at the buyer's maximum willingness to pay, which hinges on the volumes of free samples and data sold. Following the logic established in the Proof of Proposition 2, the separating equilibrium is reduced to the following form:

$$\begin{cases} \underbrace{max_{x_1,x_2 \notin \Phi} \Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_L, x_1 + x_2) - T(\sigma_L, x_1)}_{Low-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1,x_2 \in \Phi} \Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1)}_{Low-type\ seller's\ profit\ under\ deviating}, \text{(C13')} \\ \underbrace{max_{x_1,x_2 \in \Phi} \Pi_s(\sigma_H, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1)}_{High-type\ seller's\ profit\ under\ separating} \geq \\ \underbrace{max_{x_1,x_2 \notin \Phi} \Pi_s(\sigma_H, x_1 + x_2) + T(\sigma_L, x_1 + x_2) - T(\sigma_L, x_1)}_{High-type\ seller's\ profit\ under\ deviating}. \text{(C14')} \end{cases}$$

Although the seller simultaneously decides on x_1 and x_2, $\Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1)$ increases in x_2 for any x_1. For condition (C13′), the low-type seller's first-best profit under the separating equilibrium is $\Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. If $x_2 = 0$, we have $\max\Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1) = \Pi_s(\sigma_L, 0) < \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. If $x_2 = 1 - x_1$, $\max\Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1) = \Pi_s(\sigma_L, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1)$ decreases in x_1. Accordingly, if $x_1 = 0$, we have $\Pi_s(\sigma_L, 1) + T(\sigma_H, 1) > \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$, whereas if $x_1 = 1$, we have $\Pi_s(\sigma_L, 1) < \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Therefore, there exists a threshold x_{1H}^{sep}, which can be derived by letting $\Pi_s(\sigma_L, 1) + T(\sigma_H, 1) - T(\sigma_H, x_1) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. If $x_1 \le x_{1H}^{sep}$, the range of x_2 satisfying condition (C13′) is $x_2 \le \bar{x}_2(x_1)$. The result for $\bar{x}_2(x_1)$ can be derived by letting $\Pi_s(\sigma_L, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. It is easy to show that $\bar{x}_2(x_1)$ increases in x_1. Moreover, if $x_1 = 0$, $\bar{x}_2(0) = x_H^{Sep}$ is exactly the optimal solution without free samples in our main paper. If $x_1 \ge x_{1H}^{sep}$, the range of x_2 satisfying condition (C13′) is $x_2 \le 1 - x_1$. Therefore, the derivation leads to $\Phi = \{(x_1, x_2): x_2 \le \bar{x}_2(x_1)$ if $x_1 \le x_{1H}^{Sep}$, and $x_2 \le 1 - x_1$ if $x_1 \ge x_{1H}^{Sep}\}$, which is the feasible optimization region (i.e., Region II in Figure B4).

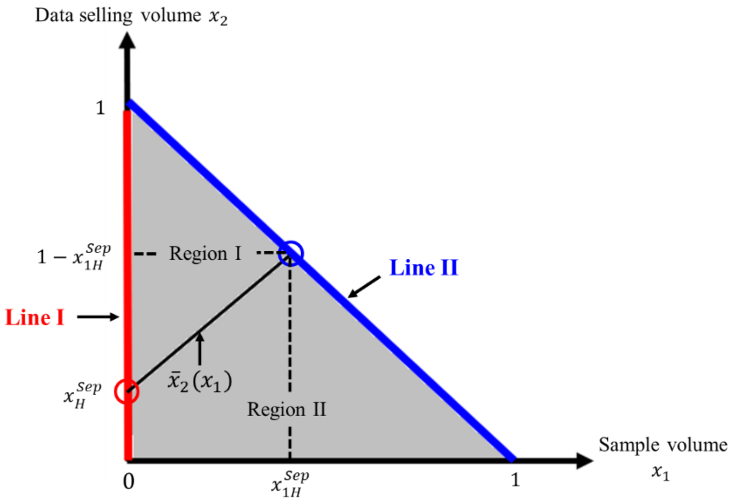

Figure B4. Illustration of the separating equilibrium.

