Supplemental Materials

Appendix A: Proofs and Supplements

A.1. Proofs and Supplements in Section 4

A.1.1. NP-hardness of the CGPO problem.

Proof of Theorem 1: We prove the NP-hardness of the CGPO problem by reduction from the SUBSET-SUM problem. Let the positive integers $b_1, \dots b_n$, and B form an instance of SUBSET-SUM. Without loss of generality, we assume that $b_i \le B$ for all $i \in \{1, 2, \dots, n\}$. Let $s_i = b_i/B$. We have $s_i \in [0, 1]$ for all i.

We construct a special instance of the CGPO problem where $\mathcal{V} = \{1, 2, \dots, n\}$, $L = 2$, $C = m$, $K \in \mathbb{Z}_+$, $A_{v,0} = 0$ and $p_v \le 1/2$ for all $v \in \mathcal{V}$. By this construction, the optimal solution satisfies $x^*_{v,2} = 0$ for all $v \in \mathcal{V}$. We show this claim by contradiction. Assume the optimal solution $x^*_{v,2} > 0$. We construct a feasible solution: $x'_{v,1} = x^*_{v,1} + x^*_{v,2}$ and $x'_{v,2} = 0$. Let $x^*_{v,1} + x^*_{v,2} = c_v \le 1$. The optimal adoption is given by

$$A^*_{v,2} = -p_v^2 q_v m {x^*_{v,1}}^2 + p_v m q_v x^*_{v,1} + p_v m (x^*_{v,1} + x^*_{v,2}) = -p_v^2 q_v m {x^*_{v,1}}^2 + p_v m q_v x^*_{v,1} + p_v m c_v \,.$$

Similarly, the adoptions with $x'_{v,1}$ and $x'_{v,2}$ is $A'_{v,2} = -p_v^2 q_v m {x'_{v,1}}^2 + p_v m q_v x'_{v,1} + p_v m c_v$. Given $p_v \le 1/2$, we have $A'_{v,2} - A^*_{v,2} = p_v q_v m x^*_{v,2}(1 - p_v(x'_{v,1} + x^*_{v,1})) \ge p_v q_v m x^*_{v,2}(1 - 2p_v) \ge 0$, contradicting with the assumption.

Therefore, we can omit the variables $\boldsymbol{x}_{:,2}$ and write $\boldsymbol{x} = \boldsymbol{x}_{:,1}$ for simplicity. Let $\alpha_v = p_v^2 q_v m$ and $\beta_v = p_v m (q_v + 1)$. Under this construction, the CGPO problem can be expressed as

$$\begin{aligned} \max_{\boldsymbol{x}, U \subseteq \mathcal{V} : |U| \le K} \quad & \sum_{v \in \mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v \\ \text{s.t.} \quad & \sum_{v \in \mathcal{V}} x_v \le 1, \\ & 0 \le x_v \le \mathbb{1}\{v \in U\}, \forall v \in \mathcal{V}. \end{aligned}$$

We assume that there exists values of $\boldsymbol{p}$, $\boldsymbol{q}$ and m such that for all $v \in \mathcal{V}$, $-2\alpha_v s_v + \beta_v = d_1$ and $\alpha_v s_v^2 = d_2$ for some $d_1, d_2 \in \mathbb{R}_+$. We justify the existence of $\boldsymbol{p}$, $\boldsymbol{q}$, and m at the end of this proof. Here, we let $\ell(\boldsymbol{x}) = \sum_{v \in \mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v$. We claim that

$$\ell(\boldsymbol{x}^*) \ge d_1 + d_2 K \,, \Longleftrightarrow \text{ there exists } U \subseteq \mathcal{V} \text{ with } |U| = K \text{ and } \sum_{v \in U} s_v = 1 \,.$$

To prove this claim, we first express the objective value as

$$\begin{aligned} \ell(\boldsymbol{x}) &= \sum_{v \in \mathcal{V}} -\alpha_v x_v^2 + \beta_v x_v \overset{(a)}{=} \sum_{v \in U} -\alpha_v x_v^2 + \beta_v x_v \\ &\overset{(b)}{=} \sum_{v \in U} -\alpha_v (x_v - s_v)^2 + d_1 x_v + d_2 \overset{(c)}{\le} d_1 \sum_{v \in U} x_v + d_2 |U| \overset{(d)}{\le} d_1 + d_2 K \,. \end{aligned}$$

where (a) follows from constraint $x_v \le \mathbb{1}\{v \in U\}$; (b) follows from the definition of d_1 and d_2; (c) follows since $\alpha_v \ge 0$; (d) follows from the constraint $\sum_{v \in \mathcal{V}} x_v \le 1$. If there exists $U \subseteq \mathcal{V}$ such that $|U| = K$ and $\sum_{v \in U} s_v = 1$, then we can let $U^* = U$ and $x^*_v = s_v \mathbb{1}\{v \in U\}$ for all $v \in \mathcal{V}$. Then, it is easy to verify that $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$. On the other hand, if $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$, then the (c) implies that we must have $x^*_v = s_v \mathbb{1}\{v \in U^*\}$; (d) implies that $\sum_{v \in U^*} x^*_v = \sum_{v \in U^*} s_v = 1$ and $|U^*| = K$.

This claim allows us to reduce SUBSET-SUM to our problem as follows. If there exists a subset $I \subseteq \{1, \dots, n\}$ such that $\sum_{i \in I} s_i = 1$, then the objective value $\ell(\boldsymbol{x}^*)$ is equal to $d_1 + d_2 K$ for $K = |I|$. If there exists $K \in \{1, 2, \dots, n\}$ such that $\ell(\boldsymbol{x}^*) = d_1 + d_2 K$, one can find $I = U^*$ such that $\sum_{i \in I} s_i = 1$.

We then proceed by providing an example of the values of m, d_1, and d_2, which ensures that our construction serves as a valid CGPO instance. Let $\underline{s} = \min_{v\in\mathcal{V}} s_v$, $m = 64/\underline{s}^2$, $d_1 = m/2$ and $d_2 = 1$. Our construction assumes

$$\begin{cases} -2p_v^2 q_v m s_v + p_v m(q_v+1) & = d_1 \\ p_v^2 q_v m s_v^2 & = d_2 \end{cases}, \Longrightarrow \begin{cases} p_v q_v + p_v & = \frac{1}{2} + \frac{2}{m s_v} \\ p_v q_v \cdot p_v & = \frac{1}{m s_v^2} \end{cases}.$$

Therefore, for any given s_v, we can solve p_v and q_v as demonstrated in (EC.1):

$$p_v = \frac{\frac{1}{2} + \frac{2}{m s_v} + \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}}{2} \text{ and } q_v = \frac{\frac{1}{2} + \frac{2}{m s_v} - \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}}{\frac{1}{2} + \frac{2}{m s_v} + \sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}}}. \tag{EC.1}$$

To ensure that p_v and q_v provided in (EC.1) are valid within the context of P-BDM, we need further justifications. First, we need to ensure (EC.1) has real value solutions, that is,

$$\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2} = \frac{1}{4} + \frac{2}{m s_v} + \left(\frac{4}{m^2} - \frac{4}{m}\right)\frac{1}{s_v^2} \geq \frac{1}{4} + \frac{2}{m} + \left(\frac{4}{m^2} - \frac{4}{m}\right)\frac{1}{\underline{s}^2} = \frac{3}{16} + \frac{33}{16m} \geq 0.$$

Next, we need to validate that $p_v, q_v \geq 0$, $p_v \leq 1/2$, and $p_v + q_v \leq 1$. It is obvious that $p_v, q_v \geq 0$ by (EC.1). In order to show $p_v + q_v \leq 1$, it suffices to show that $p_v, q_v \leq 1/2$. To show $p_v \leq 1/2$, we have

$$\sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}} \leq \frac{1}{2} - \frac{2}{m s_v} \Longleftrightarrow \frac{1}{2} - \frac{2}{m s_v} \geq 0 \text{ and } \sqrt{\left(\frac{1}{2} - \frac{2}{m s_v}\right)^2 + \frac{4}{m s_v} - \frac{4}{m s_v^2}} \leq \frac{1}{2} - \frac{2}{m s_v},$$

where the inequalities follow since $0 \leq s_v \leq 1$. To show $q_v \leq 1/2$, we have

$$\frac{1}{2} + \frac{2}{m s_v} \leq 3\sqrt{\left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{4}{m s_v^2}} \Longleftrightarrow \left(\frac{1}{2} + \frac{2}{m s_v}\right)^2 - \frac{9}{2 m s_v^2} \geq \frac{23}{128} + \frac{33}{16m} \geq 0.$$

Therefore, the construction of this specific CGPO instance is valid. Such construction exists for every instance of SUBSET-SUM. In conclusion, the CGPO problem is NP-hard. ☐

A.1.2. Proofs for the PO-CR problem.

Proof of Theorem 2: To establish the equivalence of two problems, we need to show that for all optimal solutions of the PO-CR problem, equalities hold for all constraints (5). We will show this by contradiction.

Let $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ and R^* be the optimal solution and optimal value of the PO-CR problem. Assume there exists $u \in U$ and $\tau \in \{1, \ldots, L\}$ such that the following inequality holds: $A^*_{u,\tau} < A^*_{u,\tau-1} + p_v m x^*_{u,\tau} + \frac{q_u}{m} A^*_{u,\tau-1}(m - A^*_{u,\tau-1})$.

By the optimality of $(\boldsymbol{x}^*, \boldsymbol{A}^*)$, it is straightforward to confirm that $(\boldsymbol{x}^*_{U,\tau+1:L}, \boldsymbol{A}^*_{U,\tau+1:L})$ is also the optimal solution of the following problem (EC.2). Consequently, the optimal value of problem (EC.2) is equal to R^*.

$$\max_{\boldsymbol{x}\geq \boldsymbol{0}, \boldsymbol{A}_{U,\tau+1:L}} \quad \sum_{v\in U} A_{v,L} \tag{EC.2a}$$

$$\text{s.t.} \quad A_{v,\tau+1} \leq A^*_{v,\tau} + p_v m x_{v,\tau+1} + q_v \frac{A^*_{v,\tau}}{m}(m - A^*_{v,\tau}), \quad \forall v \in U, \tag{EC.2b}$$

$$x_{v,\tau+1} \leq 1 - \frac{A^*_{v,\tau}}{m}, \quad \forall v \in U, \tag{EC.2c}$$

$$A_{v,t} \leq A_{v,t-1} + p_v m x_{v,t} + q_v \frac{A_{v,t-1}}{m}(m - A_{v,t-1}), \forall v \in U \quad \forall t = \tau+2, \ldots, L, \tag{EC.2d}$$

$$x_{v,t} \leq 1 - \frac{A_{v,t-1}}{m}, \quad \forall v \in U \quad \forall t = \tau+2, \ldots, L, \tag{EC.2e}$$

$$m \sum_{t=\tau+1}^{L} \sum_{v\in U} x_{v,t} \leq C - m \sum_{t=1}^{\tau} \sum_{v\in U} x^*_{v,t}. \tag{EC.2f}$$

We then construct a feasible solution $(\boldsymbol{x}', \boldsymbol{A}')$ for the PO-CR problem and let R' be the objective value:

(i) When $t = 1, 2, \ldots, \tau-1$, $\boldsymbol{x}'_{U,t} := \boldsymbol{x}^*_{U,t}$; When $t = 0, 1, \ldots, \tau-1$, $\boldsymbol{A}'_{U,t} := \boldsymbol{A}^*_{U,t}$.

