Optimizing Marketing Subsidies via Counterfactual Learning with Asymmetric Reward Function

Xiang Li
Peking University
Beijing, China
lilixiang222@gmail.com

Yanghao Xiao
University of Chinese Academy of Sciences, Beijing, China
xiaoyanghao22@mails.ucas.ac.cn

Chunyuan Zheng
Peking University
Beijing, China
cyzheng@stu.pku.edu.cn

Qian Zou
Meituan
Beijing, China
zouqian03@meituan.com

Bing Cheng
Meituan
Beijing, China
bing.cheng@meituan.com

Wei Lin
Meituan
Beijing, China
lwsaviola@163.com

Haoxuan Li*
State Key Lab of General AI, School of Intelligence Science and Technology, Peking University, Beijing, China
hxli@stu.pku.edu.cn

Zhouchen Lin*
State Key Lab of General AI, School of Intelligence Science and Technology, Peking University, Beijing, China
zlin@pku.edu.cn

Abstract

In marketing, optimizing subsidy allocation to maximize overall profits is of substantial economic importance. Prior research has employed treatment effect estimation techniques to identify subsidy-sensitive items and design corresponding allocation strategies. However, more accurate treatment effect estimations do not necessarily lead to better allocations, underscoring the critical influence of decision boundaries in decision-making. This paper argues that optimal allocation fundamentally depends on predicting the expected optimal subsidy, a challenge distinct from conventional treatment effect estimation or causal decision-making, which existing approaches fail to address. To fill this gap, we introduce a two-stage **Co**unterfactual optimal subsidy **L**earning method with an **A**symmetric reward (**CoLA**). In the first stage, we derive a coarse estimate of the expected subsidy threshold by exploiting order information and the conditional independence between expected and observed subsidies. In the second stage, we refine these estimates using an asymmetric loss function, leading to more robust predictions. Under practical budget constraints, we prioritize candidates based on their Sharpe ratios to determine the final subsidy allocation strategy. Experiments on three public datasets and an online A/B test show that our method achieves significant performance improvements, yielding the highest total profit and incremental leverage ratios.

CCS Concepts

• **Information systems → Information retrieval**; • **Computing methodologies → Machine learning**; • **Applied computing → Electronic commerce**.

Keywords

Online Subsidy Allocation, Uplift Model, Causal Effect Estimation

ACM Reference Format:
Xiang Li, Yanghao Xiao, Chunyuan Zheng, Qian Zou, Bing Cheng, Wei Lin, Haoxuan Li*, and Zhouchen Lin*. 2026. Optimizing Marketing Subsidies via Counterfactual Learning with Asymmetric Reward Function. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '26), July 20–24, 2026, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3805712.3809750

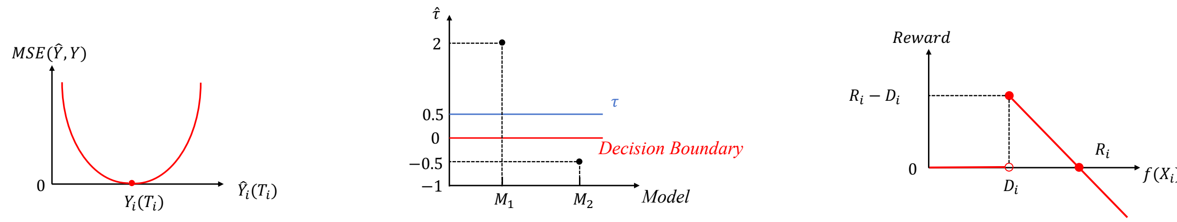
1 Introduction

Advanced marketing technologies are increasingly adopted across industries, with online platforms using subsidies such as coupons or cash incentives to actively stimulate conversion of item [10, 35, 61, 64]. A key challenge in marketing is determining how to allocate the budget for item-level subsidies to achieve the highest returns [73, 74]. In essence, subsidy allocation is closely coupled with recommendation systems [29, 30, 39, 60, 69], as the recommender controls item exposure while the subsidy module determines the attached incentive, and both jointly drive conversion behavior. An effective subsidy allocation aims to predict the optimal subsidy for each item so as to maximize the platform's total profit. To address the problem, prior research building on causal inference principles conceptualizes subsidies as treatment and predicts treatment effects to identify items that benefit most [19, 36, 70]. Based on the identified marketing-sensitive items, it enables the development of marketing strategies that enhance the effectiveness of subsidy allocation to achieve maximum revenue.

*Haoxuan Li and Zhouchen Lin are the corresponding authors.

SIGIR '26, Melbourne, VIC, Australia.

ACM ISBN 979-8-4007-2599-9/2026/07
https://doi.org/10.1145/3805712.3809750



(a) Treatment Effect Estimation. **(b) Causal Decision Making.** **(c) Optimal Subsidy Prediction.**

Figure 1: The comparison of treatment effect estimation, causal decision making and optimal subsidy prediction problem. (a) In treatment effect estimation, symmetric regression loss is used for potential outcome modeling. (b) In causal decision making, decision boundary plays a more crucial role, where τ and $\hat{\tau}$ are the true and estimated causal effect, respectively. (c) In optimal subsidy prediction, the target is to estimate expected subsidy D_i for profit maximization.

Prior studies simplify subsidy allocation to treatment effect estimation, focusing on accurate effect prediction. For instance, in binary settings where treatment is defined as whether to allocate a subsidy, various causal inference methods have been applied, including propensity score based methods [9, 44, 46], tree based methods [7, 17, 21], representation learning methods [24, 48], and generative model based methods [37, 62]. In continuous treatment scenarios, where the treatment is defined as the specific amount of subsidy allocated, also known as dose-response curve estimation, existing methods include Generalized Propensity Score (GPS) based methods [38, 42], DRNet [47], VCNet [40], and SCIGAN [5]. Based on predicted treatment effects, items are grouped and subsidies are allocated only to those with positive profit gains.

However, recent literature suggests that more accurate treatment effect estimations do not necessarily improve the decision making [12]. The key insight is that causal decision making belongs to a classification task based on decision boundaries, while treatment effect estimation is a regression task towards the ground truth effect [13]. In this work, we argue that optimal subsidy prediction in marketing is fundamentally neither a treatment effect estimation problem nor a causal decision-making problem. The essence of this task is to predict the expected subsidy threshold, specifically the optimal subsidy required to trigger item conversion. Concretely, if the predicted subsidy falls below a certain threshold, the subsidy fails to convert the item, resulting in zero reward; if it exceeds the expected subsidy threshold, the subsidy successfully converts the item, but the reward diminishes as the subsidy increases. In summary, optimal subsidy prediction relies on an asymmetric reward function based on the expected optimal subsidy, whereas treatment effect estimation is grounded in symmetric regression losses, and causal decision making centers on identifying decision boundaries. We illustrate the distinctions among these three problems in Figure 1. Recognizing these differences, existing methods based on treatment effect estimation or causal decision-making cannot achieve optimal subsidy prediction.

To address this gap, we propose a two-stage **Co**unterfactual robust optimal subsidy **L**earning framework with an **A**symmetric reward function (**CoLA**). Note that precisely estimating the expected subsidy threshold is inherently challenging, our method focuses on generating robust subsidy predictions that ideally fall within the right-side neighborhood of this unknown threshold. In the first stage, we integrate order information with the conditional independence between expected and observed subsidies to obtain a coarse estimate of the expected subsidy threshold. In the second stage, we refine these estimates using an asymmetric loss function, intentionally producing slightly overestimated predictions. This design encourages predictions to reside in the right-side neighborhood of the true optimal expected subsidy. Given practical budget constraints, we compute the Sharpe ratio of predicted subsidies to measure unit return and allocate subsidies to items with the highest ratios. We conduct semi-simulated experiments on three public datasets and an online A/B test. Results demonstrate that the proposed method consistently outperforms existing causal effect estimation approaches, delivering superior subsidy allocation strategies. The key contributions of this study are summarized as follows.

- To the best of our knowledge, this is the first work to highlight that optimal subsidy allocation in marketing fundamentally relies on predicting the expected subsidy threshold and cannot be reduced to a treatment effect estimation or causal decision-making problem.
- We propose a two-stage counterfactual robust optimal subsidy learning method grounded in asymmetric reward function to produce robust predictions expected to lie within the right-side neighborhood of the true expected threshold.
- Extensive experiments on three public datasets and an online A/B test demonstrate that our method surpasses prior treatment effect estimation-based approaches, delivering superior subsidy allocation strategies.

