Debiased Recommendation Beyond the Positive Propensity Assumption

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

Hao Wang
Zhejiang University
Hangzhou, China
haohaow@zju.edu.cn

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

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

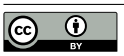
Haoxuan Li*
State Key Lab of General AI, School of Intelligence Science and Technology, Peking University, Beijing, China
hxli@stu.pku.edu.cn

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

Abstract

Post-click conversion rate (CVR) prediction is a central task in recommender systems, yet selection bias creates a severe distributional gap between the clicked training samples and the entire inference space. To address selection bias, propensity-based methods such as inverse propensity scoring (IPS) and doubly robust (DR) have been adopted, which aim to estimate the unbiased learning objective from biased training samples. However, these approaches assume strictly positive propensities, implying every user-item pair has a nonzero probability of interaction. In practice, such positivity assumption maybe violated, for example, in food-delivery platforms, some restaurants located more than 10 kilometers away will be blocked for recommendation. In this study, we theoretically show that when such zero-propensity samples, termed extrapolation samples exist, both IPS and DR estimators become biased. To overcome this limitation, we propose ExtraDebias method, which enables debiased recommendation in both non-extrapolation and extrapolation samples. Specifically, we first train a propensity model to identify extrapolation samples with extremely small propensity estimates, then estimate their pseudo-label intervals, and derive an upper bound of the learning objective for extrapolation samples. By minimizing the derived upper bound, debiased learning on extrapolation samples is ensured, while unbiased learning on non-extrapolation samples is achieved by standard IPS. Experiments on four real-world offline datasets and one online A/B test show that ExtraDebias effectively minimizes prediction errors on extrapolation samples and achieves optimal performance.

*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/10.1145/3805712.3809636

CCS Concepts

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

Keywords

Recommender Systems, Selection Bias, Extrapolation

ACM Reference Format:
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 *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, 11 pages. https://doi.org/10.1145/3805712.3809636

1 INTRODUCTION

Accurate estimation of the Post-click Conversion Rate (CVR) is a critical component of modern recommendation systems [8, 28, 69]. CVR quantifies the conditional probability that a user performs a specific conversion action after clicking on an item. These conversion actions represent the ultimate business objectives across diverse industrial scenarios, encompassing product purchases in e-commerce [23, 24, 29], software downloads in app markets [3], and channel subscriptions in content delivery platforms [4, 65, 68]. Due to its direct correlation with core business metrics, such as user satisfaction and Gross Merchandise Value (GMV) [56], CVR has become a ubiquitous metric, driving continuous advancements in recommendation algorithms.

A critical challenge in CVR prediction is selection bias [57]. Specifically, during training, the dataset is composed exclusively of clicked samples, as conversion labels are only observed given a user click. In contrast, at inference time the model is required to estimate CVRs for all candidates and select the Top-K items for recommendation. Because user clicks are preference-driven rather than random, for example, users are more likely to click items with

attractive titles, samples are not randomly involved in the training data [66]. This induces a distribution shift between training and inference data, causing CVR models trained on clicked-only data to perform poorly at inference [41].

To address selection bias, the inverse propensity scoring (IPS) method [40, 41] estimates the propensity score, defined as the probability of observing a user-item interaction (e.g., a click), and applies inverse-propensity weighting to each observed sample in the training loss. By assigning higher weights to underrepresented samples, IPS theoretically achieves unbiased training objective estimator when propensity scores are accurate. To further enhance robustness, doubly robust (DR) methods [38, 51] combine IPS with an error imputation model. DR enjoys the property of double robustness, ensuring unbiasedness if either the propensity scores or the imputation errors are accurate. Subsequent research has made various improvements to IPS and DR estimators, including enhancing the accuracy of propensity score estimation [1, 18, 19, 52], mitigating hidden confounding bias [6, 16, 17], and integrating with multi-task learning to alleviate data sparsity [47, 60].

Despite their success, such propensity-based debiasing methods assume that the propensity scores of all samples are strictly positive. However, this assumption may be violated in real-world scenarios. For example, on food-delivery platforms, restaurants located more than 10 kilometers away may be blocked from recommendation due to limited delivery capacity; in video recommendation, copyright restrictions may prevent certain videos from being recommended to users in specific regions. In this study, we theoretically show that when there are extrapolation samples, i.e., samples with zero propensities, both IPS and DR estimators become biased. Some might argue that estimating CVR for extrapolation samples is meaningless, as they are currently excluded from exposure by system rules. However, system policies are dynamic and often evolve, for instance, through delivery capacity expansion or new copyright acquisitions, accurately predicting feedback for extrapolation samples carries substantial commercial value. Yet, research addressing this problem remains lacking.

To fill this gap, we propose the Extrapolation Debiasing (ExtraDebias) method, which achieves debiased recommendation on both non-extrapolation and extrapolation samples. The method consists of three main steps as shown in Figure 1. First, we train a propensity model and partition all samples into *interpolation* samples with estimated propensity scores above a threshold, and *extrapolation* samples with scores below the threshold. Second, for interpolation samples, we apply standard debiased learning techniques to ensure unbiased training and compute pseudo-labels for all interpolation samples. Third, for extrapolation samples, we estimate pseudo-label intervals under a Lipschitz continuity assumption, derive an upper bound on the unbiased CVR training objective, and minimize this bound to control the CVR prediction error on extrapolation samples, thereby enabling effective debiased learning. Our contributions can be summarized as follows.

- To our knowledge, we are the first to consider the presence of zero propensity in debiased recommendations, and theoretically prove that previous propensity-based methods such as IPS and DR fail in this scenario.

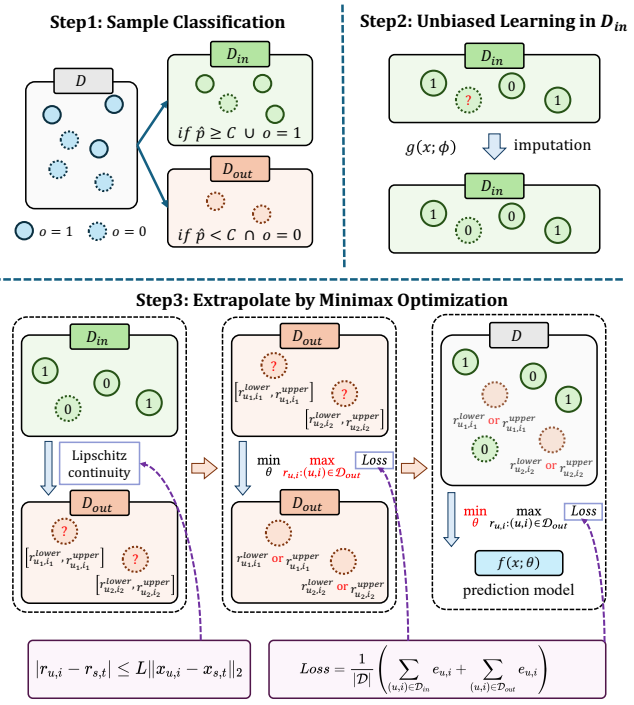


Figure 1: The ExtraDebias framework consists of three steps: Step 1 identifies extrapolation samples, Step 2 performs debiasing on interpolation samples and imputes missing outcomes, and Step 3 estimates pseudo-label intervals for extrapolation samples and minimizes the worst-case risk.

- We propose the ExtraDebias method to enable debiased learning on both interpolation and extrapolation samples, with rigorous theoretical guarantees.
- We conduct experiments on four offline real-world datasets and one online A/B test to validate the effectiveness of ExtraDebias.

2 PRELIMINARY

2.1 Problem Setup

Consider a user set $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$ and an item set $\mathcal{I} = \{i_1, i_2, \ldots, i_m\}$ with n users and m items, respectively. The target population is defined as all samples (user-item pairs), denoted by $\mathcal{D} = \mathcal{U} \times \mathcal{I}$. For each sample $(u, i) \in \mathcal{D}$, let $x_{u,i}$ denote the features or embeddings of sample (u, i). In addition, we denote $y_{u,i} \in \{0, 1\}$ as the conversion label of (u, i). Denote the CVR prediction model as $f(x_{u,i}; \theta)$ and the predicted label as $\hat{y}_{u,i}$. Ideally, if the conversion outcomes $\{y_{u,i} : (u, i) \in \mathcal{D}\}$ are fully observed, we can minimize the following ideal loss for training the CVR prediction model

$$\mathcal{L}_{\text{ideal}}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} e_{u,i}, \tag{1}$$

where $e_{u,i} = \ell(\hat{y}_{u,i}, y_{u,i})$ denotes the prediction error with $\ell(\cdot, \cdot)$ as an arbitrary loss functions such as cross-entropy loss or mean square error. However, computing the ideal loss is impractical due to the missingness of conversion labels for most samples in the impression space.

To formulate the missingness of collected data, we denote the observational indicator as $o_{u,i} \in \{0, 1\}$. Specifically, $o_{u,i} = 1$ means the $y_{u,i}$ is observed and vice versa. Based on this, we obtain observational data and denote it as $O = \{(u, i) \mid (u, i) \in \mathcal{D}, o_{u,i} = 1\}$. The propensity score $p_{u,i}$ is defined as the probability that a user-item interaction being observed, *i.e.,* $p_{u,i} = \mathbb{P}(o_{u,i} = 1 \mid x_{u,i})$. A straightforward approach to train a CVR prediction model involves minimizing the following naive loss

$$\mathcal{L}_{\text{naive}}(\theta) = \frac{1}{|O|} \sum_{(u,i)\in O} e_{u,i}. \tag{2}$$

However, the observational data suffers from selection bias [2, 39], causing the observed population O to be unrepresentative of the target population $\mathcal{D}$. Consequently, the naive estimator is a biased estimator of the ideal loss $\mathbb{E}[\mathcal{L}_{\text{naive}}(\theta)] \neq \mathcal{L}_{\text{ideal}}$.

