Appendix

Guocheng Liao, Xu Chen, and Jianwei Huang

1 Proof of Theorem 2

Assumption 1. *There are K data reporters' types, i.e., the type set is $\mathcal{B} = \{\beta_1^*, \beta_2^*, ..., \beta_K^*\}$ where $\beta_1^* < \beta_2^* < ... < \beta_K^*$. The probability of each type for each data reporter is $Pr(\beta_j = \beta_k^*) = P_k, k = 1, 2, ..., K$ for all $j \in \mathcal{M}$, where $\sum_{k=1}^{K} P_k = 1$.*

Theorem 1. *Under Assumption 1, the symmetric pure BNE of the Bayesian data reporting game has the following structure:*

- *Case 1: $a_g < a_{g1}$. The threshold $\tilde{\beta} < \beta_1^*$, and the BNE is*

$$s^*(\beta_k^*) = A_k, \ 1 \le k \le K. \tag{1}$$

Here $A_k, k = 1, ..., K$, satisfies

$$r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{K} P_i \exp(-A_i) \right)^{M-1} \beta_k^*. \tag{2}$$

- *Case 2: $a_g \in [a_{g\hat{k}}, a_{g(\hat{k}+1)})$ where $1 \le \hat{k} < K$. The threshold is $\tilde{\beta} = \beta_{\hat{k}}^*$, and the BNE is*

$$s^*(\beta_k^*) = \begin{cases} 0, & \text{if } \ 1 \le k \le \hat{k}, \\ A_k, & \text{if } \ \hat{k} < k \le K. \end{cases} \tag{3}$$

Here $A_k, k = \hat{k}+1, ..., K$ satisfies

$$r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{\hat{k}} P_i + \sum_{i=\hat{k}+1}^{K} P_i \exp(-A_i) \right)^{M-1} \beta_k^*. \tag{4}$$

- *Case 3: $a_g \ge a_{gK}$. The threshold is $\tilde{\beta} > \beta_K^*$, and the BNE is*

$$s^*(\beta_k^*) = 0, \ 1 \le k \le K. \tag{5}$$

Proof. We prove Theorem 2 by checking that every type of data reporter is making a best response to others' strategies. Based on the first-order condition, a type $\beta_k^*, 1 \le k \le K$ data reporter's best response function under symmetric BNE is as follows:

$$s^*(\beta_k^*) = \max \left\{ (M-1) \ln \left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*)) \right) + \ln(\frac{\beta_k^*}{r_d}) - a_g, 0 \right\}. \tag{6}$$

The value of $s^*(\beta_k^*)$ depends on the data collector's strategy a_g. We define some boundaries of intervals that a_g may possibly lie in. Let

$$\begin{aligned}
a_{gk} &= (M-1)\ln\left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{K} P_i \frac{\beta_k^*}{\beta_i^*}\right) + \ln\left(\frac{\beta_k^*}{r_d}\right), 1 \le k \le K-1, \\
a_{gK} &= \ln\left(\frac{\beta_K^*}{r_d}\right).
\end{aligned} \tag{7}$$

Now we check that the following strategies indeeds satisfy the best response function under different values of a_g, and thus, constitutes the BNE.

1.1

We prove that when $a_g \ge a_{gK} = \ln\left(\frac{\beta_K^*}{r_d}\right)$, the following strategy constitutes a BNE:

$$s^*(\beta_k^*) = 0, k = 1, ..., K. \tag{8}$$

For the type $\beta_k^*, k = 1, ..., K$, we have

$$\begin{aligned}
&(M-1)\ln\left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*))\right) + \ln(\frac{\beta_k^*}{r_d}) - a_g \\
\le &\ln(\frac{\beta_k^*}{r_d}) - \ln\left(\frac{\beta_K^*}{r_d}\right) \\
\le &0.
\end{aligned} \tag{9}$$

The first inequality is due to $a_g \ge \ln\left(\frac{\beta_K^*}{r_d}\right)$. The second inequality is due to $\beta_k^* \le \beta_K^*$.

So we have $s^*(\beta_k^*) = 0$, for $k = 1, ..., K$, satisfies the best response function (6) and thus, constitutes the BNE.