(ii) When $t = \tau$, $\boldsymbol{x}'_{U,t} := \boldsymbol{x}^*_{U,t}$ and $A'_{v,t} := \begin{cases} A^*_{v,t-1} + p_v m x^*_{u,t-1} + q_v \frac{A^*_{v,t-1}}{m}(m - A^*_{v,t-1}), & \text{when } v \in \{u\}, \\ A^*_{v,t}, & \text{when } v \in U \setminus \{u\}. \end{cases}$

(iii) When $t=\tau+1,\dots,L$, $(\boldsymbol{x}'_{U,t},\boldsymbol{A}'_{U,t})$ is defined as the optimal solution of the following problem (EC.3):

$$\max_{\boldsymbol{x}\geq \boldsymbol{0},\boldsymbol{A}_{U,\tau+1:L}} \quad \text{(EC.2a)}$$

$$\text{s.t.} \quad A_{v,\tau+1}\leq A'_{v,\tau}+p_v m x_{v,\tau+1}+\frac{q_v}{m}A'_{v,\tau}(m-A'_{v,\tau}), \forall v\in U, \tag{EC.3a}$$

$$x_{v,\tau+1}\leq 1-\frac{A'_{v,\tau}}{m}, \quad \forall v\in U, \tag{EC.3b}$$

$$\text{(EC.2d)} - \text{(EC.2f)}.$$

As a consequence, the optimal value of problem (EC.3) is also equal to R'.

In the following, we focus on problems (EC.2) and (EC.3). We aim to show that $R'>R^*$, which contradicts with the optimality of $(\boldsymbol{x}^*,\boldsymbol{A}^*)$. Firstly, to characterize the optimal solution for problem (EC.2), we use Lagrangian multipliers to introduce the constraints (EC.2b), (EC.2c) and (EC.2f) into the objective function. Let Ω denote the feasible region constructed by constraints (EC.2d), (EC.2e). The dual problem can thus be expressed as

$$\min_{\boldsymbol{\lambda}\geq\boldsymbol{0},\boldsymbol{\mu}\geq\boldsymbol{0},\theta\geq 0} r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)+\sum_{v\in U}\lambda_v\left(A^*_{v,\tau}+\frac{q_v}{m}A^*_{v,\tau}(m-A^*_{v,\tau})\right)+\sum_{v\in U}\mu_v\left(1-\frac{A^*_{v,\tau}}{m}\right)+\theta C\,, \tag{EC.4}$$

where $r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)$ is the optimal value function of a parameterized maximization problem as shown in (EC.5):

$$r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta):=\max_{\substack{\boldsymbol{x}_{U,\tau+1:L}\geq 0,\\ (\boldsymbol{x},\boldsymbol{A})_{U,\tau+1:L}\in\Omega}}\sum_{v\in U}\left[A_{v,L}-\lambda_v A_{v,\tau+1}+(\lambda_v m p_v-\mu_v-\theta m)x_{v,\tau+1}-\theta m\left(\sum_{t=1}^{\tau}x^*_{v,t}+\sum_{t=\tau+2}^{L}x_{v,t}\right)\right]. \tag{EC.5}$$

Notice that, in the maximization problem (EC.5), variable $\boldsymbol{x}_{U,\tau+1}$ is not constrained by any other variables, but only by a constant 0. Thus, we can split problem (EC.5) into two subproblems:

$$\max_{\substack{\boldsymbol{x}_{U,\tau+2:L}\geq 0,\\ (\boldsymbol{x},\boldsymbol{A})_{U,\tau+1:L}\in\Omega}}\sum_{v\in U}\left[A_{v,L}-\lambda_v A_{v,\tau+1}-\theta m\left(\sum_{t=1}^{\tau}x^*_{v,t}+\sum_{t=\tau+2}^{L}x_{v,t}\right)\right] \text{ and } \max_{\boldsymbol{x}_{U,\tau+1}\geq 0}\sum_{v\in U}(\lambda_v m p_v-\mu_v-\theta m)\,x_{v,\tau+1}\,.$$

We then optimize two subproblems separately. For the former subproblem, we denote its optimal value as $s(\boldsymbol{\lambda},\theta)$. For the latter subproblem, the maximization problem can be further decomposed for each content piece $v\in U$ and we can easily derive the optimal solution and the value of function $r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)$ as follows:

$$x^*_{v,\tau+1}=\begin{cases}0 & \text{when } \mu_v\geq m(\lambda_v p_v-\theta),\\ +\infty & \text{when } \mu_v< m(\lambda_v p_v-\theta),\end{cases} \text{ and } r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)=\begin{cases}s(\boldsymbol{\lambda},\theta) & \text{when } \theta>\min_{v\in U}(\lambda_v p_v-\frac{\mu_v}{m}),\\ +\infty & \text{o.w.}\end{cases}$$

We then substitute the value function of $r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)$ into the dual problem (EC.4). To minimize the dual value, it is obvious that $r(\boldsymbol{\lambda},\boldsymbol{\mu},\theta)$ should not be infinity. Hence, the dual problem (EC.4) can be reformulated as (EC.6),

$$\min_{\boldsymbol{\lambda}\geq 0,\theta\geq 0}\left[s(\boldsymbol{\lambda},\theta)+\sum_{v\in U}\lambda_v\left(A^*_{v,\tau}+\frac{q_v}{m}A^*_{v,\tau}(m-A^*_{v,\tau})\right)+\theta C+\sum_{v\in U}\min_{\substack{\mu_v\geq 0,\\ \mu_v\geq m(\lambda_v p_v-\theta)}}\mu_v\left(1-\frac{A^*_{v,\tau}}{m}\right)\right]. \tag{EC.6}$$

The inner minimization of problem (EC.6) can be represented as (EC.7). For any $v\in U$,

$$\min_{\substack{\mu_v\geq 0,\\ \mu_v\geq m(\lambda_v p_v-\theta)}}\mu_v\left(1-\frac{A^*_{v,\tau}}{m}\right). \tag{EC.7}$$

Since $A^*_{v,\tau}\leq m$ always holds, when given the value of $(\boldsymbol{\lambda},\theta)$, the optimal solution of μ_v should be $\mu^*_v(\boldsymbol{\lambda},\theta)=\max\{0,m(\lambda_v p_v-\theta)\}$. Finally, we define the dual value function as

$$u(\boldsymbol{\lambda},\theta;\boldsymbol{A}^*_{U,\tau}):=s(\boldsymbol{\lambda},\theta)+\sum_{v\in U}\lambda_v\left(A^*_{v,\tau}+\frac{q_v}{m}A^*_{v,\tau}(m-A^*_{v,\tau})\right)+\theta C+\sum_{v\in U}\left(1-\frac{A^*_{v,\tau}}{m}\right)\max\{0,m(\lambda_v p_v-\theta)\}.$$

Similarly, for problem (EC.3), we can also have the dual value function as

$$u(\boldsymbol{\lambda},\theta;\boldsymbol{A}'_{U,\tau}):=s(\boldsymbol{\lambda},\theta)+\sum_{v\in U}\lambda_v\left(A'_{v,\tau}+\frac{q_v}{m}A'_{v,\tau}(m-A'_{v,\tau})\right)+\theta C+\sum_{v\in U}\left(1-\frac{A'_{v,\tau}}{m}\right)\max\{0,m(\lambda_v p_v-\theta)\}.$$

Let $(\boldsymbol{\lambda}^*, \theta^*)$ and $(\boldsymbol{\lambda}', \theta')$ be the optimal dual variables for the dual problems of (EC.2) and (EC.3). Given that problems (EC.2) and (EC.3) are convex optimization problems, by invoking Slater's condition, strong duality holds for both problems. Thus, the optimal values for problems (EC.2) and (EC.3) can be represented as $R^* = u(\boldsymbol{\lambda}^*, \theta^*; \boldsymbol{A}^*_{U,\tau})$ and $R' = u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_{U,\tau})$. Finally, we compare between R^* and R' and show that $R^* \le R'$.

$$
\begin{aligned}
R^* - R' &= u(\boldsymbol{\lambda}^*, \theta^*; \boldsymbol{A}^*_\tau) - u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_\tau) \le u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}^*_\tau) - u(\boldsymbol{\lambda}', \theta'; \boldsymbol{A}'_\tau) && \text{(EC.8a)}\\
&= \sum_{v\in U} \lambda'_v \left(A^*_{v,\tau} + \frac{q_v}{m} A^*_{v,0}(m - A^*_{v,\tau}) \right) + \sum_{v\in U} \left(1 - \frac{A^*_{v,\tau}}{m}\right) \max\{0, m(\lambda'_v p_v - \theta')\}\\
&\quad - \left[\sum_{v\in U} \lambda'_v \left(A'_{v,\tau} + \frac{q_v}{m} A'_{v,\tau}(m - A'_{v,\tau}) \right) + \sum_{v\in U} \left(1 - \frac{A'_{v,\tau}}{m}\right) \max\{0, m(\lambda'_v p_v - \theta')\} \right]\\
&= \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \left[\lambda'_v \left(1 + q_v - \frac{q_v}{m}(A^*_{v,\tau} + A'_{v,\tau})\right) - \frac{1}{m}\max\{0, m(\lambda'_v p_v - \theta')\} \right]\\
&\le \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \left[\lambda'_v \left(1 + q_v - \frac{q_v}{m}(A^*_{v,\tau} + A'_{v,\tau})\right) - \lambda'_v p_v \right] && \text{(EC.8b)}\\
&< \sum_{v\in U} (A^*_{v,\tau} - A'_{v,\tau}) \lambda'_v (1 - q_v - p_v) && \text{(EC.8c)}\\
&\le 0, && \text{(EC.8d)}
\end{aligned}
$$

where (EC.8a) follows since $(\boldsymbol{\lambda}^*, \theta^*)$ is the optimal solution to dual problem (EC.4); (EC.8b) follows due to the construction of $\boldsymbol{A}^*_\tau \le \boldsymbol{A}'_\tau$ and the inherent non-negativity of dual variables (i.e., $\boldsymbol{\lambda}' \ge \mathbf{0}$ and $\theta' \ge 0$); (EC.8c) follows since by definition $q_v \ge 0$, $A^*_{u,\tau} < A'_{u,\tau} \le m$ and $A^*_{v,\tau} = A'_{u,\tau} \le m$ for all $v \in U \setminus \{u\}$; (EC.8d) follows from the definition $p_v + q_v \le 1$.