2 Related Work

Causal Effect Estimation and Decision-Making. For estimating treatment effect [58, 68, 71, 72], a wide body of literature has focused on binary treatment scenarios, which can be broadly categorized into tree-based methods [7, 17, 21, 53], propensity score-based methods [9, 44–46, 51], covariate balancing methods [3, 11, 23, 57], doubly robust methods [4, 41, 43], representation learning methods [8, 24, 25, 36, 48, 50, 70], and generative model based methods [37, 62]. When the treatment is continuous variable such as subsidy value, estimating the treatment effect is commonly referred to as dose-response curve estimation. By modeling conditional density of treatment given features, the Generalized Propensity Score (GPS)-based methods adopt matching [38] and weighting [42] for treatment effect estimation. DRNet extends traditional representation-based methods to continuous treatment settings, which discretizes the continuous treatment values into multiple intervals and models each interval separately [47]. VCNet takes one step further by introducing a varying coefficient prediction head network [40], guaranteeing the continuity in the predicted dose-response curve. SCIGAN extends the traditional generative modeling methods, which introduces both treatment type discriminator and dosage discriminators to improve the quality of the generated counterfactuals through adversarial learning [5]. Different from causal effect estimation, causal decision making aims to determine whether a specific treatment should be applied to a unit to achieve an optimal outcome. For binary treatments, the goal is to identify individuals whose outcomes would change from negative to positive if treated [12]. Fernández-Loría and Provost [13] emphasize that accurate effect estimation is not always necessary for effective causal decision making. The key difference is that causal decision making is a classification problem, while causal effect estimation is a regression task [63, 66].

Subsidy Allocation in Marketing. In recent years, subsidy allocation methods have gained widespread attention in marketing, which share close connections with recommender systems [28, 31, 34, 55], as both aim to estimate counterfactual responses and select the most proper items [32, 54, 56, 59]. Traditional heuristic approaches [16, 65, 67] focus on predicting causal effects but lack a clear optimization framework, limiting their effectiveness. The mainstream budget allocation method is a two-stage approach [1, 15], where causal inference is used in the first stage to predict treatment effects, and integer programming is used in the second stage for optimal allocation. Based on this, Li et al. [33] propose a personalized optimization method based on counterfactual estimation and multi-task learning, while Chen et al. [6] estimate treatment effects unbiasedly using inverse propensity weighting and optimize exposure strategies. He et al. [19] address the long-tail distribution problem in income regression by normalizing and adjusting extreme data points. Sun et al. [52] improve model robustness by solving the issue of over-reliance on features through adversarial desensitization, and Huang et al. [22] address treatment misalignment by integrating contextual information. Note that existing marketing algorithms typically adopt symmetric loss functions like MSE for treatment effect estimation. However, more accurate causal effect estimates do not guarantee a better subsidy strategy. To address this issue, we develop a two-stage subsidy learning method to achieve better subsidy allocation.

3 Problem Setup

We consider a merchant-facing subsidy allocation scenario, where the platform allocates limited marketing budgets to item-level candidates. Since this paper focuses on the subsidy on the item, we ignore the user subscript. For a specific user, let $\mathcal{I} = \{i_1, \ldots, i_n\}$ denote the set of item candidates, where each candidate can correspond to an item, a product campaign, or a merchant-item opportunity eligible for subsidy support. For each item candidate $i \in \mathcal{I}$, denote $X_i \in \mathbb{R}^P$ as the observed features, including item attributes, merchant-side information, campaign context, and aggregated demand signals. Let $b_i \in \mathbb{R}^+$ be the maximum subsidy budget reserved for item candidate i, $T_i \in [0, b_i]$ be the subsidy amount assigned to item candidate i, and $Y_i \in \mathbb{R}$ be the realized profit.

We adopt the potential outcome framework, in which the subsidy T_i is defined as the treatment, and the potential outcome $Y_i(t)$ denotes the profit that would be observed if item candidate i were assigned subsidy level t. In practice, each item candidate can receive only one realized subsidy level, resulting in only one observed outcome. We assume the consistency assumption holds, meaning that the observed outcome corresponds to the potential outcome under the actual subsidy, i.e., $Y_i = Y_i(T_i)$. Additionally, we assume the Stable Unit Treatment Value Assumption (SUTVA) holds at the item-candidate level, which states that there are no alternative forms of the same subsidy treatment and that item candidates are independent after conditioning on their observed features. Denote $\mathcal{I}_{tr}$ and $\mathcal{I}_{te}$ as the training and test sets of item candidates.

In the marketing scenario, our goal is to stimulate transactions for item candidates through subsidies while respecting practical budget constraints. Let $D_i \in \mathbb{R}^+ \cup \{0\}$ denote the item-specific subsidy value of item candidate i. This value is determined by the intrinsic characteristics of the item candidate, such as item price, category, merchant margin, campaign context, and historical aggregated demand signals. Therefore, different item candidates naturally correspond to different subsidy values. When the actual subsidy satisfies $T_i \geq D_i$, the subsidized offer converts; otherwise, it does not convert. The potential outcome $Y_i(T_i)$ can be expressed as follows:

$$Y_i(T_i) = \begin{cases} R_i - T_i, & \text{if } T_i \geq D_i, \\ 0, & \text{if } T_i < D_i, \end{cases} \tag{1}$$

where $R_i \in \mathbb{R}^+$ is the known net profit without subsidy.

Ideally, the optimal treatment is the item-specific subsidy value, i.e., $T_i^* = D_i$. An ideal goal is to learn a prediction model that accurately estimates the unknown D_i from observed item-level features X_i. However, given the difficulty of precise point estimation, this study instead aims to learn a prediction function $f : \mathbb{R}^P \to \mathbb{R}^+ \cup \{0\}$ such that the predicted value $f(X_i)$ falls in the right-side neighborhood of the unknown D_i, achieving robust predictions by satisfying $f(X_i) \geq D_i$ while remaining close to D_i. Random variables are denoted without subscripts. Under a platform-level budget $B \leq \sum_{i \in \mathcal{I}} b_i$, the final allocation further selects item candidates whose predicted subsidies yield the highest profit per unit cost.

4 Methodology

4.1 Two-stage Learning Method

We decompose the robust item-level subsidy prediction task into two stages. In Stage 1, we use historical subsidy and conversion information to obtain a coarse estimate of the item-specific subsidy value. In Stage 2, we refine this estimate with an asymmetric loss that encourages slight overestimation, guiding the final prediction to fall within the right-side neighborhood of the unknown optimal subsidy value and thereby enhancing robustness.

4.1.1 Stage 1: Coarse-grained Expected Subsidy Estimation. In the first stage, we develop an auxiliary item-specific subsidy estimation model $f_D(X_i) = \hat{D}_i$, designed to estimate the item-specific subsidy value D_i, distinct from the final prediction model f. Here, D_i is determined by the intrinsic characteristics of item candidate i, such as item price, category, merchant margin, campaign context, and historical aggregated demand signals. Therefore, different item candidates naturally correspond to different subsidy values. The supervision signal is derived from two sources.

• First, by leveraging historical subsidy and conversion data in the training set, we can infer a valid range for the item-specific subsidy value, serving as a factual supervision signal.

• Second, since the item-specific subsidy value should be determined by the observed item, merchant, and campaign features, and conditionally independent of historical subsidy assignments, we introduce a regularizer to enforce the conditional independence of D and T given X as an additional supervision signal.

Based on the collected training data, if an item-level opportunity converts with $Y_i > 0$ after receiving a subsidy T_i, it indicates that the assigned subsidy is sufficient for item candidate i under its current product and campaign characteristics. In other words, the factual subsidy T_i reaches or exceeds the potential item-specific subsidy value D_i. Conversely, if the opportunity does not convert with $Y_i = 0$, it suggests that the current subsidy level is still insufficient under this item's product and campaign characteristics, and a higher subsidy may be required. According to this, we propose the following hinge loss:

$$L_{\text{hinge}}(f_D) = \frac{1}{|\mathcal{I}_{tr}|}\Big[\sum_{\substack{i\in\mathcal{I}_{tr}\\ Y_i=0}} \max(T_i - f_D(X_i), 0) + \sum_{\substack{i\in\mathcal{I}_{tr}\\ Y_i>0}} \max(f_D(X_i) - T_i, 0)\Big]. \tag{2}$$

The item-specific subsidy value D typically depends on item, merchant, and campaign features, such as item price, merchant margin, category, and campaign context, and is independent of the actual subsidy T assigned by either an algorithm or random allocation after conditioning on X. Formally, this implies that the conditional covariance is zero, i.e.,

$$\mathbb{E}\big[(D - \mathbb{E}[D \mid X]) \cdot (T - \mathbb{E}[T \mid X]) \mid X\big] = 0, \tag{3}$$

and we propose using this conditional covariance to measure the target conditional dependence.

