Entire Space Counterfactual Learning for Reliable Content Recommendations

Hao Wang, Zhichao Chen, Zhaoran Liu, Haozhe Li, Degui Yang, Xinggao Liu, and Haoxuan Li, *Member, IEEE*

***Abstract*—Post-click conversion rate (CVR) estimation is a fundamental task in developing effective recommender systems, yet it faces challenges from data sparsity and sample selection bias. To handle both challenges, the entire space multitask models are employed to decompose the user behavior track into a sequence of exposure → click → conversion, constructing surrogate learning tasks for CVR estimation. However, these methods suffer from two significant defects: (1) intrinsic estimation bias (IEB), where the CVR estimates are higher than the actual values; (2) false independence prior (FIP), where the causal relationship between clicks and subsequent conversions is potentially overlooked. To overcome these limitations, we develop a model-agnostic framework, namely Entire Space Counterfactual Multitask Model (ESCM^2), which incorporates a counterfactual risk minimizer within the entire space multitask framework to regularize CVR estimation. Experiments conducted on large-scale industrial recommendation datasets and an online industrial recommendation service demonstrate that ESCM^2 effectively mitigates IEB and FIP defects and substantially enhances recommendation performance.**

***Index Terms*—Debiased recommendation, multitask learning, conversion rate estimation.**

I. Introduction

RECOMMENDATION systems play an essential role in customizing content delivery across various industries such as e-commerce [1], advertising [2], and social media [3], serving as a cornerstone in information management and dissemination [4], [5]. They typically operate through a two-phase pipeline, in Fig. 1, involving an offline phase and an online phase. In the offline phase, user profiles, item attributes, and user-item interactions are extracted from logs to train a ranking model. In the online phase, this model ranks candidate items based on criteria of interest such as click-through rate (CTR), post-click conversion rate (CVR), and click-through and conversion rate (CTCVR), and exposes top items to users to meet their preferences. The model is continuously refined based on real-time user feedback [1].

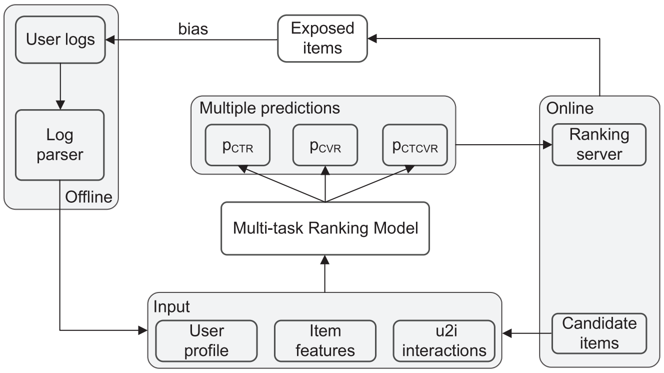


Fig. 1. Overview of a two-stage industrial recommender, which involves the offline and online phases.

User behavior in recommendation systems often follows a trajectory of exposure → click → conversion [1]. In this context, CTR, CVR, and CTCVR respectively quantify the transition probabilities from exposure to click, click to conversion, and exposure to conversion. Advances in feature interaction and deep learning have established CTR estimation as a good practice. However, click-through feedback is often interfered by unexpected factors such as clickbait, which distorts genuine user preferences [6]. In contrast, CVR, which indicates post-click user behavior, offers a more accurate reflection of user preferences, attracting focused attention within the recommendation community [7], [8], [9].

A naïve yet common approach [10] to obtaining CVR estimators is to train models solely with samples where click happens. Where conversion labels are fully observable [10]. This naïve approach introduces two major problems: sample selection bias and data sparsity. Sample selection bias arises because the training dataset consists only of clicked samples, whereas the inference needs to consider all exposures [7]. Since samples with lower CVR are more likely to be excluded from the click space [7], [11], training data is missing not at random [10], which leads to a distribution shift between the training and inference spaces [12]. Data sparsity arises due to low click-through rates (e.g., 4% in the

Received 17 April 2024; revised 26 August 2024 and 7 October 2024; accepted 13 November 2024. Date of publication 12 December 2024; date of current version 12 February 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 623B2002, Grant 62073288, Grant 12075212, and Grant 12105246. The associate editor coordinating the review of this article and approving it for publication was Prof. Edgar Weippl. *(Corresponding author: Haoxuan Li.)*

Hao Wang, Zhichao Chen, Zhaoran Liu, Haozhe Li, and Xinggao Liu are with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: Ho-ward@outlook.com; 12032042@zju.edu.cn; 22032057@zju.edu.cn; lihaozhe@zju.edu.cn; lxg@zju.edu.cn).