As a result, $(\boldsymbol{x}', \boldsymbol{A}')$ is a feasible solution to the PO-CR problem while the resulting objective value R' is greater than R^*. This indicates that $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ cannot be an optimal solution to the PO-CR problem.

In conclusion, the original PO problem (4) and the relaxed problem are equivalent. □

A.1.3. Lemmas and proofs for the single-variable reformulation. In the following, we include lemmas and proofs to validate the single-variable reformulation and to show that it remains a convex program.

LEMMA EC.1 **(Redundant Constraint)**. *Constraint* (6c) *is redundant to the reformulation* (6).

Proof of Lemma EC.1: Let $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ be the optimal solution of the PO problem (4). By P-BDM dynamics (4b), we can deduce that the optimal adoption number $\boldsymbol{A}^*$ is non-decreasingly evolves over time (i.e., $A_{v,0} \le A^*_{v,1} \le \cdots \le A^*_{v,L}$ for all $v \in U$). By Theorem 2, $(\boldsymbol{x}^*, \boldsymbol{A}^*)$ is also the optimal solution to the PO-CR problem. Therefore, for all $v \in U, t = 1, 2, \ldots, L$, constraint $x_{v,t} \le 1 - A_{v,0}/m$ is redundant compared with $x_{v,t} \le 1 - A_{v,t-1}/m$ in the PO-CR problem. □

Construction of Feasible Solution to Problem (7)**.**

For any $v \in U$, given $\boldsymbol{x}_{v,:} \in [0, 1 - A_{v,0}/m]^L$, a feasible solution can be constructed by the following three steps:

Step 1: Set $t := 1$. Let $A_{v,1} := A_{v,0} + p_v m x_{v,1} + q_v A_{v,0}(m - A_{v,0})/m$. Then, increment t by 1, i.e., $t := t+1$.

Step 2: Let $A_{v,t} = \min\{m(1 - x_{v,t+1}), A_{v,t-1} + p_v m x_{v,t} + q_v A_{v,t-1}(m - A_{v,t-1})/m\}$.

Step 3: Increment t by 1, i.e., $t := t+1$. Repeat step 2, until $t = L+1$. □

To demonstrate that single-variable reformulation (6) remains a convex program, it is sufficient to show that the objective is concave, given that all constraints are linear.

Proof of Lemma 1: For simplicity, we omit the subscript v here. For any $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, let $\boldsymbol{x}^{(\lambda)} = \lambda \boldsymbol{x}^{(1)} + (1-\lambda)\boldsymbol{x}^{(2)}$, and we want to show that $\lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda) f(\boldsymbol{x}^{(2)}) \le f(\boldsymbol{x}^{(\lambda)})$ holds for any $0 \le \lambda \le 1$.

Suppose $\boldsymbol{A}^{(1*)}$, $\boldsymbol{A}^{(2*)}$ and $\boldsymbol{A}^{(\lambda*)}$ are the optimal solutions to problem (7) with regard to $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$, and $\boldsymbol{x}^{(\lambda)}$, respectively. Let $\boldsymbol{A}^{(\lambda)} = \lambda\boldsymbol{A}^{(1*)} + (1-\lambda)\boldsymbol{A}^{(2*)}$. We first show that $\boldsymbol{A}^{(\lambda)}$ is a feasible solution to problem (7) with regard to $\boldsymbol{x}^{(\lambda)}$ by verifying that it satisfies constraints (5) and (4c).

For constraint (5), we have

$$\begin{aligned}
& A_{t-1}^{(\lambda)} + pmx_t^{(\lambda)} + \frac{q}{m}A_{t-1}^{(\lambda)}(m - A_{t-1}^{(\lambda)}) - A_t^{(\lambda)} \\
=& \lambda A_{t-1}^{(1*)} + (1-\lambda)A_{t-1}^{(2*)} + pm\left[\lambda x^{(1)} + (1-\lambda)x^{(2)}\right] \\
& + \frac{q}{m}\left[\lambda A_{t-1}^{(1*)} + (1-\lambda)A_{t-1}^{(2*)}\right]\left[m - \lambda A_{t-1}^{(1*)} - (1-\lambda)A_{t-1}^{(2*)}\right] - \left[\lambda A_t^{(1*)} + (1-\lambda)A_t^{(2*)}\right] \\
=& \lambda\left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + qA_{t-1}^{(1*)} - A_t^{(1*)}\right] + (1-\lambda)\left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + qA_{t-1}^{(2*)} - A_t^{(2*)}\right] - \frac{q}{m}\left[\lambda A_{t-1}^{(1*)} + (1-\lambda)A_{t-1}^{(2*)}\right]^2 \\
\geq& \lambda\left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + qA_{t-1}^{(1*)} - A_t^{(1*)}\right] + (1-\lambda)\left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + qA_{t-1}^{(2*)} - A_t^{(2*)}\right] \\
& - \frac{q}{m}\left[\lambda {A_{t-1}^{(1*)}}^2 + (1-\lambda){A_{t-1}^{(2*)}}^2\right](\lambda + 1 - \lambda) \\
=& \lambda\left[A_{t-1}^{(1*)} + pmx_{t-1}^{(1)} + \frac{q}{m}A_{t-1}^{(1*)}\left(m - A_{t-1}^{(1*)}\right) - A_t^{(1*)}\right] + (1-\lambda)\left[A_{t-1}^{(2*)} + pmx_{t-1}^{(2)} + \frac{q}{m}A_{t-1}^{(2*)}\left(m - A_{t-1}^{(2*)}\right) - A_t^{(2*)}\right] \\
\geq& 0\,,
\end{aligned}$$

where the first inequality follows from Cauchy–Schwarz inequality, the second inequality follows since $\boldsymbol{A}^{(1*)}$ and $\boldsymbol{A}^{(2*)}$ satisfy the constraint (5) with regard to $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$.

For constraint (4c), we have $1 - \frac{A_{t-1}^{(\lambda)}}{m} - x_t^{(\lambda)} = \lambda\left[1 - \frac{A_{t-1}^{(1*)}}{m} - x_t^{(1)}\right] + (1-\lambda)\left[1 - \frac{A_{t-1}^{(2*)}}{m} - x_t^{(2)}\right] \geq 0$, where the inequality follows since $\boldsymbol{A}^{(1*)}$ and $\boldsymbol{A}^{(2*)}$ satisfy the constraint (4c) with regard to $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$.

Next, by the optimality of $\boldsymbol{A}^{(\lambda*)}$, we have $f(\boldsymbol{x}^{(\lambda)}) = A_L^{(\lambda*)} \geq A_L^{(\lambda)} = \lambda A_L^{(1*)} + (1-\lambda)A_L^{(2*)} = \lambda f(\boldsymbol{x}^{(1)}) + (1-\lambda)f(\boldsymbol{x}^{(2)})$.

In conclusion, $f_v(\boldsymbol{x}_{v,:})$ is a concave function on the range $[0, 1 - A_{v,0}/m]^L$. □

We then outline the optimality condition of the single-variable reformulation in Lemma EC.2. To facilitate the characterization of subgradient, we introduce the convex function $\tilde{f}_v := -f_v$. Let $\partial\tilde{f}_v(\boldsymbol{x}_v)$ be the subgradient set at $\boldsymbol{x}_v$:

LEMMA EC.2 **(Optimality Condition)**. *Given $\theta \geq 0$, the optimal solution $\boldsymbol{x}^*(\theta)$ to the inner maximization problem* (9) *satisfies the following condition.*

$$\forall v \in U, t = 1, 2, \ldots, L, \boldsymbol{g}_{v,:}(\theta) \in \partial\tilde{f}_v(\boldsymbol{x}_v^*(\theta)), \begin{cases} g_{v,t}(\theta) \geq -\theta & \textit{when } x_{v,t}^*(\theta) = 0, \\ g_{v,t}(\theta) = -\theta & \textit{when } 0 < x_{v,t}^*(\theta) < A_{v,0}, \\ g_{v,t}(\theta) \leq -\theta & \textit{when } x_{v,t}^*(\theta) = A_{v,0}. \end{cases} \tag{EC.9}$$

Proof of Lemma EC.2: Define $r_v(\boldsymbol{x}_{v,:};\theta) := -\tilde{f}_v(\boldsymbol{x}_{v,:}) - \theta m\sum_{t=1}^L x_{v,t}$ for all $v \in U$. The problem is separable by content piece v, so we focus on a specific $v \in U$ in the following and omit the subscription for clarity.

When $x_t^*(\theta) = 0$, we construct a feasible solution $\boldsymbol{x}'(\theta) := \boldsymbol{x}^*(\theta) + \epsilon\boldsymbol{e}_t$ where ϵ is a sufficiently small positive constant and $\boldsymbol{e}_t$ is a vector with 1 in the t-th entry and 0 in all other entries. By the concavity of r, we have

$$r(\boldsymbol{x}'(\theta);\theta) \geq r(\boldsymbol{x}^*(\theta);\theta) - (\boldsymbol{g}(\theta) + \theta\boldsymbol{1})^\top(\boldsymbol{x}'(\theta) - \boldsymbol{x}^*(\theta)) = r(\boldsymbol{x}^*(\theta);\theta) - (g_t(\theta) + \theta)\epsilon\,,$$

where $\boldsymbol{1}$ is the all one vector. Given the optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) \geq -\theta$ should hold.