In practice, $\mathbb{E}[D \mid X]$ is modeled by the target function f_D, while $\mathbb{E}[T \mid X]$ is implemented as an auxiliary function f_T, which takes item-level features X_i as input to predict the assigned subsidy T_i. Since D_i is unobserved, direct computation of the conditional covariance is infeasible. To address this, we generate pseudo-labels for D_i by scaling the net profit without subsidy, R_i, by a factor k, reflecting the intuition that item candidates with higher margins can support larger effective subsidy values. Based on this construction, the conditional dependence loss is derived as:

$$\begin{aligned} &L_{\text{cd}}(f_D, f_T)\\ &= \frac{1}{|\mathcal{I}_{tr}|}\sum_{i\in\mathcal{I}_{tr}} (\bar{D}_i - f_D(X_i)) \cdot (T_i - f_T(X_i))\\ &= \frac{1}{|\mathcal{I}_{tr}|}\sum_{i\in\mathcal{I}_{tr}} (k \cdot R_i - f_D(X_i)) \cdot (T_i - f_T(X_i)), \end{aligned} \tag{4}$$

where $\bar{D}_i = k \cdot R_i$ is the pseudo-label, and $0 < k < 1$ is a hyper-parameter. The auxiliary model f_T is jointly optimized using the conditional dependence loss in Equation 4.

In summary, in Stage 1, the final loss function $L_{\text{1-stage}}$ of the subsidy estimation model f_D is:

$$L_{\text{1-stage}}(f_D, f_T) = L_{\text{hinge}}(f_D) + \alpha \cdot L_{\text{cd}}(f_D, f_T), \tag{5}$$

where α is a hyper-parameter. The learned model f_D provides a coarse-grained estimate of D_i, denoted as $\hat{D}_i$.

4.1.2 Stage 2: Robust Prediction Based on Asymmetric Loss. Since perfectly predicting the unobserved item-specific subsidy value is highly challenging or even impossible, the first-stage output $\hat{D}_i = f_D(X_i)$ inevitably contains errors. If the final subsidy prediction falls slightly below the true item-specific subsidy value, the assigned subsidy may still be insufficient and the opportunity receives zero reward. In contrast, predictions slightly above this value can still trigger conversion, although at a higher subsidy cost. This motivates the use of an asymmetric loss function that deliberately biases predictions upward to ensure the final prediction $f(X_i) \geq \hat{D}_i$, providing a more robust subsidy.

Given the estimated item-specific subsidy value $\hat{D}_i$ learned from Stage 1, a natural idea is to use the naive negative reward function shown in Figure 1(c) as the Stage-2 loss, given as:

$$-\hat{R}(f) = \frac{1}{|\mathcal{I}_{tr}|}\sum_{i\in\mathcal{I}_{tr}} (f(X_i) - R_i)\mathbb{I}(f(X_i) \geq \hat{D}_i). \tag{6}$$

However, directly optimizing the negative reward function $-\hat{R}(f)$ has two main drawbacks. First, the objective function is discontinuous at $f(X_i) = \hat{D}_i$. Second, when $f(X_i) < \hat{D}_i$, the objective function value always remains zero, thereby blocking gradient flow and preventing effective parameter updates.

To address these issues, we introduce the following approximation function to replace the indicator function $\mathbb{I}(\cdot)$:

$$g(x) = \begin{cases} x + \hat{D}_i - R_i, & \text{if } x \geq 0,\\ -|\hat{D}_i - R_i| \cdot e^{\tau \cdot x}, & \text{if } x < 0, \end{cases} \tag{7}$$

where $\tau > 0$ is a large constant. Based on this, we define the refinement loss $L_{\text{2-stage}}(f)$ as:

$$L_{\text{2-stage}}(f) = \frac{1}{|\mathcal{I}_{tr}|}\sum_{i\in\mathcal{I}_{tr}} g(f(X_i) - \hat{D}_i). \tag{8}$$

Using $L_{\text{2-stage}}(f)$ for training possesses several desired properties. First, this construction ensures continuity at $f(X_i) = \hat{D}_i$. Second, it generates nonzero gradients whenever $f(X_i) < \hat{D}_i$, guiding the model to provide a subsidy that is not lower than the estimated item-specific subsidy value. Third, as $\tau \to \infty$, the $L_{\text{2-stage}}(f)$ loss converges to the original negative reward function, thus preserving the crucial asymmetric penalty property.

The resulting model f then produces the final item-level robust subsidy predictions. In the absence of budget constraints, these predictions directly constitute the subsidy policy for each item-level opportunity.

4.2 Theoretical Analysis

We now establish theoretical guarantees for our two-stage learning framework. Our goal is to prove that the regret of the learned policy relative to the ideal optimal policy is bounded under finite samples.

We define the reward function for any prediction function f as:

$$\pi(f) = \mathbb{E}[(R_i - f(X_i))\mathbb{I}(f(X_i) \geq D_i)], \tag{9}$$

where $\pi(f)$ represents the expected profit, R_i is the net profit without subsidy, $f(X_i)$ is the predicted subsidy, D_i is the true item-specific subsidy value, and $\mathbb{I}(\cdot)$ is the indicator function.

The regret is defined as:

$$\text{Regret}(f(\hat{D})) = \pi(f^*) - \pi(f(\hat{D})), \tag{10}$$

where f^* is the ideal optimal policy with access to true D_i, and $f(\hat{D})$ is our learned policy based on estimated $\hat{D}_i$ from Stage 1.

Theorem 4.1 (Regret Bound). *With probability at least $1 - \delta$, the regret satisfies:*

$$\pi(f^*) - \pi(f(\hat{D})) \leq 2(1 + \max_i |\hat{D}_i - R_i|) \cdot Ra(\mathcal{F}_D) + 4T_{\max}\sqrt{\frac{2\log(4/\delta)}{|\mathcal{I}|}},$$

where the Rademacher complexity $Ra(\mathcal{F}_D)$ measures the function class capacity, defined as:

$$Ra(\mathcal{F}_D) = \mathbb{E}_{\sigma \sim \{-1,+1\}^{|\mathcal{I}|}}\left[\sup_{f_D \in \mathcal{F}_D} \frac{1}{|\mathcal{D}|}\sum_i \sigma_i \mathcal{L}_{hinge}(f_D(X_i))\right],$$

where σ_i are independent Rademacher random variables taking values in $\{-1, +1\}$ with equal probability, and $\mathcal{L}_{hinge}$ is the hinge loss from Stage 1. $\max_i |\hat{D}_i - R_i|$ captures the worst-case estimation bias, $T_{\max}$ is the maximum subsidy value, $|\mathcal{I}|$ is the number of items, and δ is the confidence parameter.

The regret bound consists of two terms. The first term $2(1 + \max_i |\hat{D}_i - R_i|) \cdot Ra(\mathcal{F}_D)$ captures the estimation error from Stage 1: smaller Rademacher complexity and more accurate subsidy-value estimates lead to smaller regret. The second term $4T_{\max}\sqrt{\frac{2\log(4/\delta)}{|\mathcal{I}|}}$ represents statistical fluctuation that decays at rate $O(1/\sqrt{n})$ as sample size increases.

This theorem establishes that our two-stage framework achieves provably near-optimal subsidy allocation with high probability. The bound demonstrates that: (1) accurate Stage 1 estimation (via hinge loss and conditional independence) directly improves final performance; (2) the asymmetric loss in Stage 2 provides robustness by ensuring predictions fall in the right-side neighborhood of true item-specific subsidy values; and (3) the regret vanishes as sample size grows, guaranteeing consistency.

4.3 Allocation Strategy with Budget Constraint

In real-world subsidy allocation scenarios, decisions must always be made under a predefined budget constraint B. To this end, we propose a subsidy allocation strategy based on the Sharpe Ratio (SR) [49], which assumes each predicted subsidy $\hat{T}_i$ triggers a conversion and evaluates the profitability per unit of subsidy, given as:

$$SR(\hat{T}_i) = \frac{R_i - \hat{T}_i}{\hat{T}_i}, \tag{11}$$

Algorithm 1: Two-stage Robust Optimal Subsidy Learning with Budget

Input: Subsidy prediction model f; Expected subsidy estimation model f_D; Factual subsidy prediction model f_T; Training dataset $\{X_i, T_i, Y_i, R_i\}_{i \in \mathcal{I}_{tr}}$; Test dataset$\{X_i, R_i\}_{i \in \mathcal{I}_{te}}$; Budget B.