Blindly minimizing the naive loss leads to a biased CVR prediction model with suboptimal performance. Therefore, constructing unbiased estimator of the ideal loss is critical to achieveing unbiased learning of prediction model for accurate CVR prediction.

2.2 Fundamental Debiasing Estimators

To address selection bias, the inverse propensity score (IPS) estimator [41] aims to estimate the propensity scores, adopts inverse propensity weighting on the observed population, and train the prediction model on the weighted samples, formulated as

$$\mathcal{L}_{\text{IPS}}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}}, \tag{3}$$

where $\hat{p}_{u,i} = \pi(x_{u,i}; \psi)$ represents the estimated propensity score, and the propensity model is denoted as $\pi(x_{u,i}; \psi)$. When the estimated propensity scores are accurate, i.e., $\hat{p}_{u,i} = p_{u,i}$, the IPS estimator is unbiased, *i.e.,* $\mathbb{E}[\mathcal{L}_{\text{IPS}}(\theta)] = \mathcal{L}_{\text{ideal}}$.

The doubly robust (DR) estimator [51] further imputes missing data on top of IPS weighting of observed samples, expressed as

$$\mathcal{L}_{\text{DR}}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left(\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\hat{p}_{u,i}} \right), \tag{4}$$

where $\hat{e}_{u,i} = m(x_{u,i}; \xi)$ is the imputed error for sample (u, i), obtained through the imputation model $m(x_{u,i}; \xi)$, and the training objective of imputation model $m(x_{u,i}; \xi)$ is often given as

$$\mathcal{L}_{\text{imp}}(\xi) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}(\hat{e}_{u,i} - e_{u,i})^2}{\hat{p}_{u,i}}. \tag{5}$$

The DR estimator is unbiased *i.e.,* $\mathbb{E}[\mathcal{L}_{\text{DR}}(\theta)] = \mathcal{L}_{\text{ideal}}$, if either the propensity score is accurately estimated ($\hat{p}_{u,i} = p_{u,i}$), or the error imputation is accurately estimated ($\hat{e}_{u,i} = e_{u,i}$).

Note that existing propensity-based methods, including IPS, DR estimators, and their variants, assume $p_{u,i} > 0$ for all $(u, i) \in \mathcal{D}$ to ensure that denominators in the weights are nonzero. In the next section we will show that when this assumption is violated, both IPS and DR estimators are no longer unbiased.

3 Methodology

The organization of this section is as follows: Section 3.1 first presents the motivation of this work, Section 3.2 details the proposed ExtraDebias method along with its theoretical analysis, and Section 3.3 describes the detailed workflow and learning algorithm.

3.1 Motivation

• **Prevalence of Extrapolation Samples.** Zero-propensity samples are prevalent in industrial recommendation systems. This strict inaccessibility stems from explicit systemic rules and hard constraint embedded within the platform's serving logic, rather than user disinterest. For instance, on food-delivery platforms, a restaurant with high potential preference for a user may be systematically excluded from the candidate set due to rigid delivery radius constraints (e.g., exceeding 10 km). Similarly, in video streaming, highly relevant content can be rendered inaccessible to users in specific regions due to copyright regulations. In such instances, the exposure probability is forced to zero ($p_{u,i} = 0$) by system policies.

• **Significance of Prediction on Extrapolation Samples.** One might overlook these extrapolation samples since they are currently excluded by system rules. However, accurately predicting their potential feedback (e.g., CVR) has substantial commercial value, as platform policies are dynamic and constraints frequently change, for instance, a delivery platform may expand its service radius, or a video platform may acquire new copyrights. When such shifts occur, previously blocked items become viable candidates. A model that generalizes to these extrapolation samples enables counterfactual evaluation, allowing platforms to estimate potential return on investment before deployment. In contrast, a model that fails to generalize leaves strategic decisions unguided, resulting in missed market opportunities.

• **Failure of Existing Estimators.** Despite the critical importance of these extrapolation samples, existing debiasing paradigms are ill-equipped to handle them due to their strict reliance on the positivity assumption. Next, we theoretically demonstrate that this violation renders standard estimators, such as IPS and DR, biased. To illustrate this, we explicitly decompose the ideal learning objective over the entire space $\mathcal{D}$ into two disjoint parts

$$\mathcal{L}_{\text{ideal}}(\theta) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} e_{u,i} = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{in}}} e_{u,i} + \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i}, \tag{6}$$

where $\mathcal{D}_{\text{in}} = \{(u, i) \mid p_{u,i} > 0\}$ and $\mathcal{D}_{\text{out}} = \{(u, i) \mid p_{u,i} = 0\}$.

Theorem 3.1. *In the presence of extrapolation samples, given that $\hat{p}_{u,i} = p_{u,i}$ for all $(u, i) \in \mathcal{D}_{\text{in}}$,*
(a) the bias of IPS estimator is $|\sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i}|/|\mathcal{D}|$,
(b) the bias of DR estimator is $|\sum_{(u,i)\in\mathcal{D}_{\text{out}}} (\hat{e}_{u,i} - e_{u,i})|/|\mathcal{D}|$.

Proof. For (a), with $\hat{p}_{u,i} = p_{u,i}$ for all $(u, i) \in \mathcal{D}_{\text{in}}$, we have

$$\begin{aligned}
&\text{Bias}(\mathcal{L}_{\text{IPS}}(\theta)) = |\mathbb{E}[\mathcal{L}_{\text{IPS}}(\theta)] - \mathcal{L}_{\text{ideal}}(\theta)| \\
&= \left| \frac{1}{|\mathcal{D}|} \mathbb{E}\left[\sum_{(u,i)\in\mathcal{D}_{\text{in}}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}} + \sum_{(u,i)\in\mathcal{D}_{\text{out}}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}} \right] - \mathcal{L}_{\text{ideal}}(\theta) \right|
\end{aligned}$$

$$= \left| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{in}}} \frac{p_{u,i}e_{u,i}}{\hat{p}_{u,i}} - \mathcal{L}_{\text{ideal}}(\theta) \right|$$

$$= \left| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{in}}} \frac{(p_{u,i}-\hat{p}_{u,i})e_{u,i}}{\hat{p}_{u,i}} - \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i} \right|$$

$$= \left| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i} \right| > 0.$$

For (b), with $\hat{p}_{u,i} = p_{u,i}$ for all $(u,i) \in \mathcal{D}_{\text{in}}$, we have

$$\begin{aligned}
&\text{Bias}(\mathcal{L}_{\text{DR}}(\theta)) = |\mathbb{E}[\mathcal{L}_{\text{DR}}(\theta)] - \mathcal{L}_{\text{ideal}}(\theta)| \\
&= \left| \frac{1}{|\mathcal{D}|}\mathbb{E}\left[\sum_{(u,i)\in\mathcal{D}_{\text{in}}} \left(\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}} \right) \right] + \right. \\
&\quad \left. \frac{1}{|\mathcal{D}|}\mathbb{E}\left[\sum_{(u,i)\in\mathcal{D}_{\text{out}}} \left(\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}} \right) \right] - \mathcal{L}_{\text{ideal}}(\theta) \right| \\
&= \left| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{in}}} \left(\hat{e}_{u,i} - e_{u,i} + \frac{p_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}} \right) + \right. \\
&\quad \left. \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{out}}} (\hat{e}_{u,i} - e_{u,i}) \right| = \left| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}_{\text{out}}} (\hat{e}_{u,i} - e_{u,i}) \right|.
\end{aligned}$$

Note that accurately imputing the outcomes (or errors) for extrapolation samples is nearly infeasible. As discussed in Section 2.2, existing imputation-based training methods in Equation 5 still rely on nonzero propensities and therefore cannot guarantee accurate imputation for extrapolation samples. Consequently, the bias is non-zero, i.e., $\text{Bias}(\mathcal{L}_{\text{DR}}(\theta)) > 0$.

□

This theoretical limitation highlights a critical gap in current literature, underscoring the urgent need for novel methodologies capable of ensuring robust generalization even in the presence of zero-propensity extrapolation samples.

3.2 ExtraDebias

Figure 1 illustrates the overall framework of proposed ExtraDebias method, and we begin the detailed description by introducing the core Lipschitz continuity assumption.

Assumption 1 (*L*-Lipschitz). *For all samples (u,i) and (s,t) in $\mathcal{D}$, we have $|r_{u,i} - r_{s,t}| \le L\|x_{u,i} - x_{s,t}\|_2$, where x and r denotes the learned continuous latent embedding and CVR.*

It is important to clarify that this assumption applies to the relationship between the **continuous latent embeddings** and the underlying **CVR**, rather than between discrete raw features and binary observed labels. Specifically, discrete features (e.g., user/item IDs, age, country) are mapped to continuous embeddings, where the Lipschitz condition guarantees that similar embeddings imply similar CVRs, even though the observed binary labels sampled from these probabilities may differ due to stochasticity. This assumption effectively bridges zero and positive propensity samples by bounding the feasible outcomes for samples in $\mathcal{D}_{\text{out}}$. Next, we describe in detail the three steps of ExtraDebias.