When $x_t^*(\theta) = A_{v,0}$, we construct a feasible solution $\boldsymbol{x}''(\theta) := \boldsymbol{x}^*(\theta) - \epsilon\boldsymbol{e}_t$. By concavity of r, we have

$$r(\boldsymbol{x}''(\theta);\theta) \geq r(\boldsymbol{x}^*(\theta);\theta) - (\boldsymbol{g}(\theta) + \theta\boldsymbol{1})^\top(\boldsymbol{x}''(\theta) - \boldsymbol{x}^*(\theta)) = r(\boldsymbol{x}^*(\theta);\theta) + (g_t(\theta) + \theta)\epsilon\,.$$

Given the optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) \leq -\theta$ should hold.

When $0 < x_t^*(\theta) < A_{v,0}$, we simultaneously construct two feasible solutions $\boldsymbol{x}'(\theta)$ and $\boldsymbol{x}''(\theta)$ as previous. Similarly, by optimality of $\boldsymbol{x}^*(\theta)$, $g_t(\theta) = -\theta$ should hold.

In conclusion, we can characterize this optimality condition based on the subgradient $\boldsymbol{g}(\theta)$ and θ. □

For any $\theta \geq 0$, we denote $\left(\boldsymbol{x}^*_{v,:}(\theta) : v \in \mathcal{V}\right)$ as the optimal solution to (9), and let function $s(\theta; U) := m\sum_{v\in U}\sum_{t=1}^{L} x^*_{v,t}(\theta)$ describe the total optimal promotion times with dual variable θ. In the following, we establish the property of $s(\theta;U)$ for any given candidate set $U \subseteq \mathcal{V}$ in Lemma EC.3.

LEMMA EC.3 **(Monotonicity)**. *For any $U \subseteq \mathcal{V}$, $s(\theta;U)$ is a nonincreasing function in θ.*

Proof of Lemma EC.3: For all $\theta_1, \theta_2 \geq 0$, let $\boldsymbol{g}_v(\theta_1) \in \partial \tilde{f}_v(\boldsymbol{x}^*_{v,:}(\theta_1))$ and $\boldsymbol{g}_v(\theta_2) \in \partial \tilde{f}_v(\boldsymbol{x}^*_{v,:}(\theta_2))$ for all $v \in U$. By Lemma EC.2, for all $v \in U$, we have

$$
\begin{aligned}
&\left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \left(-\boldsymbol{g}_v(\theta_1) + \boldsymbol{g}_v(\theta_2)\right) \\
=&\left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v}(\theta_2)\right)^\top \left(\theta_1 \cdot \mathbf{1} - \theta_2 \cdot \mathbf{1}\right) + \sum_{t=1}^{L}\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(-\theta_1 - g_{v,t}(\theta_1)\right) + \sum_{t=1}^{L}\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(\theta_2 + g_{v,t}(\theta_2)\right) \\
=&(\theta_1-\theta_2)\cdot\left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \mathbf{1} + \sum_{t=1}^{L}\left(\mathbb{1}\{x^*_{v,t}(\theta_1)=0\} + \mathbb{1}\{x^*_{v,t}(\theta_1)=A_{v,0}\}\right)\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(-\theta_1 - g_{v,t}(\theta_1)\right) \\
&+\sum_{t=1}^{L}\left(\mathbb{1}\{x^*_{v,t}(\theta_2)=0\} + \mathbb{1}\{x^*_{v,t}(\theta_2)=A_{v,0}\}\right)\left(x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right)\left(\theta_2 + g_{v,t}(\theta_2)\right).
\end{aligned}
$$

We further discuss the latter two terms. We have

$$
\begin{cases}
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = -x^*_{v,t}(\theta_2) \leq 0, \text{ and } -\theta_1 - g_{v,t}(\theta_1) \leq 0, & \text{when } x^*_{v,t}(\theta_1) = 0,\\
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = A_{v,0} - x^*_{v,t}(\theta_2) \geq 0, \text{ and } -\theta_1 - g_{v,t}(\theta_1) \geq 0, & \text{when } x^*_{v,t}(\theta_1) = A_{v,0};
\end{cases}
$$

$$
\text{and }
\begin{cases}
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = x^*_{v,t}(\theta_1) \geq 0, \text{ and } \theta_2 + g_{v,t}(\theta_2) \geq 0, & \text{when } x^*_{v,t}(\theta_2) = 0,\\
x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2) = x^*_{v,t}(\theta_2) - A_{v,0} \leq 0, \text{ and } \theta_2 + g_{v,t}(\theta_2) \leq 0, & \text{when } x^*_{v,t}(\theta_2) = A_{v,0}.
\end{cases}
$$

Given concavity of f_v, we have $0 \geq \left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \left(-\boldsymbol{g}_v(\theta_1) + \boldsymbol{g}_v(\theta_2)\right) \geq (\theta_1 - \theta_2)\cdot\left(\boldsymbol{x}^*_{v,:}(\theta_1) - \boldsymbol{x}^*_{v,:}(\theta_2)\right)^\top \mathbf{1}$. By summing up over $v \in U$, we have $(\theta_1 - \theta_2)\cdot m\sum_{v\in U}\sum_{t=1}^{L}\left[x^*_{v,t}(\theta_1) - x^*_{v,t}(\theta_2)\right] = (\theta_1-\theta_2)\cdot(h(\theta_1) - h(\theta_2)) \leq 0$. In conclusion, $s(\theta;U)$ is nonincreasing. □

Proof of Lemma 2: We begin by demonstrating that $s(\theta^*(U_2);U_1) \leq s(\theta^*(U_1);U_1)$. We can decompose the function $s(\theta;U_2)$ as $s(\theta;U_1) + s(\theta;U_2\setminus U_1)$. We consider two cases based on the value of $s(\theta^*(U_1);U_1)$:

(i) $s(\theta^*(U_1);U_1) = C$. We directly have $s^*(\theta^*(U_2);U_1) \leq s^*(\theta^*(U_2);U_2) \leq C = s(\theta^*(U_1);U_1)$.

(ii) $s(\theta^*(U_1);U_1) < C$. We show this by contradiction. Assume that $s(\theta^*(U_2);U_1) > s(\theta^*(U_1);U_1)$. We can construct a feasible solution $\boldsymbol{x}'_{U_2,:}$ for the PO problem given candidate set U_2 as

$$
\boldsymbol{x}'_{v,:} = \begin{cases}
\boldsymbol{x}^*_{v,:}(\theta^*(U_1)) & \text{when } v \in U_1,\\
\boldsymbol{x}^*_{v,:}(\theta^*(U_2)) & \text{when } v \in U_2 \setminus U_1.
\end{cases}
$$

The objective value generated by $\boldsymbol{x}'_{U_2,:}$ is larger than $\boldsymbol{x}^*_{U_2,:}(\theta^*(U_2))$, given the optimality of $\boldsymbol{x}^*_{U_1,:}(\theta^*(U_1))$ for the PO problem with candidate set U_1. This contradicts with the optimality of $\boldsymbol{x}^*_{U_2,:}(\theta^*(U_2))$ for the PO problem with candidate set U_2.

Consequently, $s(\theta^*(U_2);U_1) \leq s(\theta^*(U_1);U_1)$. By Lemma EC.3, we conclude that $\theta^*(U_1) \leq \theta^*(U_2)$. □

A.1.4. Submodularity of the CGPO objective.

Proof of Theorem 3: It is easy to verify that $R(U;C) + R(\mathcal{V}\setminus U;0)$ is a monotone function. By (11), we have

$$
\begin{aligned}
&R(U\cup\{w\};C) + R(\mathcal{V}\setminus(U+\{w\});0) = \max_{0\leq c\leq C} R(U;c) + R(\{w\};C-c) + R(\mathcal{V}\setminus(U+\{w\});0)\\
\geq& R(U;C) + R(\{w\};0) + R(\mathcal{V}\setminus(U+\{w\});0) = R(U;C) + R(\mathcal{V}\setminus U;0).
\end{aligned}
$$

Next, we focus on the proof of submodularity. To prove that $R(U;C)+R(\mathcal{V}\setminus U;0)$ is a submodular function, we need to demonstrate that for any given $U_1\subseteq U_2\subseteq\mathcal{V}$ and $w\in\mathcal{V}\setminus U_2$, (10) holds. Therefore, we compare the marginal gain of content piece w when given nested content sets U_1 and U_2 as follows:

$$\begin{aligned}
&R(U_1\cup\{w\};C)-R(U_1;C)-R(\{w\};0)-[R(U_2+\{w\};C)-R(U_2;C)-R(\{w\};0)]\\
=&R(U_1;c^*(U_1))+R(\{w\};C-c^*(U_1))-R(U_1;C)-R(\{w\};0)\\
&-[R(U_2;c^*(U_2))+R(\{w\};C-c^*(U_2))-R(U_2;C)-R(\{w\};0)]\\
\geq&R(U_1;c^*(U_2))+R(\{w\};C-c^*(U_2))-R(U_1;C)-R(\{w\};0)\\
&-[R(U_2;c^*(U_2))+R(\{w\};C-c^*(U_2))-R(U_2;C)-R(\{w\};0)]\\
=&[R(U_1;c^*(U_2))-R(U_1;C)]-[R(U_2;c^*(U_2))-R(U_2;C)]=-\int_{z=c^*(U_2)}^{C}\theta^*(U_1;z)dz+\int_{z=c^*(U_2)}^{C}\theta^*(U_2;z)dz\\
=&\int_{z=c^*(U_2)}^{C}[\theta^*(U_2;z)-\theta^*(U_1;z)]\,dz\geq 0.
\end{aligned}$$

where the first inequality follows since $c^*(U_1)$ is the optimal solution that maximizes problem (11) and the last inequality follows by Lemma 2.

In conclusion, $R(U;C)+R(\mathcal{V}\setminus U;0)$ is a monotone submodular set function. □

A.2. Proofs and Supplements in Section 5

A.2.1. The OLS estimation method for the BDM. According to Bass (1969), the OLS method of the BDM works as follows. Given a sequence of adoption data $\{(a_t,A_t)\}_{t=0}^{T}$, it assumes the following relationship: $a_t=\beta_1+\beta_2\cdot A_{t-1}+\beta_3A_{t-1}^2+\epsilon_t$, where $\beta_1=pm$, $\beta_2=q-p$ and $\beta_3=-q/m$ are three different parameters to estimate; 1, A_{t-1}, and A_{t-1}^2 are considered as three observed covariates; ϵ_t is the independent white noise. For notation simplicity, we denote the covariate matrix and dependent variable as

$$Z=\begin{pmatrix}1 & A_0 & A_0^2\\ 1 & A_1 & A_1^2\\ \vdots & \vdots & \vdots\\ 1 & A_{T-1} & A_{T-1}^2\end{pmatrix}\text{ and }\boldsymbol{a}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_T\end{pmatrix}.$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ can then be derived as $\hat{\boldsymbol{\beta}}=(Z^\top Z)^{-1}Z^\top\boldsymbol{a}$. Consequently, the estimators can be obtained as

$$\hat{m}=\frac{\hat{\beta}_2\pm\sqrt{\hat{\beta}_2^2-4\hat{\beta}_1\hat{\beta}_2}}{2\hat{\beta}_3},\ \hat{p}=\frac{\hat{\beta}_1}{\hat{m}},\ \text{and }\hat{q}=-\hat{\beta}_3\hat{m}.$$

However, these estimators suffer from large variances. In extreme cases (e.g., $\hat{m}=0$), they even become invalid.