Degui Yang is with the School of Automation, Central South University, Changsha 410000, China (e-mail: degui.yang@csu.edu.cn).

Haoxuan Li is with the Center for Data Science, Peking University, Beijing 100871, China (e-mail: hxli@stu.pku.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIFS.2024.3516584, provided by the authors.

Digital Object Identifier 10.1109/TIFS.2024.3516584

Ali-CPP dataset and 3.8% in our industrial dataset), which restricts the data availability for training CVR estimators [1]. These two problems undermine the reliability of recommendation systems by hampering their generalization to unseen samples.

To handle the challenge of data sparsity and sample selection bias, the entire space multitask model (ESMM) [1] avoids training CVR estimator directly by employing a multitask approach that simultaneously optimizes CTCVR and CTR objectives [1]. Since training of both CTCVR and CTR objectives can utilize all exposure samples, this strategy alleviates data sparsity and enhances performance in practice [13]. However, its reliability in CVR estimation has been questioned due to the lack of unbiasedness guarantee [7] and overly simplistic dependency assumptions. In this study, we formalize these concerns as two defects of ESMM:

- **Intrinsic Estimation Bias (IEB)**: the CVR estimates are biased from true values.
- **False Independence Prior (FIP)**: the CTR and CVR estimates are susceptible to inappropriate assumptions of conditional independence, ignoring the causal relationship from click to conversion.

To address these defects, we propose the Entire Space Counterfactual Multitask Model (ESCM^2), a model-agnostic framework that incorporates counterfactual regularizers for CVR estimation. Both theoretical and empirical evaluations demonstrate that our regularizers effectively mitigate the IEB and FIP defects. Our main contributions are as follows:

- We identify IEB and FIP as critical defects in ESMM, with empirical results and theoretical analysis.
- We develop ESCM^2, integrating counterfactual regularizers within ESMM to enhance performance. We provide theoretical justifications and empirical validations demonstrating its effectiveness in mitigating IEB and FIP.
- We conduct extensive evaluations using industrial recommendation datasets to validate the efficacy of ESCM^2, we implement ESCM^2 on our online recommendation platform, where it achieves substantial profit increases.[1]

The remaining sections are structured as follows: Section II provides preliminaries for understanding the technical details in this work; Section III formulates the IEB and FIP defects with ESMM; Section IV introduces the ESCM^2 framework for training recommendation models, which enhances ESMM by incorporating the proposed counterfactual regularizers to handle the IEB and FIP defects; Section V presents real-world case studies to demonstrate the efficacy of ESCM^2; Section VI provides a brief overview of related works; Section VII summarizes the conclusions and outlines open questions.

II. Preliminaries

A. Notations

In this paper, uppercase letters, *e.g.,* O, represent random variables; lowercase letters, *e.g.,* o, represent the associated specific values; calligraphic letters such as $\mathcal{O}$ denote sample spaces; $\mathbb{P}(\cdot)$, $\mathbb{E}(\cdot)$, $\mathbb{V}(\cdot)$ represent probability distribution, expectation and variance, respectively.

B. Problem Statement

Denote $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$ and $\mathcal{I} = \{i_1, i_2, \dots, i_n\}$ as the respective sets of users and items in the exposure space. Let $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ be the set of user-item intersections in the exposure space. Let $\mathbf{O} \in \{0,1\}^{m\times n}$ be the click indicators where $o_{u,i} \in \{0,1\}$ indicates whether the user u clicks the item i; $\mathbf{R} \in \{0,1\}^{m\times n}$ be the conversion labels where $r_{u,i} \in \{0,1\}$ indicates whether the user u purchases the item i.