1 $Epoch \leftarrow 0$;
2 **while** *not convergent* **do**
3   Compute hinge loss $L_{\text{hinge}}(f_D)$ and the conditional dependence loss $L_{\text{cd}}(f_D, f_T)$;
4   Compute the final loss of the first stage $L_{\text{1-stage}}(f_D, f_T)$;
5   Update f_D and f_T jointly using $L_{\text{1-stage}}(f_D, f_T)$;
6   $Epoch \leftarrow Epoch + 1$;
7 **end**
8 $Epoch \leftarrow 0$;
9 **while** *not convergent* **do**
10   Compute the subsidy prediction loss $L_{\text{2-stage}}(f)$;
11   Update f using $L_{\text{2-stage}}(f)$;
12   $Epoch \leftarrow Epoch + 1$;
13 **end**
14 Obtain optimal subsidy prediction $\{f(X_i)\}_{i \in \mathcal{I}_{te}}$;
15 Compute Sharpe Ratio $\{SR(f(X_i))\}_{i \in \mathcal{I}_{te}}$;
16 Cost $C \leftarrow 0$, Subsidy prediction $\hat{T}_i \leftarrow 0$;
17 Sort the samples in descending order of their Sharpe Ratios;
18 **for** *each sample $i \in \mathcal{I}_{te}$ in sorted order* **do**
19   **if** $C + f(X_i) \leq B$ **then**
20     $\hat{T}_i \leftarrow f(X_i)$;
21     $C \leftarrow C + f(X_i)$;
22   **end**
23   **else**
24     $\hat{T}_i \leftarrow 0$;
25   **end**
26 **end**

Output: $\{\hat{T}_i\}_{i \in \mathcal{I}_{te}}$.

where $R_i - \hat{T}_i$ is the expected profit under subsidy $\hat{T}_i$ and $\hat{T}_i$ is the subsidy cost. Based on this metric, the subsidy allocation strategy comprises two steps. First, compute the Sharpe ratio $SR(\hat{T}_i)$ for each predicted subsidy $\hat{T}_i$ and sort all items in descending order. This ranking reflects the return on subsidy, where items at the top generate the greatest profit per unit cost. Second, allocate subsidies to items in this order until the cumulative subsidy expenditure reaches the budget constraint, and all remaining items receive a subsidy of zero. This procedure ensures that, under the budget constraint, total profit is maximized.

The proposed two-stage optimal subsidy prediction algorithm with budget constraints is shown in Algorithm 1. Lines 1-13 train the subsidy prediction model through two-stage learning, and Line 14 produces its output, also known as the robust estimations of expected subsidy, yielding the subsidy policy without budget constraints. Lines 15-26 then apply the Sharpe Ratio procedure to derive the budget-constrained allocation strategy.

5 Experiments

5.1 Experimental Setup

This study addresses a continuous treatment (subsidy) scenario, whereas existing public marketing datasets are configured for binary treatment and thus unsuitable. Accordingly, following prior work [33], we employ semi-synthetic data generated from public datasets, which is common in causal inference research. In the semi-synthetic experiments, the methods capable of handling continuous treatment are selected as baselines. Furthermore, we conduct an online A/B test to validate the effectiveness of the proposed method on a large-scale e-commerce platform.

5.1.1 Datasets and Semi-synthetic Data Generation. In this study, we conduct experiments on MovieLens-1M[1] [18], Yelp[2] [2] and KuaiRec[3] [14] datasets. Specifically, ML-1M contains 1,000,209 five-point scale ratings, Yelp comprises 731,671 five-point scale ratings, and KuaiRec comprises 4,676,570 video watching ratios. These datasets contain potential profit information that may reflect plausible economic relationships between item popularity and expected subsidies (e.g., higher popularity correspond to lower expected optimal subsidies).

For all datasets, we generate potential gross profit R_i, real subsidy T_i, optimal subsidy threshold D_i, and real average profit Y_i for each item as follows.

• **Generating potential gross profit** R_i: We pre-trained a Matrix Factorization (MF) [26] model to generate predicted ratings, and we treat the averaged prediction for item i as its potential gross profit R_i.

• **Generating factual subsidy** T_i: The real subsidy T_i follows a uniform distribution in the range $[0, R_i]$ with probability γ. With probability $1-\gamma$, it takes the value 0. In this experiment, we set $\gamma = 0.95$.

• **Generating optimal subsidy** D_i: We pre-trained a neural collaborative filtering (NCF) [20] model to generate predicted ratings. Let $\hat{y}_i$ be the average predicted rating for item i. For the top 5% of items based on $\hat{y}_i$, D_i is set to 0. For the remaining items, D_i follows a Gamma distribution with shape parameter θ and scale parameter $K/\hat{y}_i$. For the MovieLens-1M, Yelp, and KuaiRec datasets, we set K to 4.5, 3.5, and 1, respectively, and for all datasets, we set $\theta = 2$.

• **Generating real profit** Y_i: Based on T_i and D_i, the conversion is modeled as a non-deterministic (probabilistic) event to simulate a sparse, real-world scenario with low conversion rates. First, we define the conversion rate (CVR) using a scaled sigmoid function:

$$\mathrm{cvr}_i = \frac{1}{1000 \cdot (1 + \exp(-10 \cdot (T_i - D_i)))}.$$

This function models a permille-level probability of conversion that is dependent on the relationship between the allocated subsidy T_i and the item's expected subsidy D_i. The final real profit Y_i is then generated from a Bernoulli distribution:

$$Y_i = \begin{cases} R_i - T_i, & \text{with probability } \mathrm{cvr}_i, \\ 0, & \text{with probability } 1 - \mathrm{cvr}_i. \end{cases}$$

[1]https://grouplens.org/datasets/movielens/1M/
[2]https://www.yelp.com/dataset
[3]https://github.com/chongminggao/KuaiRec

During the training phase, only R_i, T_i, and Y_i are available, while D_i is unavailable.

5.1.2 Baselines. We compare the proposed method with both heuristic subsidy allocation strategies and continuous treatment effect estimation methods, as outlined below:

• **Heuristic-None**: This strategy does not allocate any subsidies, i.e.,$T_i = 0, \forall i \in \mathcal{I}_{te}$. The resulting profit $Reward_0$ is derived solely from the group where $D_i = 0$.

• **Heuristic-Random**: For any item i, this strategy randomly samples the subsidy T_i from a uniform distribution over $[0, R_i]$.

• **S-Learner** [27]: This is a type of meta-learners that treat the treatment T_i as a one-dimensional feature, and it is concatenated with the existing features and used to estimate the outcome.

• **GPS (Generalized Propensity Score)** [42]: This method employs the generalized propensity score for continuous treatments to re-weight the observed samples.

• **DRNet** [47]: This method first learns a shared representation, stratifies the continuous treatment T_i into discrete intervals, and trains separate prediction head models for each interval to estimate the outcome.

• **CRNet** [75]: This method learns both balancing and prognostic representations through a novel contrastive learning framework to effectively achieve unbiased estimation of heterogeneous dose-response curves.

• **VCNet** [40]: Similar to DRNet, this method initially learns a shared representation, and then trains a prediction head model with varying coefficients, which incorporates the continuous treatment T_i as an input, to estimate the outcome.

Treatment effect estimation methods, such as VCNet, are capable of predicting the outcome corresponding to specific features X_i and a given treatment T_i. To address the subsidy prediction task, we extend their application by providing multiple candidate treatments and selecting the treatment that yields the highest predicted outcome as the optimal treatment. Specifically, for each item, the candidate interventions are defined as the K-equidistant points within the interval $[0, R_i]$, where K is set to 10 across all datasets. We will release all code for the semi-synthetic experiments upon the paper's acceptance.

5.1.3 Experimental Environment. The semi-synthetic experiments are carried out on an Ubuntu 22.04 system equipped with 80GB of memory, a 32-core CPU, and 24GB NVIDIA 3090 GPUs, utilizing PyTorch 2.0.0 and CUDA 11.7. And the large-scale online A/B test experiments are implemented in TensorFlow and executed on NVIDIA A100 GPUs.

5.1.4 Evaluation Metrics. We adopt the following three evaluation metrics to assess the performance of the predicted subsidy strategy $\{T_i^{\text{pred}}\}_{i \in \mathcal{I}_{te}}$:

• **Order Rate (OR)** evaluates the proportion of successful conversions under the predicted subsidy strategy. In this study, the OR values are scaled by 10^{-3} (permille) to reflect practical industry standards, defined as:

$$OR = \frac{\sum_{i \in \mathcal{I}_{te}} \mathbb{I}(T_i^{\text{pred}} \geq D_i) \cdot \mathbb{I}(T_i^{\text{pred}} \leq R_i)}{|\mathcal{I}_{te}|},$$

Table 1: Performance comparison with and without budget, where the best results and baseline results are bolded and underlined. Asterisk (*) indicates a significant improvement over the best baseline VCNet using a paired t-test across 10 runs at $p < 0.05$. To align with real-world scenarios, the OR metric is reported in permille (10^{-3}).