A.2.2. Proofs for the asymptotic analysis of the OLS-based estimators. To streamline notation, we define the fixed-design covariate matrix for the n-th diffusion process as

$$Z_{(n)}=\begin{pmatrix}x_{1,(n)} & \bar{A}_{1,(n)}(1-\bar{A}_{1,(n)})\\ x_{2,(n)} & \bar{A}_{2,(n)}(1-\bar{A}_{2,(n)})\\ \vdots & \vdots\\ x_{n,(n)} & \bar{A}_{n,(n)}(1-\bar{A}_{n,(n)})\end{pmatrix}.$$

Proof of Theorem 4: We first consider the n-th D-OLS estimator for q, which is represented as

$$\hat{q}_{(n)}^{\text{D-OLS}}=q+\frac{\sum_{i=1}^{n}\left[A_{i,(n)}(1-x_{i,(n)}-\frac{A_{i,(n)}}{m_{(n)}})\epsilon_{i,(n)}^{\text{i}}\right]}{\sum_{i=1}^{n}\left[A_{i,(n)}(1-x_{i,(n)}-\frac{A_{i,(n)}}{m_{(n)}})\right]^2}=q+\frac{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\text{i}}}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2}.$$

Let $\bar{\boldsymbol{\epsilon}}^{\mathrm{i}}$ denote the noise distribution with finite variance $\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right) < \infty$, we can characterize the mean and variance of D-OLS estimator as follows:

$$\mathbb{E}\left[\hat{q}_{(n)}^{\text{D-OLS}}\right] = q + \frac{\frac{1}{n}\sum_{i=1}^{n} \bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\cdot\mathbb{E}\left[\bar{\epsilon}_{i,(n)}^{\mathrm{i}}\right]}{\frac{1}{n}\sum_{i=1}^{n} \bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2} = q, \text{ and}$$

$$\mathrm{var}\left(\hat{q}_{(n)}^{\text{D-OLS}}\right) = \frac{\frac{1}{n^2}\sum_{i=1}^{n} \bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2\,\mathrm{var}\left(\bar{\epsilon}_{i,(n)}^{\mathrm{i}}\right)}{\left[\frac{1}{n}\sum_{i=1}^{n} \bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2\right]^2} = \frac{\tilde{Q}_{22,(n)}}{nQ_{22,(n)}^2}\,\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right).$$

By Chebyshev's inequality, we have $\Pr\left(|\hat{q}_{(n)}^{\text{D-OLS}} - q| \geq k\right) \leq \frac{\tilde{Q}_{22,(n)}}{k^2 nQ_{22,(n)}^2}\,\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right)$. Taking limits on both sides, we get

$$\lim_{n\to\infty}\Pr\left(|\hat{q}_{(n)}^{\text{D-OLS}} - q| \geq k\right) \leq \lim_{N\to\infty}\frac{1}{k^2}\frac{\tilde{Q}_{22,(n)}\,\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right)}{nQ_{22,(n)}^2} = \frac{\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right)}{k^2}\cdot\frac{\lim_{n\to\infty}\tilde{Q}_{22,(n)}}{\lim_{n\to\infty}Q_{22,(n)}^2}\cdot\lim_{n\to\infty}\frac{1}{n} = \frac{\tilde{Q}_{22}\,\mathrm{var}\left(\bar{\epsilon}^{\mathrm{i}}\right)}{k^2 Q_{22}^2}\cdot\lim_{n\to\infty}\frac{1}{n} = 0,$$

which implies that $\lim_{n\to\infty}\hat{q}_{(n)}^{\text{D-OLS}} = q$. Similarly, we consider the n-th D-OLS estimator for p, which is

$$\hat{p}_{(n)}^{\text{D-OLS}} = p + \frac{\sum_{i=1}^{n}\left[m_{(n)}x_{i,(n)}\left((q-\hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}A_{i,(n)} + \bar{\epsilon}_{i,(n)}^{\mathrm{d}}\right)\right]}{\sum_{t=1}^{n}(m_{(n)}x_{i,(n)})^2} = p + \frac{\frac{1}{n}\sum_{i=1}^{n}\left[x_{i,(n)}\left((q-\hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}\bar{A}_{i,(n)} + \bar{\epsilon}_{i,(n)}^{\mathrm{d}}\right)\right]}{\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}^2}.$$

We consider the following two terms separately:

$$\lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(q-\hat{q}_{(n)}^{\text{D-OLS}})x_{i,(n)}^2\bar{A}_{i,(n)}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right) \text{ and } \lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\mathrm{d}}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right).$$

The first term converges to 0 since

$$\lim_{n\to\infty}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(q-\hat{q}^{\text{D-OLS}})x_{i,(n)}^2\bar{A}_{i,(n)}}{\frac{1}{n}\sum_{t=1}^{n}x_{i,(n)}^2}\right) = \lim_{n\to\infty}\left(q-\hat{q}_{(n)}^{\text{D-OLS}}\right)\cdot\lim_{n\to\infty}\left(\frac{\tilde{Q}_{11,(n)}}{Q_{11,(n)}}\right) = \frac{\tilde{Q}_{11}}{Q_{11}}\lim_{n\to\infty}\left(q-\hat{q}_{(n)}^{\text{D-OLS}}\right) = 0.$$

The second term also converges to 0, similarly as in the proof of $\hat{q}_{(n)}^{\text{D-OLS}}$. As a result, we have $\lim_{n\to\infty}\hat{p}_{(n)}^{\text{D-OLS}} = p$.

In conclusion, D-OLS estimators $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$ are consistent estimators. □

Proof of Theorem 5: We consider two different estimation methods, respectively.

(i) The D-OLS method.

First, we characterize the limiting distribution of $\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)$. We can represent it as

$$\sqrt{n}\left(\hat{q}_{(n)}^{\text{D-OLS}} - q\right) = \frac{\sqrt{n}\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\mathrm{i}}\right]}{\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\right]^2} = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\bar{\epsilon}_{i,(n)}^{\mathrm{i}}\right]}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2}.$$

We consider it as the sum of n independent random variables:

$$\sqrt{n}\left(\hat{q}_{(n)}^{\text{D-OLS}} - q\right) = \sum_{i=1}^{n} w_{i,(n)}\bar{\epsilon}_{i,(n)}^{\mathrm{i}}, \text{ where } w_{i,(n)} = \frac{\frac{1}{\sqrt{n}}\left[\bar{A}_{i,(n)}(1-x_{i,(n)}-\bar{A}_{i,(n)})\right]}{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2}.$$

We then show that this sequence satisfies

$$\lim_{n\to\infty}\max_{i=1,2,\ldots,n}|w_{i,(n)}| \leq \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{1}{\tilde{Q}_{22,(n)}} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\lim_{n\to\infty}\frac{1}{\tilde{Q}_{22,(n)}} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{1}{\tilde{Q}_{22}} = 0.$$

where the inequality follows since $0 \leq x_{i,(n)}, \bar{A}_{i,(n)} \leq 1$ for all $i = 1, 2, \ldots, n$.

This implies that Lindeberg's condition is satisfied. At last, the variance of $\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)$ is as follows:

$$\mathrm{var}\left(\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q)\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2}{\left[\frac{1}{n}\sum_{i=1}^{n}\bar{A}_{i,(n)}^2(1-x_{i,(n)}-\bar{A}_{i,(n)})^2\right]^2}\,\mathrm{var}\left(\bar{\epsilon}_{i,(n)}^{\mathrm{i}}\right) = \frac{1}{\tilde{Q}_{22,(n)}}\eta\sigma^2.$$

Let $\xi_2 = \eta Q_{22}/\tilde{Q}_{22} - 1$. By Lindeberg's central limit theorem, we have

$$\sqrt{n}(\hat{q}_{(n)}^{\text{D-OLS}} - q) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{22,(n)}}(1+\xi_2)\sigma^2\right).$$

Next, we characterize the limiting distribution of $\sqrt{n}(\hat{p}_{(n)}^{\text{D-OLS}} - p)$. We can represent it as

$$\sqrt{n}(\hat{p}_{(n)}^{\text{D-OLS}} - p) = \frac{\sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)} + \sqrt{n}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\sum_{i=1}^n x_{i,(n)}^2}$$
$$= \sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\frac{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} + \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2}.$$

We consider the two terms separately. For the former term, we can easily derive that

$$\sqrt{n}(q - \hat{q}_{(n)}^{\text{D-OLS}})\frac{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2 \bar{A}_{i,(n)}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{\eta \tilde{Q}_{11}^2}{\tilde{Q}_{22} Q_{11}^2}\sigma^2\right).$$

For the latter term, we perform a similar analysis as the previous one and get

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,(n)}\bar{\epsilon}_{i,(n)}^{\text{d}}}{\frac{1}{n}\sum_{i=1}^n x_{i,(n)}^2} \xrightarrow{d} \mathcal{N}\left(0, \frac{1-\eta}{Q_{11}}\sigma^2\right).$$

Let $\xi_1 = \eta(\tilde{Q}_{11}^2/\tilde{Q}_{22}Q_{11} - 1)$. As these two terms are independent, therefore, we can conclude that

$$\sqrt{n}(\hat{p}^{\text{D-OLS}} - p) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{Q_{11}}(1+\xi_1)\sigma^2\right).$$

(ii) The OLS method. For notation simplicity, we write the OLS formulation in matrix form. Let $\boldsymbol{\beta} = (p,q)^\top$ and $\bar{\boldsymbol{\epsilon}}_{(n)} = (\bar{\epsilon}_{1,(n)}, \bar{\epsilon}_{2,(n)}, \ldots, \bar{\epsilon}_{n,(n)})^\top$. Consider the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \boldsymbol{\beta})$, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \beta) = \sqrt{n}\left({Z_{(n)}}^\top Z_{(n)}\right)^{-1} {Z_{(n)}}^\top \bar{\boldsymbol{\epsilon}}_{(n)}.$$