If all entries $r_{u,i} \in \mathbf{R}$ are observable, the ideal learning objective for constructing CVR estimator is expressed as

$$\mathcal{P} := \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\delta\left(r_{u,i}, \hat{r}_{u,i}\right)\right], \tag{1}$$

where $\hat{r}_{u,i}$ denotes the estimate of $r_{u,i}$, δ measures the estimation error and can be specified as any classification loss function, $\delta(r_{u,i}, \hat{r}_{u,i})$ is the estimation error of CVR for a specific user and item. Following existing works [1], [7], we utilize binary cross-entropy as the loss measure:

$$\begin{aligned}\varepsilon_{u,i} :\ &= \delta\left(r_{u,i}, \hat{r}_{u,i}.\right)\\ &= -r_{u,i}\log\hat{r}_{u,i} - (1 - r_{u,i})\log(1 - \hat{r}_{u,i}),\end{aligned} \tag{2}$$

where we abbreviate the CVR estimation error $\delta(r_{u,i}, \hat{r}_{u,i})$ as $\varepsilon_{u,i}$. However, the ideal objective (1) is incomputable since $r_{u,i}$ is unobservable for samples outside the click space $\mathcal{O}$. A naive yet common shortcut is to estimate the learning objective using clicked samples in $\mathcal{O}$:

$$\mathcal{L}_{\text{naive}} := \mathbb{E}_{(u,i)\in\mathcal{O}}(\varepsilon_{u,i}) = \frac{1}{|\mathcal{O}|}\sum_{(u,i)\in\mathcal{D}}(o_{u,i}\varepsilon_{u,i}), \tag{3}$$

where $|\mathcal{O}| = \sum_{(u,i)\in\mathcal{D}}(o_{u,i})$. Nonetheless, it has been shown that (3) is a biased estimation of the ideal objective [10], [14] due to selection bias, *i.e.,*$\mathbb{E}_O[\mathcal{L}_{\text{naive}}] \neq \mathcal{P}$.

C. Entire Space Multitask Model Approach

The entire space multitask model (ESMM) [1] is prevalent in recommendation scenarios where CVR estimation plays critical roles. To bypass data sparsity and sample selection bias in training CVR estimators, ESMM avoids direct training of the CVR estimator. Specifically, according to the sequential user behavior track in Fig. 2, CVR can be represented as the quotient of CTCVR and CTR:

$$\underbrace{\mathbb{P}(r_{u,i} = 1 \mid o_{u,i} = 1)}_{\text{CVR}} = \frac{\overbrace{\mathbb{P}(r_{u,i} = 1, o_{u,i} = 1)}^{\text{CTCVR}}}{\underbrace{\mathbb{P}(o_{u,i} = 1)}_{\text{CTR}}}.$$

On this basis, ESMM constructs two predictive arms to estimate respective CTR and CVR as $\hat{o}_{u,i}$ and $\hat{r}_{u,i}$, and multiplies

[1] Building on our conference work [8], we detail the statistical properties, showing that both IPS and DR regularizers effectively handle IEB and FIP. We also demonstrate that IPS is a specialized importance sampling method.

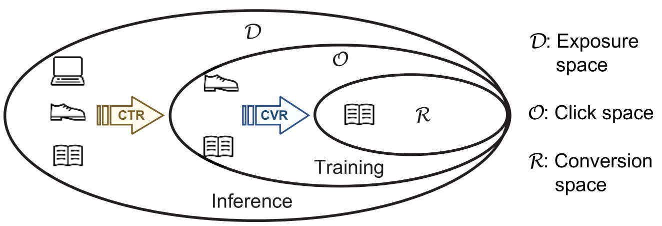

$\mathcal{D}$
$\mathcal{O}$
CTR
CVR
$\mathcal{R}$
Training
Inference
$\mathcal{D}$: Exposure space
$\mathcal{O}$: Click space
$\mathcal{R}$: Conversion space

Fig. 2. Overview of the CVR estimation task in recommendation systems, wherein online inference is executed on all exposed samples, while training is exclusively carried out on clicked samples, leading to data sparsity and sample selection bias.

them to acquire the estimate of CTCVR. During training, ESMM minimizes the empirical risk for CTR and CTCVR estimations as follows:

$$\begin{aligned}\mathcal{L}_{\text{CTR}} &:= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\delta\left(o_{u,i}, \hat{o}_{u,i}\right)\right]\\ \mathcal{L}_{\text{CTCVR}} &:= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\delta\left(o_{u,i} * r_{u,i}, \hat{o}_{u,i} * \hat{r}_{u,i}\right)\right].\end{aligned} \tag{4}$$

In the inference phase, the output from the CVR arm provides the CVR estimates. Since both CTCVR and CTR objectives can utilize all exposure samples, ESMM alleviates data sparsity and enhances performance in practice [13]. Moreover, ESMM seemingly circumvents sample selection bias by avoiding direct training of the CVR estimator. However, as elaborated in Section III-A, it remains vulnerable to selection bias, formalized as the IEB defect in this work. To adequately handle the sample selection bias, a regularization term tailored for CVR estimation seems imperative, aiming to estimate the unbiased learning objective of CVR estimator $\mathcal{P}$ with biased click data (where conversion labels are available).