We first consider the case of $x_1 \le x_{1H}^{sep}$. Because $\Pi_s(\sigma_H, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1)$ increases in x_2 for any x_1, the high-type seller chooses $\bar{x}_2(x_1)$ and maximizes $\Pi_s(\sigma_H, x_1 + \bar{x}_2(x_1)) + T(\sigma_H, x_1 + \bar{x}_2(x_1)) - T(\sigma_H, x_1)$ with respect to $x_1 \le x_{1H}^{sep}$. Because $\Pi_s(\sigma_L, x_1 + \bar{x}_2(x_1)) + T(\sigma_H, x_1 + \bar{x}_2(x_1)) - T(\sigma_H, x_1) = \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$, the problem decreases to $\max\Pi_s(\sigma_H, x_1 + \bar{x}_2(x_1)) - \Pi_s(\sigma_L, x_1 + \bar{x}_2(x_1)) + \Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$ with respect to $x_1 \le x_{1H}^{sep}$.

We let $\Pi_s(\sigma_H, x_1 + \bar{x}_2(x_1)) - \Pi_s(\sigma_L, x_1 + \bar{x}_2(x_1)) \equiv B7$ and $\frac{\partial B7}{\partial x_1} = \frac{\partial \Pi_s(\sigma_H, x_1 + \bar{x}_2(x_1))}{\partial x_1 + \bar{x}_2(x_1)} \frac{\partial x_1 + \bar{x}_2(x_1)}{\partial x_1} - \frac{\partial \Pi_s(\sigma_L, x_1 + \bar{x}_2(x_1))}{\partial x_1 + \bar{x}_2(x_1)} \frac{\partial x_1 + \bar{x}_2(x_1)}{\partial x_1} = (1 + \frac{\partial \bar{x}_2(x_1)}{\partial x_1})[\frac{\partial \Pi_s(\sigma_H, x_1 + \bar{x}_2(x_1))}{\partial x_1 + \bar{x}_2(x_1)} - \frac{\partial \Pi_s(\sigma_L, x_1 + \bar{x}_2(x_1))}{\partial x_1 + \bar{x}_2(x_1)}]$. Because $\frac{\partial^2 \Pi_s(\sigma, x)}{\partial \sigma \partial x} < 0$, we have $\frac{\partial \Pi_s(\sigma_H, x_1+\bar{x}_2(x_1))}{\partial x_1+\bar{x}_2(x_1)} - \frac{\partial \Pi_s(\sigma_L, x_1+\bar{x}_2(x_1))}{\partial x_1+\bar{x}_2(x_1)} < 0$. Moreover, $\frac{\partial \bar{x}_2(x_1)}{\partial x_1} > 0$ implies that $\frac{\partial B7}{\partial x_1} < 0$. Hence, in this case, the optimal solution is $x_1 = 0$ and $x_2 = x_H^{Sep}$. Next, we consider the case of $x_1 \geq x_{1H}^{sep}$. Because $\Pi_s(\sigma_H, x_1 + x_2) + T(\sigma_H, x_1 + x_2) - T(\sigma_H, x_1)$ increases in x_2 for any x_1, the optimal data selling volume is $x_2 = 1 - x_1$. It is easy to verify that the optimal volume of free samples is $x_1 = x_{1H}^{Sep}$.

Therefore, we conclude that if both x_1 and x_2 serve as signaling instruments, the interior solution in the parameter region cannot be realized; therefore, only the boundary solution is feasible. Accordingly, only $(0, x_H^{Sep})$ and $\left(x_{1H}^{Sep}, 1 - x_{1H}^{Sep}\right)$ are the feasible optimal solutions in this case.