- **Step 1: Identifying extrapolation samples.** We first learn a propensity model $\pi(x_{u,i};\psi)$ to generate the estimated propensity score $\hat{p}_{u,i}$ for each sample, and then identify sample (u,i) with $\hat{p}_{u,i} < C$ and $o_{u,i} = 0$ as extrapolation samples, forming the set $\mathcal{D}_{\text{out}}$. The remaining samples, i.e., those with $\hat{p}_{u,i} \ge C$ or $o_{u,i} = 1$, are regarded as interpolation samples and constitute the set $\mathcal{D}_{\text{in}}$. The threshold C is a tunable hyperparameter.
- **Step 2: Unbiased learning on interpolation samples.** For interpolation samples $(u,i) \in \mathcal{D}_{\text{in}}$, we train an unbiased imputation model $g(x_{u,i};\phi)$ using existing debiasing objectives such as IPS or DR loss. The learned model $g(x_{u,i};\phi)$ provides pseudo-labels for all interpolation samples $(u,i) \in \mathcal{D}_{\text{in}}$.
- **Step 3: Unbiased learning on extrapolation samples.** Finally, we train the CVR prediction model $f(x_{u,i};\theta)$ using both interpolation and extrapolation samples, where L-Lipschitz continuity offers CVR label interval for extrapolation samples. The CVR prediction model is then trained by minimizing the worst-case risk, formulated as the following optimization problem

$$\begin{aligned}
&\min_{\theta} \max_{r_{u,i}:(u,i)\in\mathcal{D}_{\text{out}}} \frac{1}{|\mathcal{D}|}\left(\sum_{(u,i)\in\mathcal{D}_{\text{in}}} e_{u,i} + \sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i} \right) \\
&\text{s.t. } r_{u,i} = g(x_{u,i};\phi), \quad \text{for all } (u,i) \in \mathcal{D}_{\text{in}}, \\
&\quad |r_{u,i} - r_{s,t}| \le L\|x_{u,i} - x_{s,t}\|_2, \quad \text{for all } (u,i) \text{ and } (s,t) \in \mathcal{D}, \\
&\quad e_{u,i} = \ell(f(x_{u,i};\theta), r_{u,i}), \quad \text{for all } (u,i) \in \mathcal{D}.
\end{aligned} \tag{7}$$

The unbiased pseudo-labels for all interpolation samples generated in Step 2 support unbiased learning on interpolation samples in Step 3 and provide pseudo-label intervals for extrapolation samples under the Lipschitz continuity assumption. By further minimizing the worst-case risk over extrapolation samples defined by these intervals, the CVR model minimizes prediction errors on extrapolation samples and thus achieves debiased learning.

Theoretical Justification. To efficiently solve the minimax problem in Equation 7, we derive a closed-form solution for the inner maximization step. First, we establish tight bounds for the outcomes of extrapolation samples.

Theorem 3.2 (Bounds of Extrapolation Outcomes). *For extrapolation sample $(u,i) \in \mathcal{D}_{\text{out}}$, the tight upper bound of $r_{u,i}$ is*

$$r_{u,i} \le r_{u,i}^{\text{upper}} := \min_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} + L\|x_{u,i} - x_{s,t}\|_2,$$

and the tight lower bound of $r_{u,i}$ is

$$r_{u,i} \ge r_{u,i}^{\text{lower}} := \max_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} - L\|x_{u,i} - x_{s,t}\|_2.$$

Proof. For the upper bound of $r_{u,i}$, according to Assumption 1, for any $(s,t) \in \mathcal{D}_{\text{in}}$, we have

$$r_{u,i} \le r_{s,t} + L\|x_{u,i} - x_{s,t}\|_2$$

Since this inequality holds for all $(s,t) \in \mathcal{D}_{\text{in}}$, taking the minimum over all such pairs yields

$$r_{u,i} \le \min_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} + L\|x_{u,i} - x_{s,t}\|_2 = r_{u,i}^{\text{upper}}.$$

Similarly, for the lower bound of $r_{u,i}$, for any $(s,t) \in \mathcal{D}_{\text{in}}$, we have

$$r_{u,i} \geq r_{s,t} - L\|x_{u,i} - x_{s,t}\|_2$$

Taking the maximum over all $(s,t) \in \mathcal{D}_{\text{in}}$ gives

$$r_{u,i} \geq \max_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} - L\|x_{u,i} - x_{s,t}\|_2 = r_{u,i}^{\text{lower}}.$$

□

Theorem 3.2 theoretically guarantees that the unknown ground-truth CVR outcomes for extrapolation samples are confined within a computable range derived from the interpolation samples. Building on this, we further derive the tight upper bound for the prediction errors on these extrapolation samples.

Corollary 3.3 (Tight Upper Bound of Extrapolation Prediction Errors). *For extrapolation sample $(u,i) \in \mathcal{D}_{\text{out}}$, the tight upper bound of $e_{u,i}$ is*

$$e_{u,i} \leq e_{u,i}^{\text{upper}} := \max\{\ell(r_{u,i}^{\text{upper}}, f(x_{u,i};\theta)), \ell(r_{u,i}^{\text{lower}}, f(x_{u,i};\theta))\}.$$

This corollary implies that the intractable inner maximization problem can be simplified into a direct comparison of losses at the two interval boundaries. This simplification drastically reduces computational complexity, enabling efficient end-to-end training of CVR model without iterative adversarial optimization.

In addition, determining the Lipschitz constant L is crucial for practical implementation. While L is often treated as an unknown hyperparameter, we provide a theoretical estimation for the widely-used Matrix Factorization (MF) [11] architecture.

Theorem 3.4 (Lipschitz Constant for MF). *For a Matrix Factorization model where $f(x_{u,i}) = \sigma(p_u^T q_i)$ with normalized user and item embeddings $||p_u|| = ||q_i|| = 1$, the Lipschitz constant with respect to the Euclidean distance of embeddings is exactly $L = \frac{\sqrt{2}}{4}$.*

Proof. Firstly, we derive the Lipschitz constant for the function $h(p,q) = p^T q$. Consider any $p, q, p', q' \in \mathbb{R}^d$, we have

$$\begin{aligned}
|p^T q - p'^T q'| &= |p^T q - p^T q' + p^T q' - p'^T q'| \\
&\leq |p^T (q - q')| + |(p - p')^T q'| \\
&\leq ||p|| \cdot ||q - q'|| + ||q'|| \cdot ||p - p'|| \\
&\leq ||q - q'|| + ||p - p'|| \\
&\leq \sqrt{2} \cdot \sqrt{||q - q'||^2 + ||p - p'||^2},
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second from the Cauchy–Schwarz inequality, the third from the fact that $|p| = |q| = 1$, and the final one from the inequality that $(a+b)^2 \leq 2(a^2 + b^2)$.

The final estimated CVR $\hat{r} = \sigma(h(p,q))$, where $\sigma(h) = 1/(1 + e^{-h})$. Note that $\sigma'(h) = \sigma(h)(1 - \sigma(h)) \leq 1/4$, then we have

$$\begin{aligned}
|\sigma(h(p,q)) - \sigma(h(p',q'))| &\leq \frac{1}{4}|h(p,q) - h(p',q')| \\
&\leq \frac{\sqrt{2}}{4} \cdot \sqrt{||q - q'||^2 + ||p - p'||^2}.
\end{aligned}$$

Therefore, the Lipschitz constant $L = \frac{\sqrt{2}}{4}$. □

Theorem 3.4 provides a closed-form reference value for L, eliminating the need for expensive grid searches in MF-based deployments. In practice, we introduce a scaling factor τ such that $L = \frac{\sqrt{2}}{4}\tau$, where τ acts as a risk tolerance parameter, where larger τ implies a more conservative extrapolation strategy by allowing larger outcome variations.

Algorithm 1 The ExtraDebias Learning Algorithm.

1: **Input:** Observed samples $\mathcal{O}$, all samples $\mathcal{D}$, threshold C, and Lipschitz constant L.
2: Initialize propensity model $\pi(x_{u,i};\psi)$, imputation model $g(x_{u,i};\phi)$, and prediction model $f(x_{u,i};\theta)$.
3: **while** stopping criteria is not satisfied **do**
4: Sample a batch of samples $\{(u_j, i_j)\}_{j=1}^{J}$ from $\mathcal{D}$;
5: Update ψ by descending along the gradient $\nabla_\psi \mathcal{L}_{\text{ce}}(\psi)$;
6: **end while**
7: Compute propensity $\hat{p}_{u,i} = \pi(x_{u,i};\psi)$ for (u,i) in $\mathcal{D}$;
8: Construct $\mathcal{D}_{\text{in}} = \{(u,i) \mid (u,i) \in \mathcal{D}, \hat{p}_{u,i} \geq C \text{ or } o_{u,i} = 1\}$.
9: Construct $\mathcal{D}_{\text{out}} = \{(u,i) \mid (u,i) \in \mathcal{D}, \hat{p}_{u,i} < C \text{ and } o_{u,i} = 0\}$.
10: **while** stopping criteria is not satisfied **do**
11: Sample a batch of samples $\{(u_k, i_k)\}_{k=1}^{K}$ from $\mathcal{D}_{\text{in}}$;
12: Update ϕ by descending along the gradient $\nabla_\phi \mathcal{L}_{\text{IPS}}(\phi)$;
13: **end while**
14: $r_{u,i} \leftarrow g(x_{u,i};\phi)$, for $(u,i) \in \mathcal{D}_{\text{in}}$.
15: $r_{u,i}^{\text{upper}} \leftarrow \min_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} + L||x_{u,i} - x_{s,t}||_2$, for $(u,i) \in \mathcal{D}_{\text{out}}$.
16: $r_{u,i}^{\text{lower}} \leftarrow \max_{(s,t)\in\mathcal{D}_{\text{in}}} r_{s,t} - L||x_{u,i} - x_{s,t}||_2$, for $(u,i) \in \mathcal{D}_{\text{out}}$.
17: **while** stopping criteria is not satisfied **do**
18: Sample a batch of samples $\{(u_l, i_l)\}_{l=1}^{L}$ from $\mathcal{D}$;
19: Update θ by descending along the gradient $\nabla_\theta \mathcal{L}_{\text{ext}}(\theta)$;
20: **end while**

One key distinction warrants emphasis: although Theorem 3.4 is derived for MF, this does not mean our method is limited to MF. For more complex deep learning backbones (e.g., DNNs), where analytical derivation is difficult, the Lipschitz constant L can be approximated via spectral norm estimation or just tuned as a hyperparameter, ensuring the broad applicability of our approach. In subsequent experiments, we validate this claim using the widely adopted DCN [49] backbone in industrial settings.