D. Causal Recommendation Approach

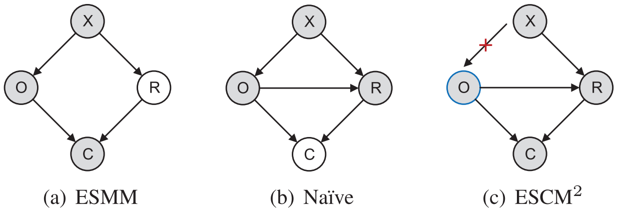
To estimate the ideal learning objective of CVR estimator, causal inference techniques have received special attention by the recommendation community [15], [16]. The core of these methods is to weight the samples in the click space to approximate the sample distribution in the exposure space. A prominent technique is the inverse propensity score (IPS) [10], which weights CVR estimation errors $\varepsilon_{u,i}$ of clicked samples using the inverse propensity score:

$$\mathcal{L}_{\text{IPS}} := \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}\varepsilon_{u,i}}{q_{u,i}}\right], \tag{5}$$

where the propensity score is specified as CTR: $q_{u,i} = \mathbb{P}(o_{u,i} = 1)$. Notably, $\mathcal{L}_{\text{IPS}}$ is an unbiased estimator of the ideal learning objective in (1), *i.e.*,$\mathbb{E}_O(\mathcal{L}_{\text{IPS}}) = \mathcal{P}$, when the propensity estimate is accurate.

However, the IPS method can exhibit high variance, particularly in sparse data scenarios like CVR estimation, where propensities are often extremely small. To address this, the doubly robust (DR) estimator [14] incorporates an error imputation technique. It constructs an imputation model $\hat{\varepsilon}_{u,i}$ to approximate CVR estimation errors in $\mathcal{D}$, and refines the imputation with $\hat{e}_{u,i} = \varepsilon_{u,i} - \hat{\varepsilon}_{u,i}$ in $\mathcal{O}$:

$$\mathcal{L}_{\text{DR}} := \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{\varepsilon}_{u,i} + \frac{o_{u,i}\hat{e}_{u,i}}{q_{u,i}}\right] \tag{6}$$



(a) ESMM (b) Naïve (c) ESCM^2

Fig. 3. Causal graphs where X, R, O, C denote the user-item intersection, conversion, click and click & conversion, respectively. Hollow and shaded nodes indicate latent and observed variables, respectively. The blue circle in (c) represents intervention, blocking the backdoor path X → O.

This formulation ensures unbiasedness as long as either the imputed error $\hat{\varepsilon}_{u,i}$ or the CTR estimate $\hat{o}_{u,i}$ is accurate, hence the term *double robustness*. The IPS and DR estimators offer an unbiased estimation of $\mathcal{P}$ with click data, free from the selection bias introduced by the non-randomness of user clicks.

III. Analysis of Entire Space Multitask Model

A. Intrinsic Estimation Bias

In this section, we delve into the IEB defect with ESMM, where the average CVR estimates exceed the actual values. Previous studies have empirically demonstrated ESMM's susceptibility to this bias [7]. Nonetheless, a formal justification of this defect has not yet been established. We define this problem formally as the IEB problem and establish its presence under mild assumptions in Theorem 1.

Theorem 1 (Existence of IEB): Suppose O, R, and C are the random variables for click, post-click conversion, and click & conversion respectively. For a specific user-item pair (u, i), let $o_{u,i}$, $r_{u,i}$, and $c_{u,i}$ denote the actual values; $\hat{o}_{u,i}$, $\hat{r}_{u,i}$, and $\hat{c}_{u,i}$ denote the estimated values. The expectation of ESMM's CVR estimates across all exposures exceeds the true CVR:

$$\text{Bias}^{\text{ESMM}} := \mathbb{E}_{\mathcal{D}}\left[\hat{R}\right] - \mathbb{E}_{\mathcal{D}}\left[R\right] > 0, \tag{7}$$

under the assumption that conversion is more likely to take place for samples within the click space [10]:

$$\mathbb{E}_{\mathcal{O}}\left[R\right] > \mathbb{E}_{\mathcal{D}}\left[R\right].$$

The existence of IEB highlights that sample selection bias cannot be effectively addressed solely by decomposing tasks within the ESMM framework. Instead, it necessitates developing an unbiased estimation approach for the ideal CVR learning objective and directly optimizing it.