Note that Line I (i.e., $\{(x_1, x_2): x_1 = 0, 0 \leq x_2 \leq 1\}$) in Figure B4 depicts the optimization range for selecting the separating equilibrium without free samples. As shown in the main model, the point $(0, x_H^{Sep})$ yields the least-cost separating equilibrium. Line II (i.e., $\{(x_1, x_2): 0 \leq x_1 \leq 1, x_2 = 1 - x_1\}$) in Figure B4 illustrates the optimization range for selecting the separating equilibrium with free samples in the main model, and $(x_{1H}^{Sep}, 1 - x_{1H}^{Sep})$ is the corresponding optimal solution. As demonstrated in the main model, the solution $(0, x_H^{Sep})$ yields a strictly better outcome for the high-type seller than does the solution $(x_{1H}^{Sep}, 1 - x_{1H}^{Sep})$. Consequently, the high-type seller will select $(0, x_H^{Sep})$ in equilibrium. □

Recall that in the main paper (see Proposition 5), with a single signaling instrument, the high-type seller *does* offer free samples in the separating equilibrium with selecting $(x_{1H}^{Sep}, 1 - x_{1H}^{Sep})$ as the outcome. In contrast, Proposition IV-4 shows that, with multiple signaling instruments, the high-type seller does *not* offer free samples (i.e., $x_1 = 0$) in the separating equilibrium with selecting $(0, x_H^{Sep})$ as the outcome. The reason is that the multidimensional signaling framework effectively unifies scenarios with and without free samples from the main model, as its two candidate optimal solutions directly correspond to the equilibrium outcomes observed with and without free samples in the baseline model. In other words, the cases examined in the main model—both with and without free samples—represent specific instances of this more general multidimensional signaling framework. Furthermore, Proposition 6 in our main paper illustrates the strategic impact of free samples on the high-type seller's profit. By comparing scenarios with and without free samples, we demonstrated that the high-type seller prefers to not provide free samples under data evaluation (in the separating equilibrium). That is, $(0, x_H^{Sep})$ dominates $\left(x_{1H}^{Sep}, 1 - x_{1H}^{Sep}\right)$ from the perspective of the high-type seller. Therefore, we obtain the equilibrium outcome as shown in Proposition IV-4 under multiple signaling instruments. In summary, the subgame separating equilibrium outcomes with the free-sample option differ between the main model and the extension model, but the final separating equilibrium outcomes under data evaluation (by comparing the equilibrium outcomes with and without the option of free samples) are the same under the two models.

From Propositions IV-2 and IV-4, we have the following remark.

Remark IV-1. *With multiple signaling instruments, the scenarios with and without free samples yield the* *same separating equilibrium outcome, i.e., both types of sellers will not provide free samples under data evaluation.*

Strategic impact of data evaluation. Next, we consider the results for the pooling equilibrium and then find the final equilibrium outcome.

Proposition IV-5. *In the pooling equilibrium (i.e., without data evaluation), both types of sellers set the same data selling volume,* $x^{pool} \in [0,1]$, *at the price* $\beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$.

Proof of Proposition IV-5. The proof is quite straightforward. By adopting a similar vein of thinking in Proof of Proposition 7, for any volumes of free samples and data sold, the high-type seller will charge the price at the buyer's maximum willingness to pay. Consequently, the price has no signaling role in determining the pooling equilibrium. Moreover, in this context, the samples provide no information value for the buyer's belief updating. Instead, providing samples undermines the seller's ability to monetize the data. Therefore, in any sustainable pooling equilibrium, sellers of both types will not provide free samples (i.e., $x_1 = 0$) but will instead monetize them to earn a higher profit. □

Proposition IV-5 corresponds to Proposition 7 in the main paper, which implies that the results without data evaluation (i.e., pooling equilibrium) do *not* change under multiple signaling instruments. The final equilibrium outcome is determined by applying the LMSE refinement and comparing the separating equilibrium (Proposition IV-4) and the pooling equilibrium (Proposition IV-5) from the high-type seller's perspective.