3.3 The Workflow of ExtraDebias

In this section, we provide the training algorithm for the ExtraDebias in Algorithm 1. The detailed procedure is described as follows.
• Firstly, we train a propensity model $\pi(x_{u,i};\psi)$ with the widely adopted cross entropy loss (lines 3-6)

$$\mathcal{L}_{\text{ce}}(\psi) = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} CE(\pi(x_{u,i};\psi), o_{u,i}) + \lambda_{\text{ce}}||\psi||_F^2, \tag{8}$$

where $CE(\cdot,\cdot)$ is the binary cross-entropy loss, $\lambda_{\text{ce}} > 0$ is a hyperparameter, and $||\psi||_F^2$ is the Frobenius norm. After training the propensity model, we impute propensity $\hat{p}_{u,i} = \pi(x_{u,i};\psi)$ for (u,i) in $\mathcal{D}$, and then divide $\mathcal{D}$ into two disjoint sets (lines 7-9)

$$\begin{aligned}
\mathcal{D}_{\text{in}} &= \{(u,i) \mid (u,i) \in \mathcal{D}, \hat{p}_{u,i} \geq C \text{ or } o_{u,i} = 1\}, \\
\mathcal{D}_{\text{out}} &= \{(u,i) \mid (u,i) \in \mathcal{D}, \hat{p}_{u,i} < C \text{ and } o_{u,i} = 0\}.
\end{aligned} \tag{9}$$

• Secondly, we learn an unbiased imputation model $g(x_{u,i};\phi)$ by minimizing an existing debiasing loss across the samples in $\mathcal{D}_{\text{in}}$ (lines 10-13). This step can accommodate any advanced debiasing method. In this paper, we employ the standard IPS method as an

illustrative example for its simplicity and effectiveness

$$\mathcal{L}_{\text{IPS}}(\phi) = \frac{1}{|\mathcal{D}_{\text{in}}|} \sum_{(u,i)\in\mathcal{D}_{\text{in}}} \frac{o_{u,i}\ell(g(x_{u,i};\phi), r_{u,i})}{\hat{p}_{u,i}} + \lambda_{\text{IPS}}||\phi||_F^2, \quad (10)$$

where $\lambda_{\text{IPS}} > 0$ is a hyperparameter. The learned imputation model $g(x_{u,i};\phi)$ outputs the imputed CVR outcomes for all interpolation samples $(u,i) \in \mathcal{D}_{\text{in}}$ and further provides interval estimates $[r_{u,i}^{\text{lower}}, r_{u,i}^{\text{upper}}]$ of the CVR outcomes for all extrapolation samples $(u,i) \in \mathcal{D}_{\text{out}}$ based on the Lipschitz continuity assumption, as stated in Theorem 3.2 (lines 14–16).

• Finally, we compute the upper bound of the prediction error for extrapolation samples using the closed-form solution in Corollary 3.3, forming the worst-case risk $\mathcal{L}_{\text{out}}(\theta)$. Combined with the base prediction loss on interpolation samples $\mathcal{L}_{\text{in}}(\theta)$, this forms the total training loss for the CVR model $f(x_{u,i};\theta)$

$$\mathcal{L}_{\text{ext}}(\theta) = \mathcal{L}_{\text{in}}(\theta) + \mathcal{L}_{\text{out}}(\theta)$$
$$= \frac{1}{|\mathcal{D}|}\left(\sum_{(u,i)\in\mathcal{D}_{\text{in}}} \ell(f(x_{u,i};\theta), g(x_{u,i};\phi)) + \sum_{(u,i)\in\mathcal{D}_{\text{out}}} e_{u,i}^{\text{upper}}\right) + \lambda_{\text{ext}}||\theta||_F^2, \quad (11)$$

where $\lambda_{\text{ext}} > 0$ is a hyperparameter. The CVR model $f(x_{u,i};\theta)$ is optimized using $\mathcal{L}_{\text{ext}}(\theta)$ (lines 17–20) to minimize prediction errors on both interpolation and extrapolation samples.

4 EXPERIMENTS

In this section, we conduct experiments to answer the following research questions (RQs):

- RQ1: Does the proposed ExtraDebias method effectively address the zero propensity challenge and outperform existing state-of-the-art (SOTA) debiasing methods?
- RQ2: Does the proposed extrapolation error control strategy (minimax optimization) improve model performance?
- RQ3: How do different components (e.g., C in extrapolation sample identification and τ for Lipschitz constant estimation) influence the performance?
- RQ4: Does ExtraDebias achieve better performance in real-world application scenarios?

4.1 Experimental Settings

In this study, we conduct experiments on four real-world offline datasets, including three public datasets and one private industrial dataset. In addition, we perform online experiments to validate the performance of the proposed method in real-world settings.

4.1.1 Public Datasets. Our experiments are conducted on three real-world datasets: Coat[1] [41], Yahoo! R3[2] [31], and KuaiRec[3] [7], which cover music, short video, and coat recommendation tasks, respectively. These datasets provide unbiased data collected through random exposure policies to validate methods' debiasing performance. Following [15, 22], we binarize the ratings for Coat and Yahoo! R3, setting ratings less than 4 to negative feedback 0 and those of 4 or higher to positive feedback 1. For KuaiRec, we binarize the watching ratio by setting values less than 2 to 0 and those equal to or greater than 2 to 1, as suggested in prior study [7]. Although these widely adopted public datasets provide unbiased data for evaluation, they lack an explicit partition of zero-propensity extrapolation samples. To simulate such setting, we use the isolation forest algorithm [26] to remove outliers from the training set while keeping the test set intact. This preprocessing creates a more compact training space, increasing the prediction difficulty for potential extrapolation samples and providing a more suitable testbed for our method.

4.1.2 Private Datasets. To further simulate real zero-propensity scenarios, we manually intervened in business data and applied system-level exposure blocking based on explicit feature semantics to create real zero-propensity regions. Specifically, using offline logs from our production recommender system, we split the data into training and test sets chronologically. We then applied policy-based blocking rules (e.g., excluding Michelin restaurants for students) to the training data, yielding a training set of 8 million interactions that contains zero-propensity samples. The test data were left unaltered, resulting in both an unrestricted test set of 3 million interactions and a dedicated extrapolation subset of 0.8 million interactions that is subject only to the blocking rules (e.g., Michelin restaurants for students). The resulting datasets comprise 56 sparse features, 306 continuous features, and both click and conversion labels.

4.1.3 Baselines. We compare ExtraDebias with the following baselines: (1) **Naive estimator**: which directly fits the observational data without debiasing shown in Equation 2; (2) **Information Bottleneck (IB) based estimator**: CVIB [54], DIB [25]; (3) **IPS based estimators**: IPS [41], SNIPS [45], ASIPS [37], ESCM^2-IPS [47], IPS-V2 [18], AKBIPS [19]; (4) **DR based estimators**: DR-JL [51], MRDR-JL [9], DR-BIAS [5], DR-MSE [5], TDR-JL [15], StableDR [22], ESCM^2-DR [47], DR-V2 [18], AKBDR [19], DCE-DR [12]. On public datasets, following prior studies [12, 15, 22, 37, 51], and to ensure fair comparison, all baseline methods adopt logistic regression for the propensity score model, while both the prediction and imputation models are implemented using Matrix Factorization (MF) [11]. On the private industrial dataset, we adopt a Deep & Cross Network (DCN) [49] as the backbone.

4.1.4 Experimental Details. To implement ExtraDebias, we set the threshold C for extrapolation sample identification as the Q-th percentile of the propensity scores over all samples, where the hyperparameter Q is tuned from the set $\{0.1, 0.5, 1, 5, 10\}$. We tune the learning rate in $\{0.01, 0.03, 0.05\}$, weight decay in $\{10^{-6}, 5\times 10^{-6}, \ldots, 10^{-1}\}$, and the scaling hyperparameter τ for Lipschitz constant estimation in $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$. On public datasets, we evaluate the prediction performance using AUC, Recall@K, and NDCG@K, where $K = 5$ for the Coat and Yahoo! R3 datasets, and $K = 50$ for the KuaiRec dataset. On the private industrial dataset, we employ AUC, KS (Kolmogorov-Smirnov), and LogLoss for evaluation. All experiments on public datasets are conducted on an NVIDIA A100 GPU using PyTorch. Experiments on the private dataset and the online A/B test are performed using a TensorFlow-based deep learning framework.