B. False Independence Prior

To estimate CTCVR, ESMM multiplies the outputs of its CTR and CVR arms:

$$\mathbb{P}(o_{u,i} = 1, r_{u,i} = 1) = \mathbb{P}(o_{u,i} = 1) * \mathbb{P}(r_{u,i} = 1 \mid o_{u,i} = 1),$$

where the CVR estimate is click-dependent, *i.e.,* conversion only occurs after the click, establishing a causal link $O \rightarrow R$ in the data generation process. However, ESMM's learning objective (4) does not explicitly capture this causal dependency, as indicated by the absent arrow $O \rightarrow R$ in Fig. 3 (a). It poses the risk that ESMM models CVR as

$\mathbb{P}(r_{u,i} = 1)$ following Fig. 3 (a), as opposed to the expected $\mathbb{P}(r_{u,i} = 1 \mid o_{u,i} = 1)$ in (III-B). This risk is formulated as false independence prior, as it confuses the targeted $\mathbb{P}(r_{u,i} = 1 \mid o_{u,i} = 1)$ in (III-B) with the unexpected $\mathbb{P}(r_{u,i} = 1)$, thereby falsely introducing independent prior in CTCVR estimation.

The naïve approach in (3) trains the CVR model within the click space, thereby explicitly incorporating the dependency $O \rightarrow R$ as depicted in Fig. 3 (b).[2] However, the backdoor path $X \rightarrow O$ introduces sample selection bias [1], [7]. From a causality perspective, the key to solving FIP without introducing backdoor path is defining CVR as a causal estimand:

$$\mathbb{P}(r_{u,i} = 1 \mid do(o_{u,i} = 1)), \tag{8}$$

where "do" represents the do-calculus [19], truncating the backdoor path $X \rightarrow O$ as shown in Fig. 3. For clicked samples, the causal estimand (8) aligns with the standard CVR definition, but for unclicked samples, it models the counterfactual problem: *What would be the likelihood of conversion if the user had clicked the item?*. Based on this formulation, the CTCVR can be redefined as:

$$\mathbb{P}\left(o_{u,i}=1, r_{u,i}=1\right)=\mathbb{P}\left(o_{u,i}=1\right)*\mathbb{P}\left(r_{u,i}=1 \mid do\left(o_{u,i}=1\right)\right),$$

which addresses both FIP and selection bias defects effectively.

IV. Methodology

In this section, we propose ESCM^2 to tackle the aforementioned IEB and FIP defects with ESMM. Section IV-A describes the implementations and properties of the proposed counterfactual regularizers; Section IV-B further demonstrates how the proposed regularizers effectively handle the IEB and FIP defects. Section IV-C develops ESCM^2 by enhancing the ESMM framework with the counterfactual regularizers, detailing the model architecture and learning objectives.

A. Counterfactual Risk Regularizers

In this section, we contextualize the implementation of two counterfactual risk regularizers: the IPS regularizer and the DR regularizer, elucidating their statistical properties. Given that post-click conversion labels are unavailable for non-clicked samples, a naïve approach involves calculating the CVR learning objective based on the estimation errors $\varepsilon_{u,i}$ from clicked samples. However, this method is prone to sample selection bias, which causes a distribution shift between the training space (click space) and the inference space (exposure space). This shift hinders the CVR estimator's ability to generalize from training to inference, leading to suboptimal performance.

To counteract the distribution shift caused by sample selection bias, the IPS regularizer, as per (5), inversely weights each clicked sample (with $o_{u,i} = 1$) with propensity score $q_{u,i}$:

$$\mathcal{R}_{\mathrm{IPS}} = \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}\varepsilon_{u,i}}{q_{u,i}}\right] = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{o_{u,i}\varepsilon_{u,i}}{\hat{o}_{u,i}}, \tag{9}$$

where the propensity score, typically the actual CTR, is unavailable; hence, the CTR estimate $\hat{o}_{u,i}$ is employed as a proxy [7]. This re-weighting strategy corrects for the overrepresentation of data that are more prone to be clicked, aligning the training data more closely with the exposure data, thereby addressing the distribution shift between training and inference space. In practice, it offers an approximation of the ideal CVR learning objective—i.e., the expected value of $\varepsilon_{u,i}$ over the dataset $\mathcal{D}$—using data from biased clicked samples.