Let $W_{(n)} = \sqrt{n}({Z_{(n)}}^\top Z_{(n)})^{-1} {Z_{(n)}}^\top$ and $\boldsymbol{w}_{i,(n)}$ be the i-th column of $W_{(n)}$ for $i = 1, 2, \ldots, n$. As a consequence, we can write $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \boldsymbol{\beta})$ as a sum of n independent random variables: $\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \beta) = \sum_{i=1}^n \boldsymbol{w}_{i,(n)}\bar{\epsilon}_{i,(n)}$. We then show that this sequence satisfies

$$\lim_{n\to\infty}\max_{i=1,2,\ldots,n}\|\boldsymbol{w}_{i,(n)}\|_2 = \lim_{n\to\infty}\|W_{(n)}\|_{\infty,2} = \lim_{n\to\infty}\left\|\left(\frac{1}{n}{Z_{(n)}}^\top Z_{(n)}\right)^{-1}\frac{1}{\sqrt{n}}{Z_{(n)}}^\top\right\|_{\infty,2}$$
$$= \lim_{n\to\infty}\left\|{Q_{(n)}}^{-1}\frac{1}{\sqrt{n}}{Z_{(n)}}^\top\right\|_{\infty,2} \le \lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|{Q_{(n)}}^{-1}\right\|_{2,\infty}\left\|{Z_{(n)}}^\top\right\|_{2,2}$$
$$\le \lim_{n\to\infty}\frac{1}{\sqrt{n}}\left\|{Q_{(n)}}^{-1}\right\|_{2,\infty} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\lim_{n\to\infty}\left\|{Q_{(n)}}^{-1}\right\|_{2,\infty} = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\left\|Q^{-1}\right\|_{2,\infty} = 0,$$

where the first inequality follows since the definition of matrix operator norm and the second inequality follows since $0 \le x_{i,(n)}, \bar{A}_{i,(n)} \le 1$ for all $i = 1, 2, \ldots, n$.

This implies that Lindeberg's condition is satisfied. Then, we calculate the variance of $\sqrt{n}(\hat{\boldsymbol{\beta}}^{\text{OLS}} - \beta)$ as follows:

$$\text{var}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \beta)\right) = \mathbb{E}\left[W_{(n)}\bar{\boldsymbol{\epsilon}}_{(n)}{\bar{\boldsymbol{\epsilon}}_{(n)}}^\top {W_{(n)}}^\top\right] = \sigma^2 W_{(n)}{W_{(n)}}^\top = n\left({Z_{(n)}}^\top Z_{(n)}\right)^{-1} = {Q_{(n)}}^{-1}.$$

By Lindeberg's central limit theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{(n)}^{\text{OLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(0, Q^{-1}\right), \text{ where } Q^{-1} = \begin{bmatrix} \frac{1}{Q_{11}}\left(1 + \frac{{Q_{12}}^2}{Q_{11}Q_{22} - Q_{12}^2}\right) & -\frac{Q_{12}}{Q_{11}Q_{22} - Q_{12}^2} \\ -\frac{Q_{12}}{Q_{11}Q_{22} - Q_{12}^2} & \frac{1}{Q_{22}}\left(1 + \frac{Q_{12}^2}{Q_{11}Q_{22} - Q_{12}^2}\right) \end{bmatrix}.$$

At last, we conclude that (i) The asymptotic variances of $\hat{p}^{\text{D-OLS}}$ and $\hat{q}^{\text{D-OLS}}$ are $(1+\xi_1)\sigma^2/Q_{11}$ and $(1+\xi_2)\sigma^2/Q_{22}$, where $\xi_1 = \eta(\tilde{Q}_{11}^2/\tilde{Q}_{22}Q_{11} - 1)$ and $\xi_2 = \eta Q_{22}/\tilde{Q}_{22} - 1$. (ii) The asymptotic variances of $\hat{p}^{\text{OLS}}$ and $\hat{q}^{\text{OLS}}$ are $(1+\kappa)\sigma^2/Q_{11}$ and $(1+\kappa)\sigma^2/Q_{22}$, where $\kappa = Q_{12}^2/|Q|$. □

Proposition EC.1. *When $\eta \le \tilde{Q}_{22}/Q_{22}$, we can show that the asymptotic variances of D-OLS estimators are smaller than those of OLS estimators, that is, $\xi_1 \le \kappa$ and $\xi_2 \le \kappa$.*

Proof for Proposition EC.1: Particularly, we have

$$\kappa-\xi_1=\frac{Q_{12}^2}{Q_{11}Q_{22}-Q_{12}^2}-\eta\left(\frac{\tilde{Q}_{11}^2}{Q_{11}\tilde{Q}_{22}}-1\right)\geq\frac{Q_{12}^2}{Q_{11}Q_{22}-Q_{12}^2}-\frac{\tilde{Q}_{11}^2}{Q_{11}Q_{22}}\geq\frac{Q_{12}^2-\tilde{Q}_{11}^2}{Q_{11}Q_{22}}=\frac{Q_{12}+\tilde{Q}_{11}}{Q_{11}Q_{22}}\left(Q_{12}-\tilde{Q}_{11}\right)$$
$$=\frac{Q_{12}+\tilde{Q}_{11}}{Q_{11}Q_{22}}\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}A_{i,(n)}(1-A_{i,(n)})-\frac{1}{n}\sum_{i=1}^{n}x_{i,(n)}^2A_{i,(n)}\right)$$
$$\geq\frac{Q_{12}+\tilde{Q}_{11}}{Q_{11}Q_{22}}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(x_{i,(n)}A_{i,(n)}\cdot x_{i,(n)}-x_{i,(n)}^2A_{i,(n)}\right)=0.$$

where the first inequality follows since $0\leq\eta\leq\tilde{Q}_{22}/Q_{22}$, the second inequality follows since $Q_{12}^2\geq 0$ and the third inequality follows since $x_{i,(n)}\leq 1-\bar{A}_{i,(n)}$. Furthermore, we have

$$\kappa-\xi_2=\frac{Q_{12}^2}{|Q|}-\left(\eta\frac{Q_{22}}{\tilde{Q}_{22}}-1\right)\ \geq\ \kappa\ \geq\ 0.$$

where the first inequality follows since $\eta\leq\tilde{Q}_{22}/Q_{22}$ and the second inequality follows by definition. □

A.2.3. Proofs for the MLE estimators.

Proposition EC.2. *When platforms cannot observe adopter types, the log-likelihood function $\mathcal{LL}^{MLE}(p,q)$ is concave.*

Proof of Proposition EC.2: To show that $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave, it is sufficient to show that the corresponding Hessian matrix is negative semi-definite.

Let $g_t=px_t/(1-A_{t-1}/m)+qA_{t-1}/m$. The partial derivatives of $\mathcal{LL}^{\text{MLE}}(p,q)$ with regard to p and q are

$$\frac{\partial\mathcal{LL}^{\text{MLE}}(p,q)}{\partial p}=\sum_{t=1}^{T}\frac{\partial g_t}{\partial p}\left(\frac{a_t}{g_t}-\frac{m-A_{t-1}-a_t}{1-g_t}\right)\quad\text{and}\quad\frac{\partial\mathcal{LL}^{\text{MLE}}(p,q)}{\partial q}=\sum_{t=1}^{T}\frac{\partial g_t}{\partial q}\left(\frac{a_t}{g_t}-\frac{m-A_{t-1}-a_t}{1-g_t}\right).$$

Let $\boldsymbol{H}_{\mathcal{LL}}$ and $\boldsymbol{H}_{g_t}$ be the Hessian matrices of $\mathcal{LL}^{\text{MLE}}$ and g_t, respectively. We have

$$\boldsymbol{H}_{\mathcal{LL}}=\sum_{t=1}^{T}\left[-\left(\frac{a_t}{g_t^2}+\frac{m-A_{t-1}-a_t}{(1-g_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial g_t}{\partial p}\\ \frac{\partial g_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial g_t}{\partial p} & \frac{\partial g_t}{\partial q}\end{pmatrix}+\left(\frac{a_t}{g_t}-\frac{m-A_{t-1}-a_t}{1-g_t}\right)\cdot\boldsymbol{H}_{g_t}\right].$$

Since $\boldsymbol{H}_{g_t}$ is a zero matrix and $a_t/g_t^2+(m-A_{t-1}-a_t)/(1-g_t)^2\geq 0$ always holds, we can conclude that $\boldsymbol{H}_{\mathcal{LL}}$ is negative semi-definite which implies $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave.

In conclusion, the log-likelihood function $\mathcal{LL}^{\text{MLE}}(p,q)$ is concave. □

Proposition EC.3. *When platforms can observe adopter types, the log-likelihood function $\mathcal{LL}^{D\text{-}MLE}(p,q)$ is concave.*

Proof of Proposition EC.3: To show that $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave, it is sufficient to show that the corresponding Hessian matrix is negative semi-definite.

Let $g_t=qA_{t-1}/m$ and $h_t=p+qA_{t-1}/m$. The partial derivatives of $\mathcal{LL}^{\text{D-MLE}}(p,q)$ with regard to p and q are

$$\frac{\partial\mathcal{LL}^{\text{D-MLE}}(p,q)}{\partial p}=\sum_{t=1}^{T}\left[\frac{\partial g_t}{\partial p}\left(\frac{a_t^{\text{i}}}{g_t}-\frac{m-A_{t-1}-mx_t-a_t^{\text{i}}}{1-g_t}\right)+\frac{\partial h_t}{\partial p}\left(\frac{a_t^{\text{d}}}{h_t}-\frac{mx_t-a_t^{\text{d}}}{1-h_t}\right)\right],\ \text{and}$$
$$\frac{\partial\mathcal{LL}^{\text{D-MLE}}(p,q)}{\partial q}=\sum_{t=1}^{T}\left[\frac{\partial g_t}{\partial q}\left(\frac{a_t^{\text{i}}}{g_t}-\frac{m-A_{t-1}-mx_t-a_t^{\text{i}}}{1-g_t}\right)+\frac{\partial h_t}{\partial q}\left(\frac{a_t^{\text{d}}}{h_t}-\frac{mx_t-a_t^{\text{d}}}{1-h_t}\right)\right].$$