1.2

We prove that when $a_g = a_{g\hat{k}}, \hat{k} = 1, ..., K-1$, the following strategy constitutes a BNE.

$$s^*(\beta_k^*) = \begin{cases} 0, & \text{if } 1 \le k \le \hat{k}, \\ \ln\left(\frac{\beta_k^*}{\beta_{\hat{k}}^*}\right), & \text{if } i > \hat{k}. \end{cases} \tag{10}$$

For type $\beta_k^*, k = 1, ..., \hat{k}$, we have

$$\begin{aligned}
&(M-1)\ln\left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*))\right) + \ln(\frac{\beta_k^*}{r_d}) - a_g \\
=&(M-1)\ln\left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{K} P_i \frac{\beta_k^*}{\beta_i^*}\right) + \ln(\frac{\beta_k^*}{r_d}) - (M-1)\ln\left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{K} P_i \frac{\beta_k^*}{\beta_i^*}\right) - \ln\left(\frac{\beta_{\hat{k}}^*}{r_d}\right) \\
=&\ln\left(\frac{\beta_k^*}{r_d}\right) - \ln\left(\frac{\beta_{\hat{k}}^*}{r_d}\right) \\
\le&0.
\end{aligned} \tag{11}$$

The equality holds when it is $\beta_{\hat{k}}^*$ type, i.e., $k = \hat{k}$. So we have $s^*(\beta_k^*) = 0$ satisfies the best response function (6) for $k = 1, ..., \hat{k}$.

For type $\beta_k^*, k = \hat{k}+1, ..., K$, we have

$$
\begin{aligned}
&(M-1)\ln\left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*))\right) + \ln(\frac{\beta_k^*}{r_d}) - a_g \\
=&(M-1)\ln\left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{K} P_i \frac{\beta_k^*}{\beta_i^*}\right) + \ln(\frac{\beta_k^*}{r_d}) - (M-1)\ln\left(\sum_{i=1}^{k} P_i + \sum_{i=k+1}^{K} P_i \frac{\beta_{\hat{k}}^*}{\beta_i^*}\right) - \ln\left(\frac{\beta_{\hat{k}}^*}{r_d}\right) \\
=&\ln\left(\frac{\beta_k^*}{\beta_{\hat{k}}^*}\right).
\end{aligned}
\tag{12}
$$

So we have $s^*(\beta_k^*) = \ln\left(\frac{\beta_k^*}{\beta_{\hat{k}}^*}\right)$ satisfies the best response function (6) for $k = \hat{k}+1, ..., K$.

In summary, the strategy (10) satisfies the best response function for all $k = 1, ..., K$, and thus, constitutes the BNE.

1.3

We prove when $a_{g\hat{k}} < a_g < a_{g(\hat{k}+1)}, \hat{k} = 1, ..., K-1$, the following strategy constitutes a BNE.

$$
s^*(\beta_k^*) = \begin{cases} 0, & \text{if } 1 \le k \le \hat{k}, \\ A_k > 0, & \text{if } \hat{k} < k \le K. \end{cases} \tag{13}
$$

Here $A_k, k = \hat{k}+1, ..., K$ satisfies

$$
r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{\hat{k}} P_i + \sum_{i=\hat{k}+1}^{K} P_i \exp(-A_i)\right)^{M-1} \beta_k^*. \tag{14}
$$

For type $\beta_k^*, k = 1, .., \hat{k}$, we have shown in previous section when $a_g = a_{a\hat{k}}$, $s^*(\beta_k^*) = 0$ satisfies best response function. Therefore, when $a_g > a_{a\hat{k}}$, we still have $s^*(\beta_k^*) = 0$ from max operation satisfies best response function.

For type $\beta_k^*, k = 1, ..., K$, from (13) we have

$$
A_k = (M-1)\ln\left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*))\right) + \ln(\frac{\beta_k^*}{r_d}) - a_g. \tag{15}
$$

So we have $s^*(\beta_k^*) = A_k$, for $k = \hat{k}+1, ..., K$, satisfies the best response function (6)

In summary, the strategy (13) satisfies the best response function for all $k = 1, ..., K$, and thus, constitutes the BNE.

1.4

We prove when $a_g < a_{g1}$, the following strategy constitutes a BNE.

$$
s^*(\beta_k^*) = A_k > 0, \; 1 \le k \le K. \tag{16}
$$

Here $A_k, k = 1, ..., K$ satisfies

$$
r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{K} P_i \exp(-A_i)\right)^{M-1} \beta_k^*. \tag{17}
$$

For type $\beta_k^*, k = 1, ..., K$, from (17) we have

$$A_k = (M-1)\ln\left(\sum_{i=1}^{K} P_i \exp(-s^*(\beta_i^*))\right) + \ln(\frac{\beta_k^*}{r_d}) - a_g. \tag{18}$$

So we have $s^*(\beta_k^*) = A_k$ satisfies the best response function (6) for $k = 1, ..., K$, and thus, constitutes the BNE.