The statistical properties of $\mathcal{R}_{\mathrm{IPS}}$ are encapsulated in Lemma 2. Specifically, given accurate CTR estimate (*i.e.,*$\hat{o}_{u,i} = q_{u,i}$), $\mathcal{R}_{\mathrm{IPS}}$ is an unbiased estimator (*i.e.,*$\mathbb{E}_O(\mathcal{R}_{\mathrm{IPS}}) = \mathcal{P}$). However, $\mathcal{R}_{\mathrm{IPS}}$ exhibits high variance when $\hat{o}_{u,i}$ values are small, which makes the training process unstable.

Lemma 2: The bias and variance of $\mathcal{R}_{\mathrm{IPS}}$ are

$$\mathrm{Bias}_{\mathrm{O}}\left(\mathcal{R}_{\mathrm{IPS}}\right) = \frac{1}{|\mathcal{D}|}\left|\sum_{(u,i)\in D}\varepsilon_{u,i}\left(\frac{q_{u,i}}{\hat{o}_{u,i}}-1\right)\right|,$$

$$\mathbb{V}_O\left(\mathcal{R}_{\mathrm{IPS}}\right) = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{q_{u,i}\left(1-q_{u,i}\right)}{\hat{o}^2_{u,i}}\left(\varepsilon_{u,i}\right)^2.$$

To mitigate the defect with IPS regularizer, the DR regularizer extends $\mathcal{R}_{\mathrm{IPS}}$ by incorporating an *imputation arm*. This arm aims to accurately impute the CVR estimation error ($\varepsilon_{u,i}$), and its output, denoted as $\hat{\varepsilon}_{u,i}$, is subsequently corrected by $\hat{e}_{u,i} = \varepsilon_{u,i} - \hat{\varepsilon}_{u,i}$. The imputation is performed in the exposure space, whereas the correction is executed in the click space where the actual $\varepsilon_{u,i}$ values are available. We implement the DR regularizer as follows:

$$\begin{aligned}\mathcal{R}^{\mathrm{err}}_{\mathrm{DR}} &= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{\varepsilon}_{u,i} + \frac{o_{u,i}\hat{e}_{u,i}}{q_{u,i}}\right]\\ &= \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\hat{\varepsilon}_{u,i} + \frac{o_{u,i}\hat{e}_{u,i}}{\hat{o}_{u,i}},\end{aligned} \tag{10}$$

where $\hat{e}_{u,i}$ is weighted with the propensity score $q_{u,i}$ to counteract sample selection bias, $\hat{o}_{u,i}$ is a proxy for the propensity score $q_{u,i}$. This strategy imputes CVR estimation errors for samples outside the click space and corrects this imputation with $\hat{e}_{u,i}$. Although $\hat{e}_{u,i}$ is only available in the click space, the re-weighting corrects for the sample selection bias and offers an equivalent expectation over the exposure space.

Similar to $\mathcal{R}_{\mathrm{IPS}}$, $\mathcal{R}^{\mathrm{err}}_{\mathrm{DR}}$ is unbiased to $\mathcal{P}$. Moreover, according to Lemma 3, $\mathcal{R}_{\mathrm{DR}}$ exhibits lower variance than $\mathcal{R}_{\mathrm{IPS}}$ when $0 < \hat{\varepsilon}_{u,i} < 2\varepsilon_{u,i}$. Besides, $\mathcal{R}_{\mathrm{DR}}$ is doubly robust since it ensures unbiasedness if either the propensity estimation or the error imputation is accurate. The accuracy of $\hat{o}_{u,i}$ can be guaranteed by an arbitrary well-trained CTR estimator, and the accuracy of $\hat{\varepsilon}_{u,i}$ can be assured by an auxiliary learning task:

$$\mathcal{R}^{\mathrm{imp}}_{\mathrm{DR}} = \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}\hat{e}^2_{u,i}}{\hat{o}_{u,i}}\right], \tag{11}$$