[1] https://www.cs.cornell.edu/~schnabts/mnar/
[2] https://www.kaggle.com/datasets/limitiao/yahoor3
[3] https://kuairec.com/

Table 1: Performance comparison on three public datasets. The best results are shown in bold, and the best baseline results are underlined. * indicates statistical significance based on a paired t-test at the 0.05 level compared to the best baseline.

Method	Coat			Yahoo! R3			KuaiRec		
	AUC	Recall@5	NDCG@5	AUC	Recall@5	NDCG@5	AUC	Recall@50	NDCG@50
Naive	$0.695_{\pm 0.009}$	$0.425_{\pm 0.012}$	$0.605_{\pm 0.013}$	$0.670_{\pm 0.002}$	$0.390_{\pm 0.003}$	$0.632_{\pm 0.003}$	$0.758_{\pm 0.001}$	$0.605_{\pm 0.002}$	$0.546_{\pm 0.004}$
CVIB	$0.719_{\pm 0.004}$	$0.434_{\pm 0.008}$	$0.619_{\pm 0.010}$	$0.688_{\pm 0.001}$	$0.425_{\pm 0.002}$	$0.656_{\pm 0.002}$	$0.773_{\pm 0.002}$	$0.634_{\pm 0.005}$	$0.556_{\pm 0.003}$
DIB	$0.714_{\pm 0.006}$	$0.430_{\pm 0.009}$	$0.621_{\pm 0.011}$	$0.690_{\pm 0.001}$	$0.421_{\pm 0.002}$	$0.654_{\pm 0.002}$	$0.770_{\pm 0.002}$	$0.636_{\pm 0.004}$	$0.560_{\pm 0.003}$
IPS	$0.716_{\pm 0.005}$	$0.435_{\pm 0.010}$	$0.615_{\pm 0.011}$	$0.679_{\pm 0.002}$	$0.416_{\pm 0.004}$	$0.639_{\pm 0.003}$	$0.763_{\pm 0.004}$	$0.634_{\pm 0.006}$	$0.554_{\pm 0.007}$
SNIPS	$0.720_{\pm 0.005}$	$0.440_{\pm 0.009}$	$0.618_{\pm 0.010}$	$0.683_{\pm 0.001}$	$0.415_{\pm 0.002}$	$0.644_{\pm 0.003}$	$0.764_{\pm 0.003}$	$0.640_{\pm 0.005}$	$0.559_{\pm 0.006}$
ASIPS	$0.712_{\pm 0.008}$	$0.438_{\pm 0.013}$	$0.622_{\pm 0.014}$	$0.678_{\pm 0.001}$	$0.414_{\pm 0.003}$	$0.637_{\pm 0.002}$	$0.762_{\pm 0.005}$	$0.628_{\pm 0.010}$	$0.555_{\pm 0.006}$
ESCM^2-IPS	$0.728_{\pm 0.004}$	$0.447_{\pm 0.009}$	$0.638_{\pm 0.008}$	$0.690_{\pm 0.001}$	$0.426_{\pm 0.003}$	$0.659_{\pm 0.003}$	$0.772_{\pm 0.002}$	$0.640_{\pm 0.004}$	$0.556_{\pm 0.005}$
IPS-V2	$0.730_{\pm 0.004}$	$0.450_{\pm 0.006}$	$0.640_{\pm 0.007}$	$0.689_{\pm 0.002}$	$0.428_{\pm 0.002}$	$0.657_{\pm 0.002}$	$0.774_{\pm 0.003}$	$0.640_{\pm 0.005}$	$0.549_{\pm 0.004}$
AKBIPS	$0.729_{\pm 0.006}$	$0.445_{\pm 0.011}$	$0.632_{\pm 0.009}$	$0.692_{\pm 0.002}$	$0.432_{\pm 0.004}$	$0.660_{\pm 0.003}$	$0.770_{\pm 0.004}$	$0.643_{\pm 0.006}$	$0.559_{\pm 0.006}$
DR-JL	$0.730_{\pm 0.004}$	$0.440_{\pm 0.009}$	$0.627_{\pm 0.010}$	$0.681_{\pm 0.002}$	$0.426_{\pm 0.003}$	$0.648_{\pm 0.002}$	$0.768_{\pm 0.004}$	$0.637_{\pm 0.006}$	$0.557_{\pm 0.009}$
MRDR-JL	$0.733_{\pm 0.003}$	$0.444_{\pm 0.009}$	$0.629_{\pm 0.008}$	$0.682_{\pm 0.002}$	$0.427_{\pm 0.003}$	$0.651_{\pm 0.003}$	$0.771_{\pm 0.004}$	$0.635_{\pm 0.006}$	$0.560_{\pm 0.007}$
DR-BIAS	$0.735_{\pm 0.003}$	$0.445_{\pm 0.008}$	$0.633_{\pm 0.007}$	$0.684_{\pm 0.002}$	$0.429_{\pm 0.003}$	$0.653_{\pm 0.002}$	$0.770_{\pm 0.003}$	$0.637_{\pm 0.005}$	$0.562_{\pm 0.006}$
DR-MSE	$0.736_{\pm 0.004}$	$0.448_{\pm 0.009}$	$0.635_{\pm 0.008}$	$0.687_{\pm 0.002}$	$0.430_{\pm 0.004}$	$0.659_{\pm 0.004}$	$0.772_{\pm 0.004}$	$0.640_{\pm 0.006}$	$0.560_{\pm 0.008}$
TDR-JL	$0.732_{\pm 0.005}$	$0.443_{\pm 0.007}$	$0.630_{\pm 0.009}$	$0.689_{\pm 0.002}$	$0.424_{\pm 0.004}$	$0.658_{\pm 0.004}$	$0.775_{\pm 0.004}$	$0.636_{\pm 0.004}$	$0.554_{\pm 0.003}$
StableDR	$0.738_{\pm 0.005}$	$0.449_{\pm 0.006}$	$0.638_{\pm 0.007}$	$0.688_{\pm 0.002}$	$0.430_{\pm 0.004}$	$0.661_{\pm 0.003}$	$0.773_{\pm 0.002}$	$0.633_{\pm 0.003}$	$0.553_{\pm 0.003}$
ESCM^2-DR	$0.738_{\pm 0.007}$	$0.448_{\pm 0.009}$	$0.642_{\pm 0.006}$	$0.689_{\pm 0.001}$	$0.433_{\pm 0.003}$	$0.663_{\pm 0.003}$	$0.779_{\pm 0.005}$	$0.644_{\pm 0.006}$	$0.562_{\pm 0.006}$
DR-V2	$0.740_{\pm 0.008}$	$0.451_{\pm 0.008}$	$0.643_{\pm 0.006}$	$0.691_{\pm 0.002}$	$0.432_{\pm 0.002}$	$0.662_{\pm 0.003}$	$0.777_{\pm 0.007}$	$0.644_{\pm 0.007}$	$0.563_{\pm 0.008}$
AKBDR	$\underline{0.743}_{\pm 0.004}$	$\underline{0.456}_{\pm 0.007}$	$\underline{0.646}_{\pm 0.005}$	$0.694_{\pm 0.002}$	$0.435_{\pm 0.004}$	$0.665_{\pm 0.004}$	$0.782_{\pm 0.004}$	$\underline{0.648}_{\pm 0.005}$	$\underline{0.567}_{\pm 0.007}$
DCE-DR	$0.740_{\pm 0.002}$	$0.450_{\pm 0.006}$	$0.640_{\pm 0.007}$	$\underline{0.697}_{\pm 0.002}$	$\underline{0.438}_{\pm 0.003}$	$\underline{0.668}_{\pm 0.004}$	$\underline{0.786}_{\pm 0.004}$	$0.647_{\pm 0.005}$	$0.565_{\pm 0.005}$
ExtraDebias	$\mathbf{0.746}_{\pm 0.003}$	$\mathbf{0.465}^{*}_{\pm 0.006}$	$\mathbf{0.658}^{*}_{\pm 0.006}$	$\mathbf{0.700}^{*}_{\pm 0.001}$	$\mathbf{0.442}^{*}_{\pm 0.002}$	$\mathbf{0.673}^{*}_{\pm 0.003}$	$\mathbf{0.796}^{*}_{\pm 0.002}$	$\mathbf{0.653}_{\pm 0.004}$	$\mathbf{0.578}^{*}_{\pm 0.007}$

Table 2: Ablation study on the training loss components of the ExtraDebias method. The competitive baseline DCE-DR is used for comparison against ExtraDebias and its ablated variants. The best result is shown in bold, and the second-best is underlined.