Let $\boldsymbol{H}_{\mathcal{LL}}$, $\boldsymbol{H}_{g_t}$ and $\boldsymbol{H}_{h_t}$ be the Hessian matrices of $\mathcal{LL}^{\text{D-MLE}}$, g_t and h_t, respectively. We have

$$\boldsymbol{H}_{\mathcal{LL}}=\sum_{t=1}^{T}\left[-\left(\frac{a_t^{\text{i}}}{g_t^2}+\frac{m-A_{t-1}-mx_t-a_t^{\text{i}}}{(1-g_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial g_t}{\partial p}\\ \frac{\partial g_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial g_t}{\partial p} & \frac{\partial g_t}{\partial q}\end{pmatrix}+\left(\frac{a_t^{\text{i}}}{g_t}-\frac{m-A_{t-1}-mx_t-a_t^{\text{i}}}{1-g_t}\right)\cdot\boldsymbol{H}_{g_t}\right]$$
$$+\sum_{t=1}^{T}\left[-\left(\frac{a_t^{\text{d}}}{h_t^2}+\frac{mx_t-a_t^{\text{d}}}{(1-h_t)^2}\right)\cdot\begin{pmatrix}\frac{\partial h_t}{\partial p}\\ \frac{\partial h_t}{\partial q}\end{pmatrix}\begin{pmatrix}\frac{\partial h_t}{\partial p} & \frac{\partial h_t}{\partial q}\end{pmatrix}+\left(\frac{a_t^{\text{d}}}{g_t}-\frac{mx_t-a_t^{\text{d}}}{1-g_t}\right)\cdot\boldsymbol{H}_{h_t}\right].$$

Since $\boldsymbol{H}_{g_t}$ and $\boldsymbol{H}_{h_t}$ are zero matrices and $a_t^{\text{i}}/g_t^2+(m-A_{t-1}-mx_t-a_t^{\text{i}})/(1-g_t)^2\geq 0$ and $a_t^{\text{d}}/h_t^2+(mx_t-a_t^{\text{d}})/(1-h_t)^2\geq 0$ always hold, we can conclude that $\boldsymbol{H}_{\mathcal{LL}}$ is negative semi-definite which implies $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave.

In conclusion, the log-likelihood function $\mathcal{LL}^{\text{D-MLE}}(p,q)$ is concave. □

Appendix B: Supplements for Numerical Experiments

B.1. Model Calibration

B.1.1. The BDM for online content adoption. This section complements our discussions on the discrepancy on the BDM and actual adoption data for online content. A common issue to fit the BDM is the notable underestimation of the diffusion coefficient q, exemplified in our case study shown in Figure 4(a). In some cases, this coefficient is even negative, as seen in Figure 4(b), leading to a deviation of the fitted BDM curve from its typical S-shaped configuration.

It is important to emphasize that these variations, while significant, do not contradict the theoretical foundation of the BDM. To clarify this point, we illustrate the complete trajectory of the fitted BDM curve, extended beyond the time horizon of our observation, for our motivating example in Figure EC.1. In Figures 1(a) and 4(a), we have

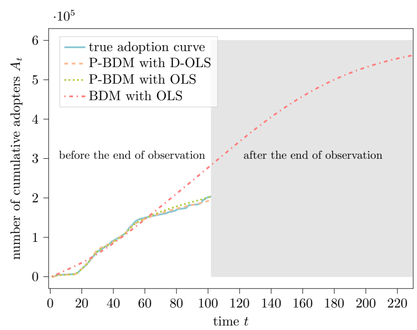


Figure EC.1 Illustration of the complete fitted BDM curve to the actual adoption for the motivating example.

only included a segment of the BDM curve that fits within the observed time frame. In Figure EC.1, the complete fitted BDM curve gradually presents an S-shape. Despite this delayed emergence of the S-shape, the BDM curve significantly diverges from the actual pattern of online content adoption. This discrepancy highlights the limitations of the BDM model in accurately capturing the pattern of online content adoption and underscores the necessity of adopting a modified model, such as the P-BDM, for a more precise representation of these dynamics.

B.1.2. Timeliness of online content diffusion. In this section, we will explore the concept of timeliness in online content and how it affects the diffusion process, resulting in a time-decay factor. We will begin by presenting our findings from data analysis and then modify the P-BDM to incorporate the time-decay factor for a better fit.

Online platforms operate in a highly dynamic and fast-paced environment, with new content being created and shared at a rapid rate. Compared to traditional markets, online platforms have a faster speed of information dissemination. As a result, the timeliness of online content plays a critical role. For example, a review video of a new movie will lose its relevance and generate fewer adoptions as the movie becomes older and less popular. Our analysis of the dataset from the video-sharing platform confirms this phenomenon. To demonstrate this, we calculate two ratios to characterize the promotion and diffusion effects:

$$\frac{a^{\mathrm{d}}_{v,t}}{m x_{v,t}} - \frac{a^{\mathrm{i}}_{v,t}}{m - A_{v,t-1}} \text{ and } \frac{a^{\mathrm{i}}_{v,t}}{\frac{A_{v,t-1}}{m}(m - A_{v,t-1})}. \tag{EC.10}$$

We would like to remark that our goal with this analysis is not to calculate precise values of p and q, but rather to provide insight into the trends of both effects in the real world. As shown in Figure EC.2, the average values of these two ratios among all content pieces are presented against the time from $t = 1$ to $t = 50$. It is apparent from Figure EC.2 that the diffusion effect exhibits a time-decay trend, while the promotion effect remains nearly constant

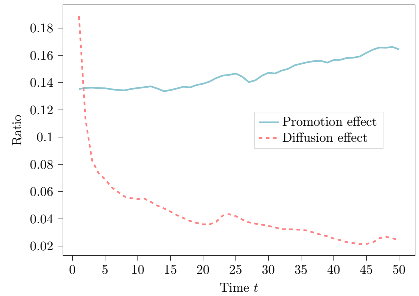


Figure EC.2 Illustration of the trends of promotion and diffusion effects. The x-axis represents the time elapsed since the video was uploaded to the platform and the y-axis represents the average values of the ratios as shown in (EC.10) among all videos at the same time step t.

throughout the entire horizon. The sensitivity to timeliness is primarily observed in the diffusion effect. As such, incorporating the time-decay factor in diffusion modeling is critical to accurately capture the content adoptions for online platforms.

Recall that in the P-BDM dynamic (2), the diffusion effect is proportional to the cumulative adopter number, given by qA_{t-1}/m. To incorporate the timeliness of online content diffusion, we introduce a time-decay multiplicative factor γ where $0 < \gamma \leq 1$. Specifically, we consider the diffusion effect to be $q\gamma^{t-1}A_{t-1}/m$ instead. Therefore, the P-BDM with a time-decay factor can be shown as follows:

$$a_t = \underbrace{\left(p + q\frac{\gamma^{t-1}A_{t-1}}{m}\right)mx_t}_{\text{Direct adopters}} + \underbrace{q\frac{\gamma^{t-1}A_{t-1}}{m}(m - A_{t-1} - mx_t)}_{\text{Indirect adopters}} = \underbrace{pmx_t}_{\text{Promotion effect}} + \underbrace{q\frac{\gamma^{t-1}A_{t-1}}{m}(m - A_{t-1})}_{\text{Diffusion effect}}. \tag{EC.11}$$

This model uses the time-decay factor to characterize the decreasing incentive to diffuse the content as time elapses since its upload. When $\gamma = 1$, this model is equivalent to the original P-BDM.

We make two remarks here. First, the exponent of γ is related to the time elapsed since the content is uploaded. It should be distinguished from the subscript t in the CGPO problem, where the latter is used to denote the time since the beginning of the L planning period. Second, when γ is given, all the results in Sections 4 and 5 still hold. From the optimization perspective, it serves as a known parameter in the CGPO problem and does not change the underlying optimal structure. From the estimation perspective, it requires preprocessing of the observations, but the same estimation methods and analyses can be applied. Therefore, the P-BDM with a time-decay factor does not add to the difficulty of the entire problem but provides flexibility in characterizing the true adoption processes.

B.1.3. Group Estimation. In the context of online platforms, estimating parameters for each content piece individually is usually impractical because of the mountainous amount of videos and the scarcity of data pertaining to a video. What makes things worse is that we often have to make promotion decisions at the early stage of a video's life cycle with minimal data available for estimation. Consequently, it is reasonable to group or cluster the videos using features, and then estimate the parameters to make sure that past estimates can be generalized to future videos and the results are precise. Due to the lack of contextual information, we focus on using category labels for estimation.

The group estimation procedure is similar to that of a single piece, except that the observations are expanded to include all content pieces in the group. Let $\mathcal{V}^c \subseteq \mathcal{V}$ be the set of content pieces in the group c. Therefore, the observations for a group can be represented as $\cup_{v\in\mathcal{V}^c}\{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$. The OLS-based and the MLE-based methods can be readily applied.

B.1.4. Calibration process. For each video category $c \in \mathcal{C}$ within the dataset, we split the observations into training, validation and test sets, separately. To avoid data corruption, we split the data based on video granularity, using a 60-20-20 split. Specifically, for the video set $\mathcal{V}^c \subseteq \mathcal{V}$ corresponding to category c, we randomly select 60% of the videos $v \in \mathcal{V}^c$ and assign the associated observations $\{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$ to the training set. The remaining videos are also randomly split into 20% and 20% for validation and test sets, respectively.

To summarize the calibration process, we present Algorithm 2. We make a remark here, for each video $v \in \mathcal{V}$, we only include observations when the promotion fraction $x_{v,t}$ is positive in our training, validation, and test sets.