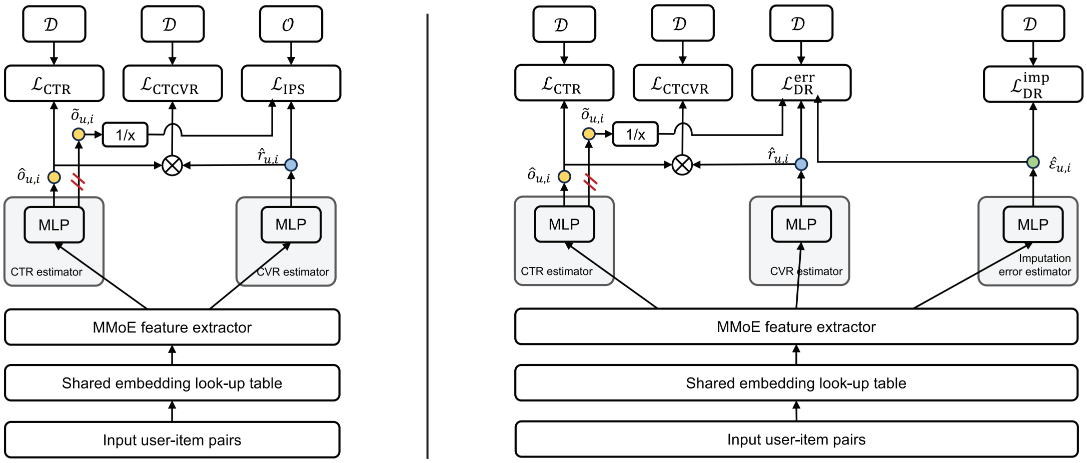
and the final learning objective of the DR regularizer is

$$\mathcal{R}_{\mathrm{DR}} = \mathcal{R}^{\mathrm{err}}_{\mathrm{DR}} + \mathcal{R}^{\mathrm{imp}}_{\mathrm{DR}}. \tag{12}$$

Lemma 3: The bias and variance of $\mathcal{R}^{\mathrm{err}}_{\mathrm{DR}}$ are

$$\mathrm{Bias}\left(\mathcal{R}^{\mathrm{err}}_{\mathrm{DR}}\right) = \frac{1}{|\mathcal{D}|}\left|\sum_{(u,i)\in D}\left(q_{u,i}-\hat{o}_{u,i}\right)\frac{\left(\varepsilon_{u,i}-\hat{\varepsilon}_{u,i}\right)}{\hat{o}_{u,i}}\right|,$$

[2] This causal graph aligns with Fig. 1 in [17] and [18].



(a) ESCM^2-IPS (b) ESCM^2-DR

Fig. 4. The core architecture of ESCM^2, where the ESCM^2-IPS involves two arms for CTR and CVR estimation; the ESCM^2-DR involves an additional arm for imputation error estimation. The breaker indicates the path where the gradients are truncated. $\mathcal{D}$ and $\mathcal{O}$ denote respective exposure and click space.

$$\mathbb{V}_{\mathcal{O}}\left(\mathcal{R}^{\text{err}}_{\text{DR}}\right) = \frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}} q_{u,i}\left(1-q_{u,i}\right)\frac{\left(\hat{\varepsilon}_{u,i}-\varepsilon_{u,i}\right)^2}{\hat{o}^2_{u,i}}.$$

B. Analytical Properties

In this section, we demonstrate that the IPS regularizer ($\mathcal{R}_{\text{IPS}}$) can effectively handle the IEB and FIP defects with ESMM. Specifically, Theorem 4 establishes that $\mathcal{R}_{\text{IPS}}$ aligns with the ideal learning objective $\mathcal{P}$ in (1), thereby handling IEB. Concurrently, Theorem 5 establishes that $\mathcal{R}_{\text{IPS}}$ promotes the estimation of the CVR as $\mathbb{P}(r_{u,i} = 1 \mid do(o_{u,i} = 1))$, as specified in (8), which explicitly models the causal link from click to conversion, thereby mitigating FIP. These theoretical results are also applicable to the DR regularizer $\mathcal{R}_{\text{DR}}$, with detailed discussions provided in Theorems 7–8 in the Supplementary material.

Theorem 4 ($\mathcal{R}_{\text{IPS}}$ handles IEB): Given accurate propensity score estimation, i.e.,$\hat{o}_{u,i} = q_{u,i}$, we have $\mathcal{R}_{\text{IPS}} = \mathcal{P}$.

Theorem 5 ($\mathcal{R}_{\text{IPS}}$ handles PIP): Suppose $\hat{r}^{\text{IPS}}_{u,i}$ is the CVR estimate that optimizes $\mathcal{R}_{\text{IPS}}$, $\mathbb{P}(r_{u,i} = 1 \mid do(o_{u,i} = 1))$ is the counterfactual conversion rate assuming the user clicked the item. For all samples in the exposure space, $\mathcal{R}_{\text{IPS}}$ encourages:

$$\hat{r}^{\text{IPS}}_{u,i} \rightarrow \mathbb{P}\left(r_{u,i} = 1 \mid do\left(o_{u,i} = 1\right)\right).$$

Corollary 6: $\mathcal{R}_{\text{IPS}}$ is typically an importance sampling, which computes the ideal expectation over the exposure space $\mathcal{D}$ using samples from the click space $\mathcal{O}$.