Proposition IV-6. *There exists a threshold* $\bar{\beta} \in (0,1)$ *such that if* $\beta \leq \bar{\beta}$, *the pooling equilibrium is selected, under which data evaluation hurts both types of sellers, i.e.,* $\Pi_{sH}^{N,sep} \leq \Pi_{sH}^{N,pool}$ *and* $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$; *otherwise, the separating equilibrium is selected, under which data evaluation benefits the high-type seller but hurts the low-type seller, i.e.,* $\Pi_{sH}^{N,sep} > \Pi_{sH}^{N,pool}$ *and* $\Pi_{sL}^{N,sep} \leq \Pi_{sL}^{N,pool}$.

Proof of Proposition IV-6. The proof is identical to that of Proposition 8 and is therefore omitted here. □

Proposition IV-6 corresponds to Proposition 8 in the main paper, which demonstrates that even with multiple signaling instruments, the final equilibrium outcome remains unchanged: the high-type seller opts for a pooling equilibrium when the market is likely to exhibit high variance (i.e., $\beta \leq \bar{\beta}$) and a separating equilibrium otherwise. Consequently, the introduction of multiple signaling instruments yields the same final equilibrium outcomes as our main model does, confirming that our findings regarding data evaluation are robust and not contingent on the assumption of a single-dimensional signaling mechanism.

5. Part V: Functional role of free samples

In our main paper, to better explore the fundamental mechanism underlying data evaluation, we focus solely on the setting in which free samples can only play a signaling role. In practice, free samples may also have functional effects and directly convey information to buyers. In this extended model, our primary goal is to verify that the functional role of free samples does not alter our main results.

Note that under the separating equilibrium, the buyer can infer the seller's type perfectly; therefore, the buyer does not need to rely on the functional role of free samples in conveying (partial) information. In other words, in the separating equilibrium, the signaling role of free samples is more effective than their functional role. Consequently, the functional role of free samples is relevant only to the pooling equilibrium. Moreover, as discussed in the main paper, in the pooling equilibrium, the sellers optimally choose not to provide free samples. However, the result is derived under the assumption that free samples have no functional role. In this section, we extend our analysis by considering a pooling equilibrium in which both sellers provide the same positive number of free samples, specifically, $x_{1H} = x_{1L} > 0$. Then, free samples perform a functional role; however, because sellers of both types provide the same quantity, the buyer cannot infer the seller's true type.

To capture the functional role of free samples while ensuring mathematical tractability, we make a simplifying assumption regarding their informational effect. Specifically, we assume that the samples from the high-type seller will update the buyer's prior belief to the extent that the seller is of a low (*resp.* high) type with probability $\beta - \varepsilon_1$ (*resp.* $1 - \beta + \varepsilon_1$). In addition, the samples from the low-type seller will update the buyer's prior belief to the extent that the seller is of a low (*resp.* high) type with probability $\beta + \varepsilon_2$ (*resp.* $1 - \beta - \varepsilon_2$). Accordingly, $\varepsilon_1 \in (0, \beta)$ and $\varepsilon_2 \in (0, 1 - \beta)$ characterize the effectiveness of the free samples' functional role. If $\varepsilon_1 = \beta$ and $\varepsilon_2 = 1 - \beta$, free samples perfectly reflect the seller's type. The values of ε_1 and ε_2 could vary with the volume of free samples; our model's fundamental mechanism requires only that samples from sellers of different types generate dispersion in the buyer's belief. We therefore model them as constants for simplicity and without a loss of generality. The other settings remain unchanged.