Algorithm 2: Calibration process with time-decay factor and group estimation.

for $c \in \mathcal{C}$ **do**
    Randomly split the video set $\mathcal{V}^c$ into $\mathcal{V}^c_{\text{train}}$, $\mathcal{V}^c_{\text{valid}}$ and $\mathcal{V}^c_{\text{test}}$, using a 60-20-20 split.
    Training set $\mathcal{D}_{\text{train}} := \cup_{v \in \mathcal{V}^c_{\text{train}}} \{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
    Valid set $\mathcal{D}_{\text{valid}} := \cup_{v \in \mathcal{V}^c_{\text{valid}}} \{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
    Test set $\mathcal{D}_{\text{test}} := \cup_{v \in \mathcal{V}^c_{\text{test}}} \{(a^{\mathrm{d}}_{v,t}, a^{\mathrm{i}}_{v,t}, A_{v,t}, x_{v,t})\}_{t=1,2,\ldots,T_v}$.
end
for $\gamma \in \{\gamma_1, \gamma_2, \ldots\}$ **do**
    for $c \in \mathcal{C}$ **do**
        Obtain $\hat{p}^c(\gamma)$ and $\hat{q}^c(\gamma)$ based on the training set $\mathcal{D}^{\text{train}}$ and time-decay factor γ.
        for $v \in \mathcal{V}^c_{\text{valid}}$ **do**
            Use $\hat{p}^c(\gamma)$ and $\hat{q}^c(\gamma)$ to predict adoptions as $\{\hat{a}_{v,t}\}_{t=1,2,\ldots T_v}$.
            $\text{WMAPE}_v(\gamma) := \sum_{t=1}^{T_v} |a_{v,t} - \hat{a}_{v,t}| / \sum_{t=1}^{T_v} a_{v,t} \times 100\%$.
        end
    end
    $\text{WMAPE}(\gamma) := \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \text{WMAPE}_v(\gamma)$.
end
$\gamma^* := \arg\max_\gamma \text{WMAPE}(\gamma)$.
for $c \in \mathcal{C}$ **do**
    for $v \in \mathcal{V}^c_{\text{test}}$ **do**
        Use $\hat{p}^c(\gamma^*)$ and $\hat{q}^c(\gamma^*)$ to predict adoptions as $\{\hat{a}_{v,t}\}_{t=1,2,\ldots T_v}$.
    end
end

We evaluate the calibration performance using the weighted mean absolute percentage error (WMAPE). In Figure EC.3, we show the WMAPE we obtain by calibrating the P-BDM with different time-decay factor γ using the D-OLS method. The minimum WMAPE is achieved when $\gamma = 0.983$.

For the sake of completeness, we also include the calibration results in Figure EC.4 when the timeliness is ignored (i.e., $\gamma = 1$). We observe that the estimated diffusion coefficient q is smaller in this case to account for the time decay in diffusion. However, the average out-of-sample WMAPE is 42.92%, which is 10% larger than when $\gamma = 0.983$. The average out-of-sample WMAPEs of the P-BDM with OLS and the BDM are 43.53% and 81.25%, respectively.

B.2. Supplementary Analysis of the AGA Policy with $L = 13$

In this section, we provide a supplementary analysis of the AGA policy. We begin by examining the AGA policy across different lifetimes, followed by the detailed procedures of the sensitivity analysis and K-Means clustering analysis.

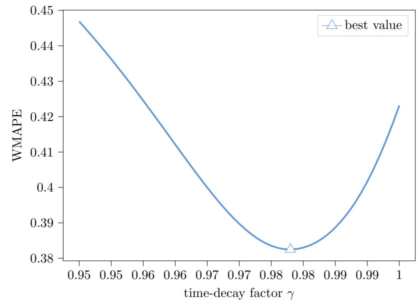


Figure EC.3 WMAPE of the validation set against time-decay factor γ.

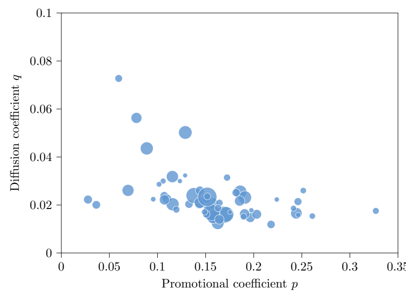
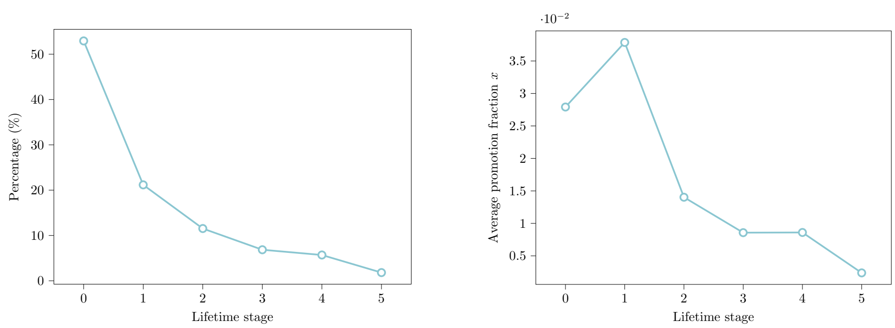


Figure EC.4 Distribution of estimated coefficients when the timeliness of content diffusion is ignored ($\gamma = 1$).



(a) Percentage of allocated budget (b) Average promotion fraction

Figure EC.5 Promotion budget allocation of the AGA policy for videos at different lifetime stages.

B.2.1. The AGA policy across different lifetime stages. We begin by presenting the AGA policy across different lifetime stages in Figure EC.5. As shown in Figure 5(a), the policy primarily promotes videos in their initial stages, dedicating about 53% of the overall budget to videos that have no adopters (stage 0). This heavy initial promotion indicates the policy aims at sparking interest in new content. Figure 5(b) further emphasizes this point by showing that, on average, the promotion fraction allocated to videos tends to decline as the lifetime progresses. However, an exception can be observed at stage 1. This anomaly occurs because only a subset of videos is advanced to later stages after the initial promotion at stage 0. It implies that the policy also acts as a filter or selection mechanism, deciding if a video shows enough promise for further promotion. As a result, the policy generally allocates the promotion budget to videos that show considerable potential in their early stages.

B.2.2. Sensitivity analysis of promotion and diffusion coefficients. In order to understand the relationship between the promotion policy and the characteristics of videos, we perform sensitivity analysis for some important characteristics. For this purpose, we specify the following regression model for the allocated promotion fraction $x_{c,v,t}$, where each observation is a video v at the beginning of time t in the experiment with promotion budget $\bar{C} = c$:

$$x_{c,v,t} = \beta_0 + \beta_1 p_v + \beta_2 q_v + \beta_3 \bar{A}_{v,t-1} + \beta_4 c + \epsilon_{v,t}. \tag{EC.12}$$

where the adoption number is normalized by the market size m before entering the regression. The coefficient $\boldsymbol{\beta}$ in (EC.12) can therefore be used to represent the impact of each characteristic on the allocated promotion fraction.

We conduct the regression on the observations from the previous AGA experiments with $L = 13$. Furthermore, in order to illustrate the difference of policy for videos at different lifetime, we perform regression within each lifetime stage separately. Table EC.1 reports the regression results. Although the R^2 is small for all the regressions, indicating

Table EC.1 Regression results of the promotion fraction of the AGA policy with regard to video characteristics

	# of Obs.	R^2	β_0 (const)	β_1 (p)	β_2 (q)	β_3 ($\bar{A}$)	β_4 ($\bar{C}$)
stage 0	72,867	0.044	-0.1412****	0.8243****	0.6687****	-	0.0055****
			(0.0043)	(0.0216)	(0.0377)	-	(0.0002)
stage 1	21,490	0.418	-1.1621****	5.8253****	3.7211****	-0.7627****	0.0576****
			(0.0245)	(0.1223)	(0.1035)	(0.0195)	(0.0245)
stage 2	31,610	0.209	-0.4399****	2.3614****	1.3477****	-0.1820****	0.0121****
			(0.0145)	(0.0794)	(0.0527)	(0.0101)	(0.0004)
stage 3	30,733	0.125	-0.2060****	0.8130****	0.5548****	0.0563****	0.0039****
			(0.0079)	(0.0304)	(0.0248)	(0.0046)	(0.0002)
stage 4	25,460	0.213	-0.1825****	1.0395****	0.7441****	-0.0714****	0.0048****
			(0.0055)	(0.0298)	(0.0247)	(0.0055)	(0.0002)
stage 5	29,340	0.085	0.0395****	0.1499****	0.1519****	-0.0843****	0.0003****
			(0.0021)	(0.0068)	(0.0080)	(0.0039)	(0.0001)

Note: Robust standard errors are reported in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$. For stage 0, $\bar{A}$ is not included in the regression since it takes 0 value for all observations.

that the linear regression model (EC.12) is not a good representation of the complicated AGA policy, all coefficients are significant at the significance level of 0.0001. Therefore, we consider the values of regression coefficients can represent the impact of video characteristics on the promotion policy.

B.2.3. K-Means clustering for the promotion policy. In order to further understand the AGA policy over the entire lifetime as a whole, we perform K-Means clustering on the average policy for different video configurations. Let $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$ be the set of lifetime stages. To summarize the clustering process, we present Algorithm 3.

As shown in Algorithm 3, the clustering is solely based on the allocated promotion fraction generated by the AGA policy, and the promotion and diffusion coefficients are not explicitly involved. In Figure EC.6, we show the average promotion policies for different video categories by their clusters.

Algorithm 3: K-Means clustering for the promotion policy of different video categories.

for $c \in \mathcal{C}$ **do**
  Classify observations $\cup_{v\in\mathcal{V}^c}\{A_{v,t-1}, x_{v,t}\}_{t=1,2,\dots,T}$ into different lifetime stages by $\bar{A}_{v,t-1}$.
  for $s \in \mathcal{S}$ **do**
    Let $\mathcal{X}_s$ be the set of promotion fractions for observations at stage s.
    $\tilde{x}_s^c := \sum_{x\in\mathcal{X}_s} x/|\mathcal{X}_s|$.
  end
  $\tilde{\boldsymbol{x}}^c := (\tilde{x}_s^c)_{s\in\mathcal{S}}^\top$. `// average promotion policy for category` c.
end
$\tilde{X} := (\tilde{\boldsymbol{x}}^c)_{c\in\mathcal{C}}$. `// feature matrix for all categories.`
Impute the missing values matrix $\tilde{X}$ using k-Nearest Neighbors, with $k=2$.
Perform K-Means clustering based on $\tilde{X}$.

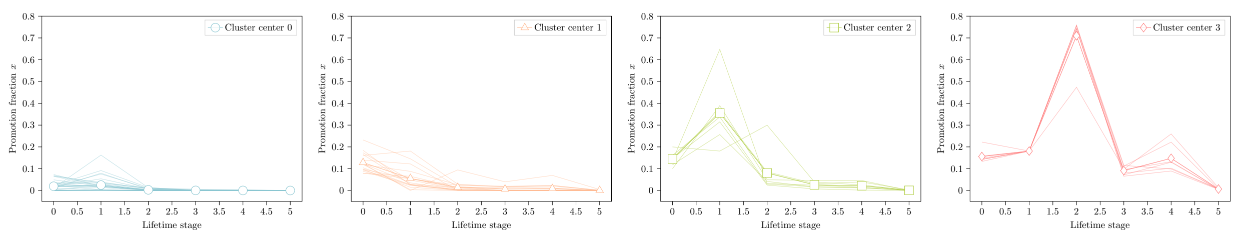


Figure EC.6 Average promotion policies of different video categories (each subfigure represents a cluster).