In summary, Section 1.4 proves Case 1 in Theorem 2, Section 1.2 and Section 1.3 prove Case 2 in Theorem 2, and Section 1.1 proves Case 2 in Theorem 2. □

2 Proof of Lemma 1

Lemma 1. *There is an unique optimal strategy a_g^* in $[a_{g1}, a_{gK}]$.*

Proof. We obtain the first-order derivative of the expected utility with respect to the decision variable a_g as follows:

$$\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[U_c]}{\partial a_g} = -r_g M \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} - r_c. \tag{19}$$

Next, we will prove that (i) the derivative (19) is positive in $a_g \in [0, a_{g1}]$ and negative in $a_g \in [a_{gK}, +\infty]$; (ii) the derivative (19) is strictly decreasing in $a_g \in [a_{g1}, a_{gK}]$.

(i). We first prove the derivative (19) is positive in $a_g \in [0, a_{g1}]$.

When $0 \le a_g < a_{g1}$, we have the following BNE according to Theorem 2:

$$s^*(\beta_k^*) = A_k > 0, \ 1 \le k \le K. \tag{20}$$

where $A_k, k = 1, ..., K$ satisfies

$$r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{K} P_i \exp(-A_i)\right)^{M-1} \beta_k^*. \tag{21}$$

From (21) we have

$$A_k = A_1 + \ln\left(\frac{\beta_k^*}{\beta_1^*}\right), k = 2, ..., K. \tag{22}$$

Taking the derivative with respect to a_g to (21), yields,

$$\frac{\partial A_1}{\partial a_g} = -\frac{1}{M}, \tag{23}$$

and thus,

$$\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} = \sum_{i=1}^{K} P_i \frac{\partial A_i}{\partial a_g} = -\frac{1}{M}. \tag{24}$$

Putting (24) to (19), we have

$$\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[U_c]}{\partial a_g} = -r_g M \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} - r_c = r_g - r_c > 0. \tag{25}$$

So the derivative is positive in $a_g \in [0, a_{g1}]$.

We then prove the derivative (19) is negative in $a_g \in [a_{gK}, +\infty]$.

When $a_g > g_{gK}$, we have the following BNE according to Theorem 2:

$$s^*(\beta_k^*) = 0, \forall k = 1, ..., K. \tag{26}$$

Thus, we have $\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} = 0$ and therefore,

$$\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}\left[U_c\right]}{\partial a_g} = -r_c < 0. \tag{27}$$

So the derivative (19) is negative in $a_g \in [a_{gK}, +\infty]$.

(ii). We will prove that the derivative (19) is strictly decreasing in $a_g \in [a_{g1}, a_{gK}]$. When $a_{g1} < a_g < a_{gK}$, we have the BNE as follows:

$$s^*(\beta_k^*) = \begin{cases} 0, & \text{if } \ 1 \le k \le \hat{k}, \\ A_k, & \text{if } \ \hat{k} < k \le K. \end{cases} \tag{28}$$

for $a_g \in [a_{g\hat{k}}, a_{g(\hat{k}+1)})$. Here $A_k, k = \hat{k}+1, ..., K$ satisfies

$$r_d \exp(a_g + A_k) = \left(\sum_{i=1}^{\hat{k}} P_i + \sum_{i=\hat{k}+1}^{K} P_i \exp(-A_i) \right)^{M-1} \beta_k^*. \tag{29}$$

Consider type β_K^*, we have

$$r_d \exp(a_g + A_K) = \left(\sum_{i=1}^{\hat{k}} P_i + \sum_{i=\hat{k}+1}^{K} P_i \exp(-A_i) \right)^{M-1} \beta_K^*. \tag{30}$$

Taking the derivative with respect to A_g to (30), yields,

$$\frac{\partial A_K}{\partial a_g} = -\frac{\sum_{i=1}^{k} P_i + \exp(-A_K) \sum_{i=k+1}^{K} \frac{\beta_K^*}{\beta_i^*} P_i}{\sum_{i=1}^{k} P_i + M \exp(-A_K) \sum_{i=k+1}^{K} \frac{\beta_K^*}{\beta_i^*} P_i}. \tag{31}$$