Proposition V-1. *Any pooling equilibrium with a positive volume of free samples cannot be sustained. Thus, the functional role of free samples vanishes.*

Proof of Proposition V-1. In this proof, we show that the above pooling equilibrium—where both sellers offer the same positive number of free samples—cannot be sustained. As a result, in the equilibrium considered in our main paper, the functional role of free samples vanishes and thus does not affect the results.

Given that both sellers provide $x_{1H} = x_{1L} \equiv \hat{x}_1$ free samples and sell the same volume of data $x_{2H} = x_{2L} \equiv \hat{x}_2$, the optimal data selling price in the pooling equilibrium is derived from the high-type seller's perspective (i.e., the one who wants to reveal the type). That is,

$$p_d{}^{Pool} = WTP_H \equiv (\beta - \varepsilon_1)[T(\sigma_L, \hat{x}_1 + \hat{x}_2) - T(\sigma_L, \hat{x}_1)] + (1 - \beta + \varepsilon_1)[T(\sigma_H, \hat{x}_1 + \hat{x}_2) - T(\sigma_H, \hat{x}_1)].$$

We let WTP_H denote the buyer's willingness to pay for the high-type seller's data. To sustain the above pooling equilibrium solution, the low-type seller must have an incentive to sustain the pooling equilibrium. However, after the buyer assesses the low-type seller's free samples, the buyer's willingness to pay for data becomes the following:

$$WTP_L \equiv (\beta + \varepsilon_2)[T(\sigma_L, \hat{x}_1 + \hat{x}_2) - T(\sigma_L, \hat{x}_1)] + (1 - \beta - \varepsilon_2)[T(\sigma_H, \hat{x}_1 + \hat{x}_2) - T(\sigma_H, \hat{x}_1)].$$

Comparing the buyer's willingness to pay and the data selling price, we have that $p_d{}^{Pool} - WTP_L = (\varepsilon_1 + \varepsilon_2)[T(\sigma_H, \hat{x}_1 + \hat{x}_2) - T(\sigma_H, \hat{x}_1) - (T(\sigma_L, \hat{x}_1 + \hat{x}_2) - T(\sigma_L, \hat{x}_1))]$. Because $\frac{\partial^2 T(\sigma,x)}{\partial x \partial \sigma} > 0$, we know that $\frac{\partial T(\sigma,\hat{x}_1+\hat{x}_2) - T(\sigma,\hat{x}_1)}{\partial \sigma} > 0$. Since $\sigma_H > \sigma_L$, we can obtain that $T(\sigma_H, \hat{x}_1 + \hat{x}_2) - T(\sigma_H, \hat{x}_1) > T(\sigma_L, \hat{x}_1 + \hat{x}_2) - T(\sigma_L, \hat{x}_1)$, implying that $p_d{}^{Pool} > WTP_L$. That is, after assessing the low-quality free samples, the buyer will not purchase the data. Thus, the low-type seller's profit under this pooling equilibrium is $\Pi_s(\sigma_L, \hat{x}_1)$, which is lower than his first-best profit $\Pi_s(\sigma_L, 1) + T(\sigma_L, 1)$. Hence, the low-type seller prefers to deviate from the pooling equilibrium, which leads to any pooling equilibrium with the provision of a positive volume of free samples becoming unsustainable. Considering this together with our main results in Proposition 7, we conclude that in any sustainable pooling equilibrium, the sellers will not provide free samples. □

In summary, even if we consider both the signaling and functional roles of free samples, our main results do not change. The reason is that in the separating equilibrium, the signaling role of free samples is more effective than their functional role. Moreover, in the pooling equilibrium, the functional role of free samples is to persuade the buyer that the seller is more likely to be of a low (high) type upon receiving samples of low (high) quality. Hence, it will reduce the buyer's willingness to pay for the low-type seller's data after receiving the samples, causing the revised willingness to pay less than the pooling price. Consequently, such a pooling equilibrium is unsustainable because the low-type seller inevitably deviates.