With Budget	ML-1M			YELP			KUAIREC		
Methods	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR
None	2.891	3212.553	-	2.956	2425.546	-	1.932	1099.504	-
Random	4.641	3754.941	0.511	5.801	2837.835	0.316	2.619	1108.885	0.324
S-Learner	4.371	4426.183	0.468	4.086	3357.425	0.298	1.934	1181.923	0.239
GPS	4.542	4610.557	0.469	6.150	3775.710	0.391	2.313	1149.406	0.481
DRNet	4.179	4905.412	0.354	7.319	4022.938	0.866	2.681	1180.909	0.855
CRNet	4.638	5120.958	1.244	6.962	4116.185	0.924	2.352	1283.901	1.046
VCNet	4.876	5450.379	1.348	6.972	3991.982	1.134	2.745	1423.771	1.281
	± 0.025	± 18.609	± 0.002	± 0.015	± 14.317	± 0.007	± 0.016	± 8.709	± 0.010
CoLA	**5.334***	**5573.352***	**1.588***	**8.728***	**4226.012***	**1.454***	**2.969***	**1471.073***	**1.533***
	± 0.052	± 20.197	± 0.014	± 0.016	± 15.086	± 0.005	± 0.024	± 6.636	± 0.020
Without Budget	ML-1M			YELP			KUAIREC		
Methods	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR
None	2.891	3212.553	-	2.956	2425.546	-	1.932	1099.504	-
Random	4.819	4113.743	0.403	5.929	2603.875	0.308	2.660	1047.884	0.315
S-Learner	4.414	3840.658	0.468	4.484	3128.653	0.271	2.081	1159.666	0.230
GPS	4.874	4303.313	0.631	6.308	3502.993	0.386	2.323	1130.326	0.262
DRNet	5.627	4426.183	0.585	6.620	3827.632	0.832	2.847	1142.892	0.147
CRNet	8.053	4307.071	0.750	7.744	3982.434	0.868	2.482	1125.878	0.412
VCNet	7.620	4957.663	1.020	7.343	3636.134	1.043	**3.152**	**1362.968**	0.880
	± 0.052	± 32.858	± 0.007	± 0.011	± 31.744	± 0.008	± 0.005	± 3.406	± 0.001
CoLA	**8.931***	**5015.779***	**1.480***	**8.956***	**4104.366***	**1.409***	3.055	1327.621	**1.106***
	± 0.047	± 34.947	± 0.009	± 0.061	± 24.104	± 0.011	± 0.005	± 6.847	± 0.002

where $\mathbb{I}(\cdot)$ is the indicator function, and D_i and R_i denote the optimal subsidy and the potential net profit, respectively.

• **Reward** measures the total profit achieved under the predicted subsidy strategy, which is calculated as:

$$Reward = \sum_{i \in \mathcal{I}_{te}} Y_i^{\text{pred}},$$

where Y_i^{pred} is profit under subsidy T_i^{pred}, determined by:

$$Y_i^{\text{pred}} = \begin{cases} R_i - T_i^{\text{pred}}, & \text{if } T_i^{\text{pred}} \geq D_i, \\ 0, & \text{if } T_i^{\text{pred}} < D_i. \end{cases}$$

• **Incremental Leverage Ratio (ILR)** quantifies the incremental profit per item subsidy relative to the no-subsidy strategy None method ($Reward_0$) and is expressed as:

$$ILR = \frac{Reward - Reward_0}{\sum_{i \in \mathcal{I}_{te}} T_i^{\text{pred}}}.$$

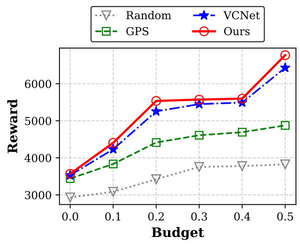
5.1.5 Implementation Details. We evaluate the performance of different methods under two scenarios: **with** and **without** budget constraints. In the budget-constrained (with budget) scenario, all models prioritize subsidy allocation to items with the highest average Sharpe Ratio, continuing this process until the cumulative subsidy expenditure reaches the predefined budget B. We define the budget B as follows:

$$B = \beta \cdot \sum_{i \in \mathcal{I}_{te}} R_i,$$

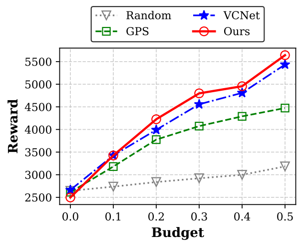
where we initialize $\beta = 0.2$. The term $\sum_{i \in \mathcal{I}_{te}} R_i$ represents the total profit under the idealized scenario, where all test item samples are converted without any subsidy, i.e., $D_i = 0, \forall i \in \mathcal{I}_{te}$. All methods are implemented using PyTorch, with the Adam optimizer employed for training. For all datasets, we tune embedding size in $\{16, 64\}$, batch size in $\{2048, 4096, 8192\}$, learning rate in $\{0.005, 0.01, 0.05\}$, and weight decay in $\{1e-6, 1e-5, 1e-4, 1e-3\}$.

5.2 Performance Comparison

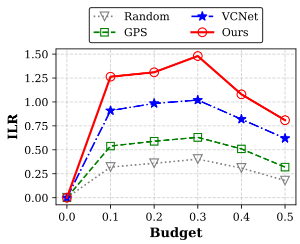
Table 1 reports the performance of all methods on three datasets under both unconstrained and budget-constrained settings. None method, which provides no subsidies, serves as the baseline for incremental leverage ratio (ILR) calculation. Random method allocates subsidies uniformly at random, yielding modest improvements in order rate and reward but low ILR due to inefficiency. S-Learner, which treats the intervention as a simple feature, consistently underestimates subsidies and performs the worst, with costs remaining below budget. GPS, DRNet, and CRNet show similar results, while



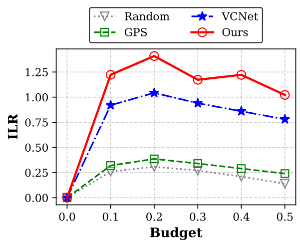
(a) Reward on ML-1M.



(b) Reward on Yelp.



(c) ILR on ML-1M.



(d) ILR on Yelp.

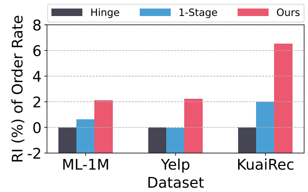
Figure 2: Effects of varying budget (tuned by hyper-parameter β) on ML-1M and Yelp datasets.

Table 2: Robust subsidy prediction accuracy within the true expected subsidy threshold right-side neighborhood.

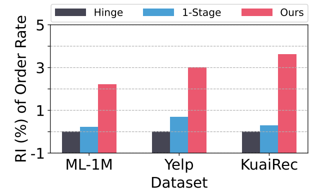
	ML-1M, $K = 1$			ML-1M, $K = 2$		
γ	0.9	0.7	0.5	0.9	0.7	0.5
VCNet	0.275	0.278	0.281	0.433	0.427	0.415
OURS	**0.490**	**0.466**	**0.453**	**0.487**	**0.489**	**0.477**
	KuaiRec, $K = 0.5$			KuaiRec, $K = 1$		
γ	0.9	0.7	0.5	0.9	0.7	0.5
VCNet	0.340	0.231	0.229	0.408	0.452	0.439
OURS	**0.498**	**0.493**	**0.466**	**0.499**	**0.488**	**0.486**

GPS produces lower rewards and ILRs under comparable order rates, suggesting potential over-subsidization. VCNet outperforms DRNet on all metrics due to its flexible varying-coefficient prediction head, making it the strongest baseline. Our method achieves the best performance across datasets, with pronounced gains under budget constraints, confirming the effectiveness of the proposed two-stage learning framework with asymmetric loss.

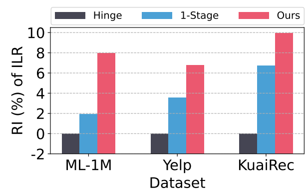
Figure 2 examines the effect of budget size, controlled by hyper-parameter β (larger β indicates higher budget). As budgets increase, all methods improve in reward, but ILR peaks at $\beta = 0.2$ and 0.3 and declines thereafter. Random performs worst due to indiscriminate allocation. GPS behaves similarly to Random at small budgets but improves with larger budgets while maintaining ILR better. VCNet is the most competitive baseline overall. Our method consistently outperforms all baselines, particularly under tight budgets, where it achieves a significantly higher ILR by accurately predicting subsidies just above expected values, enabling cost-efficient order completions and maximizing profits.



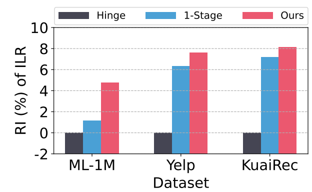
(a) OR with Budget.



(b) OR without Budget.



(c) ILR with Budget.



(d) ILR without Budget.

Figure 3: Effects of conditional dependence loss and asymmetric prediction loss across three datasets.

5.3 In-Depth Analysis

Table 2 evaluates whether the proposed robust subsidy prediction falls within the right neighborhood of the true expected subsidy threshold D_i, defined as the interval $[D_i, D_i + \sigma]$. We measure this as the probability that test samples' predictions fall within their respective true intervals, where σ corresponds to the sampling standard deviation of D_i in the semi-synthetic experiments. In these experiments, the ground-truth threshold D_i is influenced by hyper-parameter K, while the actual subsidy T_i in training data is affected by hyperparameter γ. Results across various hyperparameter settings show that our method consistently achieves a higher probability of predicting subsidies within the target interval, demonstrating its robustness.