Putting (31) to (19), we have

$$\begin{aligned} \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}\left[U_c\right]}{\partial a_g} &= -r_g M \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} - r_c \\ &= r_g \sum_{i=\hat{k}+1}^{K} P_i \cdot \left(1 + \frac{(M-1) \sum_{i=1}^{\hat{k}} P_i}{\sum_{i=1}^{\hat{k}} P_i + M \exp(-A_K) \sum_{i=\hat{k}+1}^{K} \frac{\beta_K^*}{\beta_i^*} P_i} \right) - r_c. \end{aligned} \tag{32}$$

Since A_K is decreasing in $a_g \in (a_{g1}, a_{gK})$, the term $\frac{(M-1) \sum_{i=1}^{\hat{k}} P_i}{\sum_{i=1}^{\hat{k}} P_i + M \exp(-A_K) \sum_{i=\hat{k}+1}^{K} \frac{\beta_K^*}{\beta_i^*} P_i}$ is decreasing in a_g. And the term $\sum_{i=\hat{k}+1}^{K} P_i$ is non-increasing as a_g increasing, since the threshold $\hat{k}$ would be greater. In conclusion, the derivative (19) is decreasing in a_g.

From (i) we can see that the optimal solution is in $[a_{g1}, a_{gK}]$ and from (ii), the optimal strategy a_g^* is unique in $[a_{g1}, a_{gK}]$. □

3 Proof of Lemma 2

Lemma 2. *In the sub-interval* $[a_{gk}, a_{g(k+1)}]$, $k = 1, ..., K-1$, *if*

1. $h_k(0) \ge 0$*: the expected utility is increasing in* a_g*.*

2. $h_k\left(\ln\left(\beta_{k+1}^*/\beta_k^*\right)\right) \le 0$*: the expected utility is decreasing in* a_g*.*

3. $h_k\left(0\right) < 0$ *and* $h_k\left(\ln\left(\beta_{k+1}^*/\beta_k^*\right)\right) > 0$*: the expected utility is firstly increasing in* a_g *and then decreasing in* a_g*.*

Proof In a sub-interval $[a_{gk}, a_{g(k+1)}], 1 \le k \le K-1$, we characterize the derivative as a function of $A_{k+1} \in [0, \ln(\beta_{k+1}^*/\beta_k^*)]$ using implicit function theorem:

$$\begin{aligned}
h_k(A_{k+1}) &\triangleq \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}\left[U_c\right]}{\partial a_g}\left(A_{k+1}\right) \\
&= -r_g M \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} - r_c \\
&= r_g \sum_{i=k+1}^{K} P_i \cdot \left(1 + \frac{(M-1)\sum_{i=1}^{k} P_i}{\sum_{i=1}^{k} P_i + M\sum_{i=k+1}^{K} \frac{\beta_{k+1}^*}{\beta_i^*} P_i \exp(-A_{k+1})}\right) - r_c.
\end{aligned} \tag{33}$$

Obviously the derivative $\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[U_c]}{\partial a_g}$ is increasing in $A_{k+1} \in [0, \ln(\beta_{k+1}^*/\beta_k^*)]$. It is enough to check the signs of the derivative at the boundaries to obtain the monotonicity of the expected utility. Under the BNE, we have

$$A_{k+1} = \begin{cases} \ln(\beta_{k+1}^*/\beta_k^*), & \text{if} \quad a_g = a_{gk}, \\ 0, & \text{if} \quad a_g = a_{gk+1}. \end{cases} \tag{34}$$

So the derivative value at the left (right) boundary is exactly $h_k\left(\ln\left(\beta_{k+1}^*/\beta_k^*\right)\right)$ $\left(h_k\left(0\right)\right)$. We obtains three possibilities regarding monotonicity of the expected utility, depending on the signs of the derivatives at both boundaries.

- If $h_k\left(0\right) \ge 0$, then $h_k\left(\ln(\beta_{k+1}^*/\beta_k^*)\right) > 0$. The derivate is positive in the sub-interval $[a_{gk}, a_{g(k+1)}]$, and thus the expected utility is increasing in a_g.
- If $h_k\left(\ln\left(\beta_{k+1}^*/\beta_k^*\right)\right) \le 0$, then $h_k\left(0\right) < 0$. The derivate is negative in the sub-interval $[a_{gk}, a_{g(k+1)}]$, and thus the expected utility is decreasing in a_g.
- If $h_k\left(0\right) < 0$ and $h_k\left(\ln\left(\beta_{k+1}^*/\beta_k^*\right)\right) > 0$, then the derivate changes from positive to negative in the sub-interval $[a_{gk}, a_{g(k+1)}]$, and the expected utility is firstly increasing in a_g and then decreasing in a_g. □