Method	Training loss		Coat			Yahoo! R3			KuaiRec		
	$\mathcal{L}_{\text{in}}$	$\mathcal{L}_{\text{out}}$	AUC	Recall@5	NDCG@5	AUC	Recall@5	NDCG@5	AUC	Recall@50	NDCG@50
DCE-DR	×	×	$\underline{0.740}_{\pm 0.002}$	$\underline{0.450}_{\pm 0.006}$	$0.640_{\pm 0.007}$	$\underline{0.697}_{\pm 0.002}$	$\underline{0.438}_{\pm 0.003}$	$\underline{0.668}_{\pm 0.004}$	$0.786_{\pm 0.004}$	$\underline{0.647}_{\pm 0.005}$	$0.565_{\pm 0.005}$
ExtraDebias†	✓	×	$0.737_{\pm 0.003}$	$0.449_{\pm 0.006}$	$\underline{0.646}_{\pm 0.005}$	$0.693_{\pm 0.001}$	$0.435_{\pm 0.003}$	$0.662_{\pm 0.002}$	$\underline{0.793}_{\pm 0.001}$	$0.646_{\pm 0.006}$	$\underline{0.569}_{\pm 0.006}$
ExtraDebias‡	×	✓	$0.583_{\pm 0.015}$	$0.326_{\pm 0.016}$	$0.489_{\pm 0.019}$	$0.536_{\pm 0.009}$	$0.377_{\pm 0.012}$	$0.594_{\pm 0.016}$	$0.580_{\pm 0.006}$	$0.589_{\pm 0.0016}$	$0.403_{\pm 0.021}$
ExtraDebias	✓	✓	$\mathbf{0.746}_{\pm 0.003}$	$\mathbf{0.465}^{*}_{\pm 0.006}$	$\mathbf{0.658}^{*}_{\pm 0.006}$	$\mathbf{0.700}^{*}_{\pm 0.001}$	$\mathbf{0.442}^{*}_{\pm 0.002}$	$\mathbf{0.673}^{*}_{\pm 0.003}$	$\mathbf{0.796}^{*}_{\pm 0.002}$	$\mathbf{0.653}_{\pm 0.004}$	$\mathbf{0.578}^{*}_{\pm 0.007}$

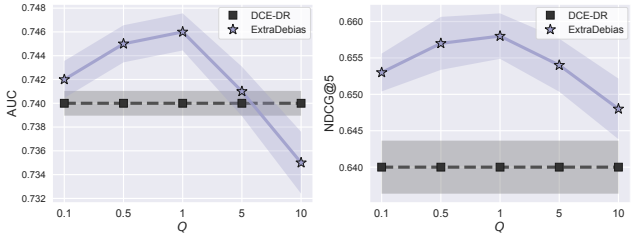
4.2 Performance Comparison (RQ1)

Table 1 presents the prediction performance of various debiasing methods across three public datasets. The Naive method, which ignores selection bias, exhibits the worst performance. To address this issue, IPS and DR-JL were proposed, both of which significantly improve prediction performance, demonstrating the effectiveness of propensity-based methods in mitigating selection bias. Building on these, classical variance reduction techniques such as SNIPS and StableDR further alleviate the high variance caused by small propensity scores, leading to enhanced prediction accuracy. Among these baselines, the most competitive methods are AKBDR and DCE-DR, both of which aim to improve propensity estimation. Specifically, AKBDR introduces a regularization term that adaptively selects kernel functions to enforce balance in the propensity model, while DCE-DR proposes a calibration loss based on the propensity model outputs and the observation indicators to guide propensity learning. In comparison, the proposed ExtraDebias method achieves the best overall performance on all three datasets. This result suggests that, compared to existing variance reduction and enhanced propensity learning strategies, ExtraDebias more effectively addresses the extrapolation error caused by zero propensity scores, thereby yielding significant performance gains.

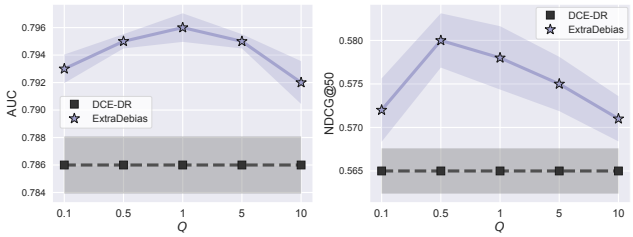
4.3 Ablation Study (RQ2)

Table 2 investigates the roles of the interpolation and extrapolation losses in CVR training within the ExtraDebias method, and the main findings are as follows:

- ExtraDebias† retains only $\mathcal{L}_{\text{in}}$ loss to ensure unbiased learning on interpolation samples but degrades the prediction ability to generalize to extrapolation samples, and its performance remains comparable to DCE-DR.
- ExtraDebias‡ relies exclusively on the extrapolation loss $\mathcal{L}_{\text{out}}$, leading to a significant performance drop, which indicates that extrapolation is infeasible without the fundamental supervision provided by high-confidence interpolation samples.

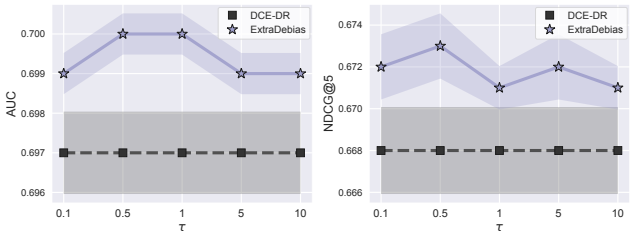


(a) Coat dataset performance.

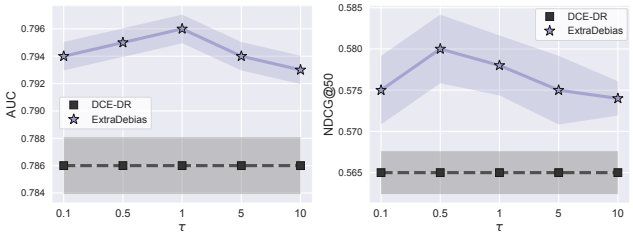


(b) KuaiRec dataset performance.

Figure 2: Effect of threshold C (tuned by hyper-parameter Q) on prediction performance.



(a) Yahoo! R3 dataset performance.



(b) KuaiRec dataset performance.

Figure 3: Effect of scaling hyper-parameter τ for Lipschitz constant estimation on prediction performance.

- ExtraDebias integrates both $\mathcal{L}_{\text{in}}$ and $\mathcal{L}_{\text{out}}$ to achieve effective debiasing on both interpolation and extrapolation samples, and demonstrates the optimal performance.

4.4 In-Depth Analysis (RQ3)

Extrapolation Sample Identification. Figure 2 illustrates the impact of the threshold C in extrapolation sample identification, where C is set as the Q-th percentile of estimated propensity scores over all samples. The lines and error bars indicate the mean performances and 90% confidence intervals. The results show that the best performance is achieved when Q takes a moderate value, for example $Q = 0.5$ or $Q = 1$. If Q is set too large, more samples are classified as extrapolation samples, introducing a substantial number of misclassified noisy-labeled samples and leading to performance degradation. Conversely, when Q is set too small, only a very limited number of samples are identified as extrapolation samples. In this case, the proposed worst-case risk on extrapolation samples has little effect, and performance gains are constrained.

Lipschitz Smoothness Strength. Figure 3 illustrates the effect of the scaling hyperparameter τ in Lipschitz constant estimation, where the lines and error bars indicate the mean performances and 90% confidence intervals. A smaller τ yields a smaller Lipschitz constant L and a narrower pseudo-label interval for extrapolation samples, corresponding to a more aggressive extrapolation strategy that may underestimate the upper bound of the prediction error. As shown in the figure, prediction performance is largely insensitive to the choice of τ. Even with substantial variation in τ, the proposed method consistently outperforms the competitive baseline DCE-DR by a notable margin. Moreover, the results indicate that the dependence on the Lipschitz continuity assumption is mild, as strong performance is maintained even for large τ values (e.g., $\tau = 10$), which correspond to a weak smoothness requirement.

4.5 Real-world Application Performance (RQ4)

Performance on Private Industrial Dataset. Table 3 compares the ability of different CVR prediction methods to handle extrapolation error in real-world applications. All methods are trained on data containing zero-propensity extrapolation regions induced by deliberate intervention, and evaluated on both an Unrestricted Test Set that includes interpolation and extrapolation samples, and an Extrapolation Test Subset consisting only of extrapolation samples. On this industrial dataset, where interactions are extremely sparse, ESCM^2-DR, which incorporates multi-task learning to alleviate data sparsity, achieves the best performance among the baselines. The proposed ExtraDebias attains the best results across all metrics and delivers substantial improvements on the extrapolation test subset. These results demonstrate that ExtraDebias effectively addresses extrapolation error in real-world settings and further enhances prediction performance.

Online A/B Test. To validate ExtraDebias in a practical industrial setting, we conducted online A/B testing on a large-scale recommendation platform. We deployed both ExtraDebias and the competitive industrial baseline, ESCM^2-DR, using a TensorFlow-based machine learning infrastructure. It is crucial to note that items restricted as zero-propensity samples during offline training may become viable candidates during online inference due to policy updates, thereby serving as a proper testbed for the model's extrapolation generalization. Table 4 reports the relative improvements of ExtraDebias over the baseline ESCM^2-DR in CVR and Return on Investment (ROI). Results over five consecutive days show that ExtraDebias

Table 3: Performance comparison on the private industrial dataset. The best results are shown in bold, and the best baseline results are underlined. "Imp. ↑" indicates the improvement of ExtraDebias over the optimal baseline ESCM^2-DR.

Method	Unrestricted Test Set			Extrapolation Test Subset		
	AUC ↑	KS ↑	LogLoss ↓	AUC ↑	KS ↑	LogLoss ↓
Naive	0.6541	0.2346	0.0650	0.6416	0.2092	0.0697
ESMM	0.6608	0.2428	0.0613	0.6376	0.1996	0.0619
IPS	0.6741	0.2742	0.0635	0.6609	0.2614	0.0638
DR	0.6839	0.2884	0.0628	0.6697	0.2735	0.0634
AKBIPS	0.6847	0.2945	0.0613	0.6718	0.2758	0.0630
AKBDR	0.6883	0.2911	0.0619	0.6723	0.2761	0.0627
DCE-DR	0.6941	0.3022	0.0581	0.6789	0.2692	0.0593
ESCM^2-IPS	0.6892	0.2981	0.0608	0.6767	0.2726	0.0616
ESCM^2-DR	0.6989	0.3131	0.0575	0.6843	0.2983	0.0592
ExtraDebias	**0.7053**	**0.3357**	**0.0569**	**0.6931**	**0.3130**	**0.0550**
Imp. ↑	**+0.92%**	**+7.22%**	**+1.04%**	**+1.29%**	**+4.93%**	**+7.09%**

Table 4: Results of the 5-day online A/B test. "Avg." denotes the average relative improvement over the 5 days.