5.4 Ablation Study

Figure 3 evaluates the effect of removing Stage 2 asymmetric loss and Stage 1 conditional independence loss. The full method is denoted as "Ours", the variant without Stage 2 loss as "1-Stage", and the variant removing both losses (trained solely with hinge loss) as "Hinge". Using "Hinge" as the baseline, we report relative improvements (RI).

Impact of Conditional Dependence Loss. A comparison between "1-Stage" and "Hinge" variants reveals that introducing the conditional independence loss L_{cd} does not significantly improve the Order Rate but leads to a notable increase in ILR and Reward. This indicates that while the hinge loss provides a reliable interval supervision, there are inevitable deviations from the true values. By incorporating the conditional independence constraint, the predictions more closely approximate the true expected subsidies.

Impact of Asymmetric Prediction Loss. A comparison between the "Ours" and "1-Stage" variants shows that incorporating

Table 3: Robustness analysis under violated conditional independence assumption. Best results are bolded. To align with real-world scenarios, the OR metric is reported in permille (10^{-3}).

With Budget	ML-1M			Yelp			KuaiRec		
Methods	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR
VCNet	4.538	**5195.723**	1.266	6.767	3649.383	1.028	2.724	**1351.307**	1.239
Ours	**5.308**	5048.316	**1.456**	**8.243**	**3946.012**	**1.434**	**2.734**	1324.388	**1.444**
Without Budget	ML-1M			Yelp			KuaiRec		
Methods	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR	OR ($\times 10^{-3}$)	Reward	ILR
VCNet	7.012	4741.640	0.919	6.711	3451.522	1.013	2.964	1235.567	0.854
Ours	**8.931**	**5015.77**	**1.480**	**8.956**	**4104.366**	**1.409**	**3.055**	**1327.621**	**1.106**

Table 4: Effects of hyper-parameter α with budget.

	ML-1M		Yelp		KuaiRec	
α	Reward	ILR	Reward	ILR	Reward	ILR
0	3729.996	1.128	3699.504	1.020	1284.238	1.338
1e-5	5191.716	1.391	4188.441	1.235	1360.912	1.422
1e-4	5243.452	1.505	4203.070	1.444	1446.626	1.416
1e-3	5486.290	1.571	**4226.012**	**1.454**	1447.836	1.432
1e-2	**5573.352**	**1.588**	3813.880	1.311	**1471.073**	**1.533**
1e-1	5339.879	1.436	3778.060	1.402	1187.237	1.334

Table 5: Effects of hyper-parameter α without budget.

	ML-1M		Yelp		KuaiRec	
α	Reward	ILR	Reward	ILR	Reward	ILR
0	3985.967	1.017	3088.032	1.123	1094.522	1.003
1e-5	4614.372	1.231	3833.403	1.327	1257.582	1.043
1e-4	4734.716	1.268	4003.370	1.348	1268.481	1.082
1e-3	4992.510	1.459	**4104.366**	**1.409**	1294.163	1.080
1e-2	**5015.779**	**1.480**	3781.571	1.379	**1327.621**	**1.106**
1e-1	4934.099	1.365	3375.989	1.240	1071.319	0.994

the asymmetric loss $L_{\text{2-stage}}$ in the second stage substantially improves the Order Rate. This improvement stems from the asymmetric loss guiding the model to produce more robust subsidy predictions, specifically higher subsidies, leading to increased conversions. While the higher Order Rate contributes to greater overall rewards, it also results in relatively higher costs. Consequently, without budget constraints, the ILR improvement is less pronounced, whereas under budget constraints, the Sharpe ratio-based high-efficiency allocation strategy mitigates this effect.

5.5 Robustness Analysis

Our estimation in Stage 1 relies on the conditional independence assumption (i.e., $D \perp T \mid X$), where D denotes the smallest subsidy level under which the item-level offer is expected to convert. However, in real-world recommendation platforms, this assumption may be violated. A common scenario is that the allocated subsidy T and the item's conversion threshold D mutually influence each other across consecutive allocation rounds. For example, the subsidy T_1 assigned to an item in a previous round shifts the item's conversion threshold D_2 in the next round (e.g., once a higher subsidy has been offered, the optimal subsidy required to trigger conversion on the same item-level offer tends to drift upward), and *at the same time*, the item's initial threshold D_1 influences the platform's next subsidy decision T_2.

To evaluate the robustness of our method under such conditions, we construct a setting that explicitly violates the assumption by simulating this two-way dependency. Let D_1 and T_1 denote the item's initial conversion threshold and the platform's initial allocated subsidy, respectively. We then generate the updated threshold D_2 and the updated subsidy T_2 via:

$$D_2 = D_1 \exp(\max\{T_1 - D_1, 0\}/10),$$
$$T_2 = T_1 \exp((D_1 - T_1)/10).$$

These functions produce two regimes based on the initial round:

- **If** $T_1 > D_1$**:** the allocated subsidy exceeds what is minimally required for conversion. The item's conversion threshold in the next round (D_2) **will rise**, since the max term is positive, reflecting an upward drift of the optimal subsidy needed to convert the same offer. Simultaneously, the platform's subsidy in the next round (T_2) **will decrease**, as the exponent ($D_1 - T_1$) is negative, reflecting a corrective pullback from over-subsidization.
- **If** $D_1 \geq T_1$**:** the allocated subsidy fails to meet the conversion threshold. The item's threshold (D_2) **remains unchanged**, since the max term is zero. However, the platform's next subsidy allocation (T_2) **will increase**, as the exponent ($D_1 - T_1$) is non-negative, reflecting an upward adjustment toward the threshold.

As shown in Table 3, even when the conditional independence assumption is violated under this two-way feedback between T and the item's conversion threshold D, our method still comprehensively outperforms the VCNet baseline on the key metrics of profit (Reward) and subsidy efficiency (ILR). This demonstrates that our proposed two-stage framework possesses strong robustness.

5.6 Sensitivity Analysis

Tables 4 and 5 investigate the effect of the conditional independence loss (L_{cd}) coefficient α in Stage 1 on prediction performance,

Table 6: Results of the online A/B test.

	Day1	Day2	Day3	Day4	Day5
ILR	**+8.06%**	**+16.70%**	**+37.44%**	-16.00%	**+30.87%**
GTV	**+3.22%**	**+2.14%**	**+1.40%**	-3.97%	**+1.38%**
OR	-1.67%	**+1.16%**	**+1.27%**	**+1.14%**	**+1.62%**
	Day6	Day7	Day8	Day9	Day10
ILR	**+41.32%**	**+7.24%**	-0.69%	**+1.96%**	-1.57%
GTV	**+0.87%**	**+2.60%**	**+1.70%**	**+1.04%**	-2.22%
OR	**+1.84%**	**+0.09%**	-1.04%	-0.78%	**+4.66%**
	Day11	Day12	Day13	Day14	Day15
ILR	**+43.62%**	**+18.49%**	**+13.76%**	**+28.81%**	**+12.35%**
GTV	**+3.95%**	-0.53%	-1.40%	**+1.77%**	**+1.18%**
OR	**+1.00%**	**+3.79%**	**+5.71%**	**+1.09%**	**+1.43%**

both with and without budget constraints. We assess Reward and ILR across three public datasets under varying α, with the best results highlighted in bold. The findings show that setting α within a moderate range effectively balances the hinge loss L_{hinge} and conditional dependence loss L_{cd}, leading to notable improvements in both metrics. Optimal performance is observed when α is set to $1e-3$ or $1e-2$. Conversely, excessively large values (e.g., $1e-1$) diminish the influence of L_{hinge}, thus negatively impacting subsidy prediction accuracy.

5.7 Online A/B Test

Table 6 presents the results of a 15-day online A/B test conducted on a large-scale e-commerce platform. The experiment was conducted in a fully compliant setting, without any user-level differentiation, and subsidies were computed solely based on item-side information. We compare our proposed method against the DRNet baseline across three key metrics: ILR, GTV, and OR. All improvements are reported as percentage gains relative to DRNet.

Our method demonstrates consistent and substantial improvements across the majority of test days. For ILR, we achieve positive gains on 11 out of 15 days and the average ILR improvement across all 15 days is +16.16%, indicating significantly higher subsidy efficiency. For GTV, our method shows positive gains on 11 out of 15 days, with an average improvement of +0.88%, demonstrating increased revenue generation. The OR metric exhibits positive improvements on 12 out of 15 days, with an average gain of +1.42%, confirming higher conversion rates.