6. Part VI: An alternative equilibrium refinement—the intuitive criterion

In this section, we prove that the pooling equilibrium outcome cannot survive the intuitive criterion but that the separating equilibrium outcome can survive it.

Lemma VI-1. *The pooling equilibrium characterized in Proposition 7 cannot survive the intuitive criterion.*

Proof of Lemma VI-1. There are two steps in implementing the intuitive criterion to refine an equilibrium. Step 1 verifies whether there are any types of senders for which the *highest* profit earned by deviating (i.e., choosing an off-equilibrium strategy) is higher than that earned by keeping the equilibrium strategy. Step 2 of the intuitive criterion checks whether, among all types of senders derived in the first step, there exists a specific type of sender, the profit of which as a result of keeping the equilibrium strategy is still lower than the *lowest* profit earned by deviating (i.e., choosing an off-equilibrium strategy), given that the receiver restricts the belief after observing such a deviation. If such a sender type exists, the equilibrium fails the intuitive criterion.

Step 1. Suppose that $x' \neq x^{pool}$ is an off-equilibrium strategy. Following the first step, we examine whether the two types of sellers' highest profits earned by adopting such an off-equilibrium strategy are greater than those obtained by keeping the equilibrium strategy. By this approach, we can derive the set of off-equilibrium strategies ($\Phi_i, i \in \{H, L\}$), with which the i-type seller will have an incentive to deviate. Note that in our model setting, for any data selling volume, the seller prefers the buyer to believe that the market is of the H-type. Hence, for any off-equilibrium strategy x', the seller's highest profit is derived if the buyer believes the market is of the H-type. We first examine the H-type seller's incentive to deviate. If the H-type seller deviates, the highest profit is given by $\Pi_s(\sigma_H, x') + T(\sigma_H, x')$. Otherwise, if the H-type seller maintains the equilibrium strategy, the equilibrium profit is given by $\Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$. Hence, we can derive that

$$\Phi_H \equiv \{x' : \Pi_s(\sigma_H, x') + T(\sigma_H, x') > \Pi_s(\sigma_H, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})\}.$$

For the low-type (L-type) seller, if he deviates from the equilibrium strategy, his profit is given by $\Pi_s(\sigma_L, x') + T(\sigma_H, x')$, which is increasing in x'. However, if the L-type seller maintains the equilibrium strategy, his equilibrium profit is given by $\Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$. Therefore,

$$\Phi_L \equiv \{x' : \Pi_s(\sigma_L, x') + T(\sigma_H, x') > \Pi_s(\sigma_L, x^{pool}) + \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})\}.$$

On the basis of the above analysis, we conclude that both types of sellers have an incentive to deviate. Thus, we proceed with the second step.

Step 2. Note that Φ_H characterizes the selling volume x' satisfying $D_H \equiv \Pi_s(\sigma_H, x') + T(\sigma_H, x') - \Pi_s(\sigma_H, x^{pool}) > \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$. In addition, Φ_L characterizes the selling volume

x' satisfying $D_L \equiv \Pi_s(\sigma_L, x') + T(\sigma_H, x') - \Pi_s(\sigma_L, x^{pool}) > \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool})$. The reason is that D_H and D_L are both increasing in x'. Hence, $\Phi_H(\Phi_L)$ actually captures the off-equilibrium selling volume satisfying $x' > \tilde{x}_H(x' > \tilde{x}_L)$, where $\tilde{x}_H$ and $\tilde{x}_L$ are derived by letting