	Day1	Day2	Day3	Day4	Day5	**Avg.**
CVR	**+1.14%**	**+1.62%**	**+1.27%**	-1.04%	**+1.09%**	**+0.82%**
ROI	**+0.78%**	**+1.29%**	**+1.83%**	-0.47%	**+2.82%**	**+1.25%**

consistently outperforms the baseline, yielding average relative improvements of 0.82% in CVR and 1.25% in ROI.

5 RELATED WORK

5.1 Debiased Recommendation

Selection bias is prevalent in recommendation systems [2, 51, 57], and ignoring such bias leads to poor prediction performance [14, 20, 21]. To address selection bias, the error imputation based (EIB) methods [30, 43] have been proposed to impute missing data and train prediction model using both observed and imputed data. The inverse propensity score (IPS) methods [27, 40, 41] estimate the probability of a sample being observed and apply inverse probability weighting to construct an estimator of the ideal loss for prediction model training. The doubly robust (DR) methods [38, 51] combine error imputation with inverse propensity weighting strategies, achieving doubly robust property. Propensity based methods such as IPS and DR have demonstrated strong performance, leading to extensive follow-up work aimed at improving them.

A major line of research focuses on improving the estimation of propensity scores and error imputation. For example, some studies introduce balancing constraints to achieve more accurate propensity score estimation [18, 19, 48], while Li et al. [15] enhance the error imputation model via targeted learning. Kweon and Yu [12] propose jointly calibrating both the propensity and imputation models to improve estimation accuracy. Moreover, when a small amount of RCT data is available, existing work adopts bi-level optimization to enhance model selection for both propensity and imputation models [1, 52]. Unobserved confounding is another active research area, aiming to address biases caused by insufficiently observed features [58, 67]. Common approaches include worst-case control methods based on sensitivity analysis [6] and model correction techniques that incorporate RCT data [16, 17]. Moreover, in practical applications, data sparsity poses additional challenges for debiasing [35, 64], for which multi-task learning and shared-bottom embedding techniques are employed to alleviate the adverse effects of sparse data [44, 47, 70].

However, existing debiasing approaches rely on the positivity propensity assumption, which requires $p_{u,i} > 0$ for all $(u, i) \in \mathcal{D}$ and may not hold in certain real-world scenarios. This study shows that when zero-propensity samples are present, both IPS and DR estimators become biased, and proposes the ExtraDebias to address this problem. Crucially, the zero-propensity challenge is fundamentally distinct from the high-variance issue caused by extremely small propensity scores. While the latter can be mitigated by variance reduction strategies such as self-normalization [45], variance minimization imputation learning [9], and bias-variance joint optimization [5, 10], these methods are ineffective when the propensity score is exactly zero, as inverse propensity weighting cannot be applied when the denominator is zero.

5.2 Cold-start Recommendation

Classical cold-start recommendation aims to provide better recommendations for new users or items with no historical interaction data [59, 61]. One prominent line of work adopts meta-learning paradigms to enable adaptation to these new entities. By simulating cold-start scenarios during training, these methods extract transferable meta-knowledge or initial parameters from limited support sets, empowering the model to effectively infer preferences for unseen samples [13, 34, 50]. Another line of work leverages generative and perturbation techniques to reconstruct collaborative signals from content features. These approaches typically synthesize virtual ID embeddings or intentionally corrupt input data during training, thereby forcing the model to learn robust representations solely from auxiliary content when interaction records are unavailable [46, 53, 62].

Crucially, the zero-propensity problem is notably distinct from the cold-start problem regarding the target samples. Existing cold-start techniques specifically target newly arriving users or items with empty history. In contrast, our work addresses systematically blocked user-item pairs dictated by platform rules (e.g., geographic restrictions), where both the user and the item may be active and mature with rich histories, yet their specific interaction is forbidden.

5.3 Distribution Robust Optimization

Distributionally Robust Optimization (DRO) assumes that the true data distribution is unknown but lies within a predefined uncertainty set, seeking to minimize the worst-case risk over this set to ensure generalization [33, 36]. Standard approaches typically construct these uncertainty sets using statistical constraints, such as the Wasserstein distance, and solve the resulting min-max problems via Lagrangian relaxation [32, 42]. In recommender systems, DRO has been actively introduced to enhance model robustness against

data heterogeneity and distribution shifts. For instance, Wen et al. [55] formulated the recommendation task as a DRO problem to optimize the worst-case user experience, thereby mitigating the unfairness arising from imbalanced user interactions. Parallelly, Zhao et al. [63] leveraged DRO to construct uncertainty sets that specifically account for popularity shifts, effectively addressing the distribution bias driven by item popularity.

However, the application of DRO in causal debiasing remains scarce, particularly for addressing the zero-propensity challenge where traditional reweighting paradigms are mathematically undefined. To bridge this gap, we reformulate the extrapolation problem as a specialized DRO task. Unlike general DRO frameworks that rely on computationally expensive Lagrangian dual optimization, we derive a closed-form, tight upper bound for the prediction error on zero-propensity extrapolation samples. This enables an exact and efficient optimization of the worst-case risk, free from the computational burden of min-max optimization.

6 CONCLUSION

This study investigates the selection bias mitigation methods in the presence of extrapolation samples, which refer to the user-item pairs with zero probability of interaction (zero propensity). Specifically, we first theoretically show that when extrapolation samples are present, the widely adopted propensity based methods such as IPS and DR estimators are biased. To tackle this problem, we propose ExtraDebias, which, to the best of our knowledge, is the first work in the recommender system field to addresses the zero propensity challenge. The method consists of three main components: extrapolation sample identification based on a pretrained propensity model, imputation model learning on interpolation samples with positive propensity using existing debiasing methods, and extrapolation error control via minimax optimization. The proposed approach is grounded in the Lipschitz continuity assumption, which enables extrapolation from regions with positive propensity to those with zero propensity. We argue and empirically demonstrate that this assumption is mild and practically reasonable. In addition, we provide theoretical analyses regarding the upper bound of extrapolation error and the estimation of the Lipschitz constant. Experimental results on three public datasets, one private industrial dataset, and an online A/B test validate the effectiveness of ExtraDebias.

One limitation is that the L-Lipschitz continuity assumption underlying extrapolation may be violated when embeddings lack sufficient discriminability (e.g., collapsing to raw features) or when the CVR function exhibits abrupt discontinuities due to hard constraints (e.g., a user purchases only if the price falls strictly below a fixed threshold). Developing effective extrapolation methods under such settings remains an important direction for future research.

Acknowledgments

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

References

[1] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. 2021. AutoDebias: Learning to Debias for Recommendation. In *International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 21–30.

[2] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Transactions on Information Systems* 41, 3 (2023).

[3] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.

[4] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *ACM conference on recommender systems*. 191–198.

[5] Quanyu Dai, Haoxuan Li, Peng Wu, Zhenhua Dong, Xiao-Hua Zhou, Rui Zhang, Rui Zhang, and Jie Sun. 2022. A Generalized Doubly Robust Learning Framework for Debiasing Post-Click Conversion Rate Prediction. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 252–262.

[6] Sihao Ding, Peng Wu, Fuli Feng, Yitong Wang, Xiangnan He, Yong Liao, and Yongdong Zhang. 2022. Addressing Unmeasured Confounder for Recommendation with Sensitivity Analysis. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 305–315.

[7] 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 *International Conference on Information and Knowledge Management*. 540–550.

[8] Chen Gao, Tzu-Heng Lin, Nian Li, Depeng Jin, and Yong Li. 2021. Cross-platform item recommendation for online social e-commerce. *IEEE Transactions on Knowledge and Data Engineering* 35, 2 (2021), 1351–1364.

[9] Siyuan Guo, Lixin Zou, Yiding Liu, Wenwen Ye, Suqi Cheng, Shuaiqiang Wang, Hechang Chen, Dawei Yin, and Yi Chang. 2021. Enhanced Doubly Robust Learning for Debiasing Post-Click Conversion Rate Estimation. In *International SIGIR Conference on Research and Development in Information Retrieval*. 275–284.

[10] Mingming Ha, Taoxuewen, Wenfang Lin, QIONGXU MA, Wujiang Xu, and Linxun Chen. 2024. Fine-Grained Dynamic Framework for Bias-Variance Joint Optimization on Data Missing Not at Random. In *Advances in Neural Information Processing Systems*.

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

[12] Wonbin Kweon and Hwanjo Yu. 2024. Doubly Calibrated Estimator for Recommendation on Data Missing Not at Random. In *International World Wide Web Conference*. ACM, 3810–3820.

[13] Hoyeop Lee, Jinbae Im, Seongwon Jang, Hyunsouk Cho, and Sehee Chung. 2019. Melu: Meta-learned user preference estimator for cold-start recommendation. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1073–1082.

[14] 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*. AAAI Press, 4417–4425.

[15] 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*.