Meanwhile, even on days with negative performance, the magnitude of decline is relatively modest compared to the substantial gains observed on other days. This pattern suggests that our method maintains robustness under varying real-world conditions while delivering consistent overall improvements. The consistency of gains across multiple metrics and the majority of test days validates the practical effectiveness of our two-stage robust optimal subsidy learning framework in production environments.

6 Conclusion

This paper investigates optimal subsidy policy learning in marketing. We highlight that predicting the expected subsidy threshold is distinct from treatment effect estimation. To bridge this gap, we propose CoLA, a two-stage robust optimal subsidy learning method. Specifically, the first stage generates a coarse estimate using factual subsidy data and conditional independence assumptions; the second stage refines it via an asymmetric loss function to introduce a robust upward bias. Experiments on three public datasets and an online A/B test show our method consistently maximizes total profit and incremental leverage.

Despite its effectiveness, a limitation of this work is that it primarily focuses on profit maximization within a single transaction, thereby treating subsidy allocation as a static decision process. However, marketing subsidies often exert long-term cumulative effects at the item level. Frequent or excessive incentives may lead to "subsidy dependency" on certain items, where the conversion threshold D drifts upward over time and the same offer increasingly relies on higher subsidies to convert, eroding its long-term profit and competitiveness. Therefore, our future work aims to extend the current single-round setting into a long-term setting, where item cumulative value will be explicitly modeled, enabling subsidy allocation strategies that remain profitable and robust over repeated rounds.

Acknowledgment

Z. Lin was supported by the NSF China (No. 62276004) and the State Key Laboratory of General Artificial Intelligence.

References

[1] Meng Ai, Biao Li, Heyang Gong, Qingwei Yu, Shengjie Xue, Yuan Zhang, Yunzhou Zhang, and Peng Jiang. 2022. LBCF: A Large-Scale Budget-Constrained Causal Forest Algorithm. In *International World Wide Web Conference*.

[2] Nabiha Asghar. 2016. Yelp Dataset Challenge: Review Rating Prediction. *arXiv preprint arXiv:1605.05362* (2016).

[3] Susan Athey, Guido W Imbens, and Stefan Wager. 2018. Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 80, 4 (2018), 597–623.

[4] Heejung Bang and James M Robins. 2005. Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics* 61, 4 (2005), 962–973.

[5] Ioana Bica, James Jordon, and Mihaela van der Schaar. 2020. Estimating the Effects of Continuous-Valued Interventions Using Generative Adversarial Networks. In *Advances in Neural Information Processing Systems*.

[6] Jiaju Chen, Wang Wenjie, Chongming Gao, Peng Wu, Jianxiong Wei, and Qingsong Hua. 2024. Treatment Effect Estimation for User Interest Exploration on Recommender Systems. In *International SIGIR Conference on Research and Development in Information Retrieval*.

[7] Hugh A Chipman, Edward I George, and Robert E McCulloch. 2010. BART: Bayesian Additive Regression Trees. *The Annals of Applied Statistics* (2010), 266–298.

[8] Alicia Curth and Mihaela Van der Schaar. 2021. On Inductive Biases for Heterogeneous Treatment Effect Estimation. In *Advances in Neural Information Processing Systems*.

[9] Rajeev H Dehejia and Sadek Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics* 84, 1 (2002), 151–161.

[10] Shuyang Du, James Lee, and Farzin Ghaffarizadeh. 2019. Improve User Retention with Causal Learning. In *ACM SIGKDD Workshop on Causal Discovery*.

[11] Jianqing Fan, Kosuke Imai, Han Liu, Yang Ning, Xiaolin Yang, et al. 2016. Improving Covariate Balancing Propensity Score: A Doubly Robust and Efficient Approach. *Technical Report* (2016).

[12] Carlos Fernández-Loría and Foster Provost. 2022. Causal Classification: Treatment Effect Estimation Vs. Outcome Prediction. *Journal of Machine Learning Research* 23, 59 (2022), 1–35.

[13] Carlos Fernández-Loría and Foster Provost. 2022. Causal Decision Making and Causal Effect Estimation Are Not the Same. . . And Why It Matters. *INFORMS Journal on Data Science* 1, 1 (2022), 4–16.

[14] Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. 2022. KuaiRec: A Fully-observed

Dataset and Insights for Evaluating Recommender Systems. In *Conference on Information and Knowledge Management*.

[15] Dmitri Goldenberg, Javier Albert, Lucas Bernardi, and Pablo Estevez. 2020. Free Lunch! Retrospective Uplift Modeling for Dynamic Promotions Recommendation Within Roi Constraints. In *ACM Conference on Recommender Systems*.

[16] Leo Guelman, Montserrat Guillén, and Ana M Pérez-Marín. 2015. Uplift Random Forests. *Cybernetics and Systems* 46, 3-4 (2015), 230–248.

[17] P Richard Hahn, Jared S Murray, and Carlos M Carvalho. 2020. Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, And Heterogeneous Effects (With Discussion). *Bayesian Analysis* 15, 3 (2020), 965–1056.

[18] F Maxwell Harper and Joseph A Konstan. 2015. The Movielens Datasets: History and Context. *Acm Transactions on Interactive Intelligent Systems* 5, 4 (2015), 1–19.

[19] Bowei He, Yunpeng Weng, Xing Tang, Ziqiang Cui, Zexu Sun, Liang Chen, Xiuqiang He, and Chen Ma. 2024. Rankability-Enhanced Revenue Uplift Modeling Framework for Online Marketing. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[20] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *International World Wide Web Conference*.

[21] Jennifer L Hill. 2011. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics* 20, 1 (2011), 217–240.

[22] Yinqiu Huang, Shuli Wang, Min Gao, Xue Wei, Changhao Li, Chuan Luo, Yinhua Zhu, Xiong Xiao, and Yi Luo. 2024. Entire Chain Uplift Modeling with Context-Enhanced Learning for Intelligent Marketing. In *International World Wide Web Conference*.

[23] Kosuke Imai and Marc Ratkovic. 2014. Covariate Balancing Propensity Score. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76, 1 (2014), 243–263.

[24] Fredrik Johansson, Uri Shalit, and David Sontag. 2016. Learning Representations for Counterfactual Inference. In *International Conference on Machine Learning*.

[25] Wenwei Ke, Chuanren Liu, Xiangfu Shi, Yiqiao Dai, S Yu Philip, and Xiaoqiang Zhu. 2021. Addressing Exposure Bias in Uplift Modeling for Large-Scale Online Advertising. In *IEEE International Conference on Data Mining*.

[26] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8 (2009), 30–37.

[27] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. 2019. Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning. *Proceedings of the National Academy of Sciences* 116, 10 (2019), 4156–4165.

[28] Haoxuan Li, Quanyu Dai, Yuru Li, Yan Lyu, Zhenhua Dong, Xiao-Hua Zhou, and Peng Wu. 2023. Multiple Robust Learning for Recommendation. In *AAAI Conference on Artificial Intelligence*.

[29] Haoxuan Li, Yan Lyu, Chunyuan Zheng, and Peng Wu. 2023. TDR-CL: Targeted Doubly Robust Collaborative Learning for Debiased Recommendations. In *International Conference on Learning Representations*.

[30] Haoxuan Li, Kunhan Wu, Chunyuan Zheng, Yanghao Xiao, Hao Wang, Zhi Geng, Fuli Feng, Xiangnan He, and Peng Wu. 2023. Removing Hidden Confounding in Recommendation: A Unified Multi-Task Learning Approach. In *Advances in Neural Information Processing Systems*.

[31] Haoxuan Li, Yanghao Xiao, Chunyuan Zheng, and Peng Wu. 2023. Balancing Unobserved Confounding with a Few Unbiased Ratings in Debiased Recommendations. In *International World Wide Web Conference*.

[32] Haoxuan Li, Chunyuan Zheng, and Peng Wu. 2023. StableDR: Stabilized Doubly Robust Learning for Recommendation on Data Missing Not at Random. In *International Conference on Learning Representations*.

[33] Haoxuan Li, Chunyuan Zheng, Peng Wu, Kun Kuang, Yue Liu, and Peng Cui. 2023. Who should be given incentives? counterfactual optimal treatment regimes learning for recommendation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[34] Haoxuan Li, Chunyuan Zheng, Yanghao Xiao, Peng Wu, Zhi Geng, Xu Chen, and Peng Cui. 2024. Debiased Collaborative Filtering with Kernel-based Causal Balancing. In *International Conference on Learning Representations*.

[35] Xiang Li, Zhao-Yu Zhang, Chunyuan Zheng, Qingying Chen, Huiyou Jiang, Haoxuan Li, and Zhouchen Lin. 2026. User Activity Modeling under Inflated Distribution. In *International SIGIR Conference on Research and Development in Information Retrieval*.