4 Proof of Theorem 3

Theorem 2. *Under Assumption 1, the Algorithm 1 computes the data collector's optimal strategy* a_g^**, i.e., the solution to her optimization problem.*

Proof. According to Lemma 1, there exits a unique optimal strategy in $[a_{g1}, a_{gK}]$. We can find the optimal strategy through the derivative:

$$\frac{\partial \mathbb{E}_{\boldsymbol{\beta}}\left[U_c\right]}{\partial a_g} = -r_g M \frac{\partial \mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*]}{\partial a_g} - r_c. \tag{35}$$

Since the term $\mathbb{E}_{\boldsymbol{\beta}}[\boldsymbol{s}^*(\boldsymbol{\beta})]$ in the derivate has different expressions when a_g lies in different intervals, according to Theorem 2. We should instead focus on the sub-intervals $[a_{gk}, a_{g(k+1)}], k = 1, ..., K-1$.

According to Lemma 2, we obtain three possibilities regarding monotonicity of the expected utility in the sub-interval. In summary, the sequential checking on the sub-intervals that finds when the derivative becomes negative can generate the optimal strategy a_g^*. □

5 Proof of Proposition 2

Proposition 1. *When there are two types of data reporters, we have*

$$S_{in} \leq S_{com}. \tag{36}$$

The equality holds when $a_g^* = a_{g2} = \ln(\beta_2^*) - \ln r_d$ under the incomplete information scenario in Theorem 2.

Proof. When there are two types, we have

$$a_{g1} = \ln\left(\beta_1^*\left(P_1 + P_2\frac{\beta_1^*}{\beta_2^*}\right)^{M-1}\right) - \ln(r_d).$$

$$a_{g2} = \ln(\beta_2^*) - \ln(r_d).$$

Calculate the derivatives at $a_g = a_{g2}$ and at $a_g = a_{g1}$, respectively:

$$D_r = \frac{P_2 r_g M}{1 - P_2 + MP_2} - r_c, \tag{37}$$

$$D_l = \frac{P_2 r_g M(1 - P_2 + \delta P_2)}{1 - P_2 + M\delta P_2} - r_c, \tag{38}$$

where $\delta \triangleq \frac{\beta_1^*}{\beta_2^*} < 1$. Note that $D_l > D_r$. According to Theorem 3, the optimal a_g^* depends on the sign of the derivatives D_r and D_l given P_1, M and β.

- Case 1: If $D_l > D_r \geq 0$, we have $a_g^* = a_{g2} = \ln(\beta_2^*) - \ln r_d$ and $A_1 = A_2 = 0$.
- Case 2: If $D_r < D_l \leq 0$, we have $a_g^* = a_{g1} = (M-1)\ln(P_1 + P_2\delta) + \ln(\frac{\beta_1^*}{r_d})$, $A_1 = 0$ and $A_2 = \ln(\frac{1}{\delta})$.
- Case 3: If $D_r < 0$ and $D_l > 0$, we have $A_2 \triangleq s^*(\beta_2^*) = \ln\left(\frac{MP_2}{\frac{(M-1)P_1P_2r_g}{r_c - P_2r_g} - P_1}\right)$, $A_1 = 0$, and

$$a_g^* = (M-1)\ln(P_1 + P_2\exp(-A_2)) + \ln(\frac{\beta_2^*}{r_d}) - A_2.$$

In Case 1, we have $S_{in} = M\mathbb{E}_\beta[s^*(\beta)] + a_g^* = \ln(\beta_2^*) - \ln r_d$ and $S_{com} = \ln(\beta_2^*) - \ln r_d$, i.e., $S_{in} = S_{com}$.
In Case 2, we have

$$\begin{aligned} S_{in} =& M\mathbb{E}_\beta[s^*(\beta)] + a_g^* \\ =& MP_2\ln(\frac{1}{\delta}) + (M-1)\ln(P_1 + P_2\delta) + \ln(\frac{\beta_1^*}{r_d}) \\ =& \ln\left((\frac{1}{\delta})^{MP_2-1}(P_1 + P_2\delta)^{M-1}\right) + \ln\left(\frac{\beta_2^*}{r_d}\right). \end{aligned} \tag{39}$$

Define $f(\delta) = (\frac{1}{\delta})^{MP_2-1}(P_1 + P_2\delta)^{M-1}$. We can check $f'(\delta) \geq 0$ in this case and that $f(\delta) < f(1) = 1$. So we have $S_{in} < \ln(f(1)) + \ln\left(\frac{\beta_2^*}{r_d}\right) = \ln\left(\frac{\beta_2^*}{r_d}\right) = S_{com}$, i.e., $S_{in} < S_{com}$.