[16] 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*.

[17] 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*. 1305–1313.

[18] Haoxuan Li, Yanghao Xiao, Chunyuan Zheng, Peng Wu, and Peng Cui. 2023. Propensity Matters: Measuring and Enhancing Balancing for Recommendation. In *International Conference on Machine Learning*, Vol. 202. 20182–20194.

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

[20] Haoxuan Li, Chunyuan Zheng, Shuyi Wang, Kunhan Wu, Eric Wang, Peng Wu, Zhi Geng, Xu Chen, and Xiao-Hua Zhou. 2024. Relaxing the Accurate Imputation Assumption in Doubly Robust Learning for Debiased Collaborative Filtering. In *International Conference on Machine Learning*.

[21] Haoxuan Li, Chunyuan Zheng, Wenjie Wang, Hao Wang, Fuli Feng, and Xiao-Hua Zhou. 2024. Debiased recommendation with noisy feedback. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1576–1586.

[22] 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*.

[23] 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 *International SIGIR Conference on Research and Development in Information Retrieval*.

[24] 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*.

[25] Dugang Liu, Pengxiang Cheng, Hong Zhu, Zhenhua Dong, Xiuqiang He, Weike Pan, and Zhong Ming. 2021. Mitigating Confounding Bias in Recommendation via Information Bottleneck. In *ACM Conference on Recommender Systems*. 351–360.
[26] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation forest. In *IEEE International Conference on Data Mining*. IEEE, 413–422.
[27] Jinwei Luo, Dugang Liu, Weike Pan, and Zhong Ming. 2021. Unbiased recommendation model based on improved propensity score estimation. *Journal of Computer Applications* (2021).
[28] Yan Lyu, Xiang Li, Tianyu Xia, Xiangnan Feng, Chunyuan Zheng, Haoxuan Li, 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*.
[29] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. 2018. Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. In *International SIGIR Conference on Research and Development in Information Retrieval*. 1137–1140.
[30] Benjamin M Marlin and Richard S Zemel. 2009. Collaborative prediction and ranking with non-random missing data. In *ACM Conference on Recommender Systems*.
[31] Benjamin M. Marlin, Richard S. Zemel, Sam T. Roweis, and Malcolm Slaney. 2012. Collaborative Filtering and the Missing at Random Assumption. *arXiv preprint arXiv:1206.5267* (2012).
[32] Peyman Mohajerin Esfahani and Daniel Kuhn. 2018. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171, 1 (2018), 115–166.
[33] Hongseok Namkoong and John C Duchi. 2016. Stochastic gradient methods for distributionally robust optimization with f-divergences. *Advances in Neural Information Processing Systems* 29 (2016).
[34] Feiyang Pan, Shuokai Li, Xiang Ao, Pingzhong Tang, and Qing He. 2019. Warm up cold-start advertisements: Improving ctr predictions via learning to learn id embeddings. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 695–704.
[35] Hang Pan, Chunyuan Zheng, Wenjie Wang, Jingang Jiang, Xueying Li, Haoxuan Li, and Fuli Feng. 2026. Batch-Adaptive Doubly Robust Learning for Debiasing Post-Click Conversion Rate Prediction Under Sparse Data. *ACM Transactions on Information Systems* 44, 3 (2026), 1–29.
[36] Hamed Rahimian and Sanjay Mehrotra. 2019. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659* (2019).
[37] Yuta Saito. 2020. Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. In *International SIGIR Conference on Research and Development in Information Retrieval*. 309–318.
[38] Yuta Saito. 2020. Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions. In *ACM Conference on Recommender Systems*. 92–100.
[39] Yuta Saito and Masahiro Nomura. 2022. Towards Resolving Propensity Contradiction in Offline Recommender Learning. In *International Joint Conferences on Artificial Intelligence*. 2211–2217.
[40] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *International Conference on Web Search and Data Mining*. 501–509.
[41] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *International Conference on Machine Learning*, Vol. 48. 1670–1679.
[42] Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John Duchi. 2017. Certifying some distributional robustness with principled adversarial training. *arXiv preprint arXiv:1710.10571* (2017).
[43] Harald Steck. 2010. Training and testing of recommender systems on data missing not at random, In ACM SIGKDD Conference on Knowledge Discovery and Data Mining. *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
[44] Hongzu Su, Lichao Meng, Lei Zhu, Ke Lu, and Jingjing Li. 2024. DDPO: Direct Dual Propensity Optimization for Post-Click Conversion Rate Estimation. In *International SIGIR Conference on Research and Development in Information Retrieval*. 1179–1188.
[45] Adith Swaminathan and Thorsten Joachims. 2015. The Self-Normalized Estimator for Counterfactual Learning. In *Advances in Neural Information Processing Systems*. 3231–3239.
[46] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. Dropoutnet: Addressing cold start in recommender systems. In *Advances in Neural Information Processing Systems*.
[47] Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. 2022. ESCM2: Entire Space Counterfactual Multi-Task Model for Post-Click Conversion Rate Estimation. In *International SIGIR Conference on Research and Development in Information Retrieval*. ACM, 363–372.
[48] 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* 43, 6 (2025), 1–24.
[49] Ruoxi Wang, Rakesh Shivanna, Derek Cheng, Sagar Jain, Dong Lin, Lichan Hong, and Ed Chi. 2021. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems. In *International World Wide Web Conference*.
[50] Wenjie Wang, Xinyu Lin, Liuhui Wang, Fuli Feng, Yinwei Wei, and Tat-Seng Chua. 2023. Equivariant Learning for Out-of-Distribution Cold-start Recommendation. In *International Conference on Multimedia*. 903–914.
[51] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *International Conference on Machine Learning*. 6638–6647.
[52] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2021. Combating Selection Biases in Recommender Systems with a Few Unbiased Ratings. In *International Conference on Web Search and Data Mining*. ACM, 427–435.
[53] Yaqing Wang, Hongming Piao, Daxiang Dong, Quanming Yao, and Jingbo Zhou. 2024. Warming Up Cold-Start CTR Prediction by Learning Item-Specific Feature Interactions. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3233–3244.
[54] Zifeng Wang, Xi Chen, Rui Wen, Shao-Lun Huang, Ercan E. Kuruoglu, and Yefeng Zheng. 2020. Information Theoretic Counterfactual Learning from Missing-Not-At-Random Feedback. *Advances in Neural Information Processing Systems* (2020).
[55] Hongyi Wen, Xinyang Yi, Tiansheng Yao, Jiaxi Tang, Lichan Hong, and Ed H Chi. 2022. Distributionally-robust recommendations for improving worst-case user experience. In *International World Wide Web Conference*. 3606–3610.
[56] 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*. 1611–1622.
[57] 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*.
[58] Yanghao Xiao, Haoxuan Li, Yongqiang Tang, and Wensheng Zhang. 2024. Addressing hidden confounding with heterogeneous observational datasets for recommendation. *Advances in Neural Information Processing Systems* 37 (2024), 130358–130383.
[59] Runsheng Yu, Yu Gong, Xu He, Yu Zhu, Qingwen Liu, Wenwu Ou, and Bo An. 2021. Personalized adaptive meta learning for cold-start user preference prediction. In *AAAI Conference on Artificial Intelligence*, Vol. 35. 10772–10780.
[60] Wenhao Zhang, Wentian Bao, Xiao-Yang Liu, Keping Yang, Quan Lin, Hong Wen, and Ramin Ramezani. 2020. Large-scale Causal Approaches to Debiasing Post-click Conversion Rate Estimation with Multi-task Learning. In *International World Wide Web Conference*.
[61] Weizhi Zhang, Yuanchen Bei, Liangwei Yang, Henry Peng Zou, Peilin Zhou, Aiwei Liu, Yinghui Li, Hao Chen, Jianling Wang, Yu Wang, et al. 2025. Cold-start recommendation towards the era of large language models (llms): A comprehensive survey and roadmap. *arXiv preprint arXiv:2501.01945* (2025).
[62] Zeyang Zhang, Xin Wang, Haibo Chen, Haoyang Li, and Wenwu Zhu. 2024. Disentangled Dynamic Graph Attention Network for Out-of-Distribution Sequential Recommendation. *ACM Transactions on Information Systems* 43, 1 (2024), 1–42.
[63] Jujia Zhao, Wenjie Wang, Xinyu Lin, Leigang Qu, Jizhi Zhang, and Tat-Seng Chua. 2023. Popularity-aware distributionally robust optimization for recommendation system. In *International Conference on Information and Knowledge Management*. 4967–4973.
[64] Chunyuan Zheng, Hang Pan, Yang Zhang, and Haoxuan Li. 2025. Adaptive structure learning with partial parameter sharing for post-click conversion rate prediction. In *International SIGIR Conference on Research and Development in Information Retrieval*. 233–243.
[65] 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*, Vol. 40. 16468–16476.
[66] 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*.
[67] 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*. 2113–2123.
[68] Chuan Zhou, Lina Yao, Haoxuan Li, and Mingming Gong. 2025. Counterfactual implicit feedback modeling. In *Advances in Neural Information Processing Systems*.
[69] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
[70] Feng Zhu, Mingjie Zhong, Xinxing Yang, Longfei Li, Lu Yu, Tiehua Zhang, Jun Zhou, Chaochao Chen, Fei Wu, Guanfeng Liu, and Yan Wang. 2023. DCMT: A Direct Entire-Space Causal Multi-Task Framework for Post-Click Conversion Estimation. In *International Conference on Data Engineering*. 3113–3125.