[36] Dugang Liu, Xing Tang, Han Gao, Fuyuan Lyu, and Xiuqiang He. 2023. Explicit Feature Interaction-Aware Uplift Network for Online Marketing. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[37] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. 2017. Causal Effect Inference with Deep Latent-Variable Models. In *Advances in Neural Information Processing Systems*.

[38] Bo Lu, Elaine Zanutto, Robert Hornik, and Paul R Rosenbaum. 2001. Matching with Doses in an Observational Study of a Media Campaign Against Drug Abuse. *J. Amer. Statist. Assoc.* 96, 456 (2001), 1245–1253.

[39] Yan Lyu, Haoxuan Li, Tianyu Xia, Xiang Li, Xiangnan Feng, Chunyuan Zheng, and Xiao-Hua Zhou. 2026. Hierarchical Denoising Entire Space Multi-Task Model for Post-Click Conversion Rate Prediction with Noisy Labels. In *International SIGIR Conference on Research and Development in Information Retrieval*.

[40] Lizhen Nie, Mao Ye, Dan Nicolae, et al. 2021. VCNet and Functional Targeted Regularization For Learning Causal Effects of Continuous Treatments. In *International Conference on Learning Representations*.

[41] James Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. 2007. Comment: Performance of Double-Robust Estimators When" Inverse Probability" Weights Are Highly Variable. *Statist. Sci.* 22, 4 (2007), 544–559.

[42] James M Robins, Miguel Angel Hernan, and Babette Brumback. 2000. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 11, 5 (2000), 550–560.

[43] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. 1994. Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *J. Amer. Statist. Assoc.* 89, 427 (1994), 846–866.

[44] Paul R Rosenbaum. 1987. Model-Based Direct Adjustment. *J. Amer. Statist. Assoc.* 82, 398 (1987), 387–394.

[45] Paul R Rosenbaum and Donald B Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70, 1 (1983), 41–55.

[46] Paul R Rosenbaum and Donald B Rubin. 1984. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *J. Amer. Statist. Assoc.* 79, 387 (1984), 516–524.

[47] Patrick Schwab, Lorenz Linhardt, Stefan Bauer, Joachim M Buhmann, and Walter Karlen. 2020. Learning Counterfactual Representations for Estimating Individual Dose-Response Curves. In *AAAI Conference on Artificial Intelligence*.

[48] Uri Shalit, Fredrik D Johansson, and David Sontag. 2017. Estimating Individual Treatment Effect: Generalization Bounds and Algorithms. In *International Conference on Machine Learning*.

[49] William F Sharpe. 1966. Mutual Fund Performance. *The Journal of business* 39, 1 (1966), 119–138.

[50] Claudia Shi, David Blei, and Victor Veitch. 2019. Adapting Neural Networks for the Estimation of Treatment Effects. In *Advances in Neural Information Processing Systems*.

[51] Elizabeth A Stuart. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25, 1 (2010), 1.

[52] Zexu Sun, Bowei He, Ming Ma, Jiakai Tang, Yuchen Wang, Chen Ma, and Dugang Liu. 2023. Robustness-Enhanced Uplift Modeling with Adversarial Feature Desensitization. In *IEEE International Conference on Data Mining*.

[53] Stefan Wager and Susan Athey. 2018. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *J. Amer. Statist. Assoc.* (2018).

[54] Hao Wang, Zhichao Chen, Zhaoran Liu, Haozhe Li, Degui Yang, Xinggao Liu, and Haoxuan Li. 2024. Entire Space Counterfactual Learning for Reliable Content Recommendations. *IEEE Transactions on Information Forensics and Security* (2024).

[55] Hao Wang, Zhichao Chen, Honglei Zhang, Zhengnan Li, Licheng Pan, Haoxuan Li, and Mingming Gong. 2025. Debiased Recommendation via Wasserstein Causal Balancing. *ACM Transactions on Information Systems* (2025).

[56] Jun Wang, Haoxuan Li, Chi Zhang, Dongxu Liang, Enyun Yu, Wenwu Ou, and Wenjia Wang. 2023. CounterCLR: Counterfactual contrastive learning with non-random missing data in recommendation. In *IEEE International Conference on Data Mining*.

[57] Raymond KW Wong and Kwun Chuen Gary Chan. 2018. Kernel-Based Covariate Functional Balancing for Observational Studies. *Biometrika* 105, 1 (2018), 199–213.

[58] Anpeng Wu, Haoxuan Li, Chunyuan Zheng, Kun Kuang, and Kun Zhang. 2025. Classifying Treatment Responders: Bounds and Algorithms. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[59] Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, and Xiao-Hua Zhou. 2022. On the Opportunity of Causal Learning in Recommendation Systems: Foundation, Estimation, Prediction and Challenges. In *International Joint Conferences on Artificial Intelligence*.

[60] Yanghao Xiao, Hao Wang, Xiang Li, Qian Zou, Bing Cheng, Wei Lin, Haoxuan Li, and Zhouchen Lin. 2026. Debiased Recommendation Beyond the Positive Propensity Assumption. In *International SIGIR Conference on Research and Development in Information Retrieval*.

[61] Peng Ye, Julian Qian, Jieying Chen, Chen-hung Wu, Yitong Zhou, Spencer De Mars, Frank Yang, and Li Zhang. 2018. Customized Regression Model for Airbnb Dynamic Pricing. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

[62] Jinsung Yoon, James Jordon, and Mihaela Van Der Schaar. 2018. GANITE: Estimation of Individualized Treatment Effects Using Generative Adversarial Nets. In *International Conference on Learning Representations*.

[63] Bianca Zadrozny. 2003. *Policy Mining: Learning Decision Policies from Fixed Sets of Data.* University of California, San Diego.

[64] Yang Zhang, Bo Tang, Qingyu Yang, Dou An, Hongyin Tang, Chenyang Xi, Xueying Li, and Feiyu Xiong. 2021. BCORLE (λ): An Offline Reinforcement Learning and Evaluation Framework for Coupons Allocation in E-commerce Market. In *Advances in Neural Information Processing Systems*.

[65] Yan Zhao, Xiao Fang, and David Simchi-Levi. 2017. Uplift Modeling with Multiple Treatments and General Response Types. In *IEEE International Conference on*

Data Mining.
[66] Yingqi Zhao, Donglin Zeng, A John Rush, and Michael R Kosorok. 2012. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *J. Amer. Statist. Assoc.* 107, 499 (2012), 1106–1118.
[67] Zhenyu Zhao and Totte Harinen. 2019. Uplift Modeling for Multiple Treatments with Cost Optimization. In *IEEE International Conference on Data Science and Advanced Analytics*.
[68] Chunyuan Zheng, Anpeng Wu, Chuan Zhou, Taojun Hu, Qingying Chen, Hongyi Liu, Chenxi Li, Huiyou Jiang, Haoxuan Li, and Zhouchen Lin. 2026. Uplift Modeling with Delayed Feedback: Identifiability and Algorithms. In *AAAI Conference on Artificial Intelligence*.
[69] Chunyuan Zheng, Haocheng Yang, Haoxuan Li, and Mengyue Yang. 2025. Unveiling Extraneous Sampling Bias with Data Missing-Not-At-Random. In *Advances in Neural Information Processing Systems*.
[70] Kailiang Zhong, Fengtong Xiao, Yan Ren, Yaorong Liang, Wenqing Yao, Xiaofeng Yang, and Ling Cen. 2022. DESCN: Deep Entire Space Cross Networks for Individual Treatment Effect Estimation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
[71] Chuan Zhou, Yaxuan Li, Chunyuan Zheng, Haiteng Zhang, Min Zhang, Haoxuan Li, and Mingming Gong. 2025. A Two-Stage Pretraining-Finetuning Framework for Treatment Effect Estimation with Unmeasured Confounding. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
[72] Chuan Zhou, Lina Yao, Haoxuan Li, and Mingming Gong. 2025. Counterfactual Implicit Feedback Modeling. In *Advances in Neural Information Processing Systems*.
[73] Hao Zhou, Rongxiao Huang, Shaoming Li, Guibin Jiang, Jiaqi Zheng, Bing Cheng, and Wei Lin. 2024. Decision Focused Causal Learning for Direct Counterfactual Marketing Optimization. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
[74] Hao Zhou, Shaoming Li, Guibin Jiang, Jiaqi Zheng, and Dong Wang. 2023. Direct Heterogeneous Causal Learning for Resource Allocation Problems in Marketing. In *AAAI Conference on Artificial Intelligence*.
[75] Minqin Zhu, Anpeng Wu, Haoxuan Li, Ruoxuan Xiong, Bo Li, Xiaoqing Yang, Xuan Qin, Peng Zhen, Jiecheng Guo, Fei Wu, et al. 2024. Contrastive Balancing Representation Learning for Heterogeneous Dose-Response Curves Estimation. In *AAAI Conference on Artificial Intelligence*.