In Case 3, we have

$$
\begin{aligned}
S_{in} =& M\mathbb{E}_\beta[s^*(\beta)] + a_g^* \\
=& A_2(MP_2-1) + (M-1)\ln(P_1 + P_2\exp(-A_2)) + \ln\left(\frac{\beta_2^*}{r_d}\right) \\
=& \ln\left(\left(\frac{MP_2(r_c - P_2 r_g)}{(MP_2 r_g - r_c)(1-P_2)}\right)^{MP_2-1}\left(\frac{(M-1)(1-P_2)r_c}{M(r_c - P_2 r_g)}\right)^{M-1}\right) \\
&+ \ln\left(\frac{\beta_2^*}{r_d}\right).
\end{aligned}
\tag{40}
$$

Define $g(P_2) = \left(\frac{MP_2(r_c-P_2r_g)}{(MP_2r_g-r_c)(1-P_2)}\right)^{MP_2-1}\left(\frac{(M-1)(1-P_2)r_c}{M(r_c-P_2r_g)}\right)^{M-1}$. We can check that $g(P_2)$ is an increasing function in this case. We have $P_2 < \frac{1}{1+M(\frac{r_g}{r_c}-1)}$ from $D_r < 0$. So $g(P_2) < f(\frac{1}{1+M(\frac{r_g}{r_c}-1)}) = 1$. So we have $S_{in} < \ln(g(1)) + \ln\left(\frac{\beta_2^*}{r_d}\right) = \ln\left(\frac{\beta_2^*}{r_d}\right) = S_{com}$, i.e., $S_{in} < S_{com}$.

In summary, we have $S_{in} \leq S_{com}$ and the equality holds in Case 1. ☐

6 Proof of Proposition 3

Proposition 2. *When there are two types of data reporters, we have*

$$
S_{com} - S_{in} < \ln\left(\frac{\beta_2^*}{\beta_1^*}\right). \tag{41}
$$

Proof. In the case 2 in the proof of Proposition 2, we have

$$
S_{com} - S_{in} = -\ln\left((\frac{1}{\delta})^{MP_2-1}(P_1 + P_2\delta)^{M-1}\right). \tag{42}
$$

In case 3 in the proof of Proposition 2, from $D_l > 0$ we have

$$
\delta < \frac{(MP_2r_g - r_c)(1-P_2)}{MP_2(r_c - P_2r_g)}. \tag{43}
$$

Recall that $f(\delta) = (\frac{1}{\delta})^{MP_2-1}(P_1+P_2\delta)^{M-1}$ is an increasing function. Thus we have $f(\delta) < f\left(\frac{(MP_2r_g-r_c)(1-P_2)}{MP_2(r_c-P_2r_g)}\right) = \left(\frac{MP_2(r_c-P_2r_g)}{(MP_2r_g-r_c)(1-P_2)}\right)^{MP_2-1}\left(\frac{(M-1)(1-P_2)r_c}{M(r_c-P_2r_g)}\right)^{M-1}$.

Recall that we have

$$
S_{com} - S_{in} = -\ln\left(\left(\frac{MP_2(r_c - P_2r_g)}{(MP_2r_g - r_c)(1-P_2)}\right)^{MP_2-1}\left(\frac{(M-1)(1-P_2)r_c}{M(r_c - P_2r_g)}\right)^{M-1}\right). \tag{44}
$$

in the case 3 in the proof the Theorem 4. This means that the difference $S_{com} - S_{in}$ in Case 2 is greater than that in Case 3. So it suffices to focus on Case 2 to find the upper bound of $S_{com} - S_{in} = -\ln\left((\frac{1}{\delta})^{MP_2-1}(P_1 + P_2\delta)^{M-1}\right)$. Define $h(P_2) = (\frac{1}{\delta})^{MP_2-1}(1 - P_2 + P_2\delta)^{M-1}$. We can check that $h(P_2)$ is an increasing function in Case 2. Thus, we have $h(P_2) > h(P_2 = 0) = \ln(\delta) = \ln(\frac{\beta_1^*}{\beta_2^*})$. Thus, $S_{com} - S_{in} < \ln(\frac{\beta_2^*}{\beta_1^*})$. ☐