$$D_H = \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool});$$

$$D_L = \beta T(\sigma_L, x^{pool}) + (1-\beta)T(\sigma_H, x^{pool}).$$

Obviously, both $\tilde{x}_H$ and $\tilde{x}_L$ are lower than x^{pool}. We next compare the sellers' incentives in deviating (i.e., compare $\tilde{x}_H$ and $\tilde{x}_L$). $D_H - D_L = \Pi_s(\sigma_H, x') - \Pi_s(\sigma_H, x^{pool}) - [\Pi_s(\sigma_L, x') - \Pi_s(\sigma_L, x^{pool})]$. Since $\frac{\partial^2 \Pi_s(\sigma_i,x')}{\partial \sigma_i \partial x'} < 0$ can be easily proven, we conclude that $\frac{\partial \Pi_s(\sigma_i,x')}{\partial \sigma_i}$ is decreasing in x' . Hence, $\frac{\partial[\Pi_s(\sigma_i,x') - \Pi_s(\sigma_i,x^{pool})]}{\partial \sigma_i} > 0$ for any $x' < x^{pool}$. Therefore, $D_H > D_L$, which implies that $\tilde{x}_H < \tilde{x}_L$. The above analysis shows that there exists a parameter range $x' \in (\tilde{x}_H,\tilde{x}_L)$, wherein the H-type seller deviates but the L-type seller does not. After observing such a deviation, the buyer will restrict the belief that such a deviation cannot be made by the L-type seller. Even though the buyer believes that the deviating seller is of the H-type, the H-type seller still has an incentive to deviate. By definition, the pooling equilibrium cannot survive the intuitive criterion. This concludes the proof. □

Lemma VI-2. *The separating equilibria characterized in propositions 2 and 5 can survive the intuitive criterion.*

Proof of Lemma VI-2. In this proof, we demonstrate that the separating equilibria under the scenarios with and without free samples can survive the intuitive criterion. We first consider the separating equilibrium without free samples (i.e., characterized in Proposition 2). In **step 1** of the intuitive criterion, we examine whether sellers have an incentive to deviate by choosing an off-equilibrium strategy regardless of the buyer's belief. For the H-type seller, suppose again that $x'_H \neq x_H^{sep}$ is an off-equilibrium strategy. Let Φ_H be the set of off-equilibrium strategies, with which the H-type seller's highest profit earned by deviating is higher than that attained by keeping the equilibrium strategy. The H-type seller's highest profit is earned if the buyer believes that the market is of the H-type. Hence, it is easy to verify that $\Phi_H \equiv \{x'_H : x_H^{sep} < x'_H < 1\}$. Similarly, for the L-type seller, suppose that $x'_L \neq x_L^{sep}$ is an off-equilibrium strategy. Additionally, let Φ_L be the set of off-equilibrium strategies in which the L-type seller would choose to deviate. Recall that, as seen in the Proof of Proposition 2, if the buyer believes that the market is of the H-type, the L-type seller's profit is increasing in the selling volume. Therefore, we conclude that $\Phi_L \equiv \{x'_L : x_H^{sep} < x'_L < 1\}$. The reason is that both types of sellers' highest profits when deviating are greater than those earned by keeping equilibrium strategies. Hence, we proceed with **step 2** of the intuitive criterion, which checks whether there exists one specific type of seller with the lowest profit in the case of deviating that is still higher than that earned under the equilibrium strategy. If so, the equilibrium fails the intuitive criterion. Sellers' lowest profits in the case of deviating are earned if the buyer believes that the market is of the L-type. Then, it is easy to verify that the H-type seller's profit under deviation is lower than that under the equilibrium strategy (i.e., condition $C2'$ in the Proof of Proposition 2 holds). For the L-type seller, if the buyer believes that the market is of the L-type, any off-equilibrium strategy will lead to a strictly lower profit than his equilibrium profit. The reason is that keeping the equilibrium strategy will lead to his first-best profit. With that said, after the buyer restricts the belief, the sellers derived in step 1 would have no incentive to deviate. Thus, the separating equilibrium without free samples survives the intuitive criterion. By following a similar reasoning, one can easily show that the separating equilibrium with free samples also survives the intuitive criterion. This concludes the proof. □

In summary, even under the alternative equilibrium refinement of the intuitive criterion, our main results regarding the interplay between data evaluation and free sample provision remain unchanged.