While Theorem 2 initially identified the unbiasedness of $\mathcal{R}_{\text{IPS}}$, Theorem 4 presents a stronger unbiasedness, emphasizing a strong alignment with importance sampling principles [20]. Building on this, Corollary 6 posits that $\mathcal{R}_{\text{IPS}}$ typically operates as a typical *importance sampling*. This interpretation suggests that the IPS estimator can be enhanced using advanced importance sampling techniques [20], [21], [22], for superior statistical properties.

C. Architecture and Learning Objective

While the proposed counterfactual regularizers effectively approximate the ideal learning objective, they do not yield a deployable recommendation model. To bridge this gap, we introduce ESCM^2, which employs the counterfactual risk regularizers for training recommendation models. The detailed steps are outlined in Algorithm 1 and are explained as follows.

First, we construct a multi-arm estimator f to estimate the CTR, CVR and imputation error (step 1). The architecture of f in Fig. 4 involves an embedding lookup table, a feature extractor, and task-specific prediction arms. To mitigate data sparsity, the embedding lookup table is shared across different tasks, and f is implemented using a MMoE model [23].

Subsequently, we compute the CVR loss $\mathcal{L}_{\text{CVR}}$ using the counterfactual regularizers (steps 2–6). When the IPS regularizer is employed, $\mathcal{L}_{\text{CVR}}$ corresponds to $\mathcal{R}_{\text{IPS}}$ as defined in (9). Crucially, we truncate the gradient of $\mathcal{R}_{\text{IPS}}$ with respect to $\hat{o}_{u,i}$ (step 2) because the CTR estimate acts only as a coefficient in this context; optimizing it alongside the CVR objective would degrade CTR estimation performance. Alternatively, if the DR regularizer is used, $\mathcal{L}_{\text{CVR}}$ corresponds to $\mathcal{R}_{\text{DR}}$ as defined in (12). For training stability, we stop the gradient flow from $\hat{\varepsilon}_{u,i}$ when calibrating it with true errors and from $\varepsilon_{u,i}$ when optimizing its estimation.

Finally, we define the learning objective of ESCM^2, which comprises three terms (steps 7-8):

$$\mathcal{L}_{\text{ESCM}^2} := \mathcal{L}_{\text{CTR}} + \lambda_{\text{c}}\mathcal{L}_{\text{CVR}} + \lambda_{\text{g}}\mathcal{L}_{\text{CTCVR}}, \tag{13}$$

where $\mathcal{L}_{\text{CTR}}$ and $\mathcal{L}_{\text{CTCVR}}$ represent the CTR and CTCVR estimation risks defined in (4); $\mathcal{L}_{\text{CVR}}$ is the counterfactual CVR risk; λ_{c} and λ_{g} are weighting factors. This configuration enables ESCM^2 to leverage the ESMM structure to counteract data sparsity while effectively handling the IEB and FIP defects with ESMM using counterfactual regularizers.

V. Experiments

In this section, we conduct experiments to investigate the research questions as follows:

RQ1: How does ESCM^2 perform compared to the prevalent CVR and CTCVR estimators in offline and online scenarios?

RQ2: Does ESMM suffer from the intrinsic estimation bias on CVR estimation? Does ESCM^2 effectively reduce the bias?

Algorithm 1 The Computational Procedure for ESCM^2

Input: $(u, i) \in \mathcal{D}$: the user-item pairs in the exposure space; $o_{u,i}$: the click label in the exposure space; $r_{u,i}$: the conversion label in the click space.

Parameter: λ_c: the weight of the counterfactual risk; λ_g: the weight of the global risk.

Output: $\mathcal{L}_{\text{ESCM}^2}$: the learning objective of ESCM^2.

```
$\hat{o}_{u,i}, \hat{r}_{u,i}, \hat{\varepsilon}_{u,i} \leftarrow f(u, i)$.
$\tilde{o}_{u,i} \leftarrow \text{StopGradient}(\hat{o}_{u,i})$.
if model is ESCM$^2$-IPS then
    Calculate $\mathcal{L}_{\text{CVR}}$ as $\mathcal{R}_{\text{IPS}}$ in Eq.(9).
else if model is ESCM$^2$-DR then
    Calculate $\mathcal{L}_{\text{CVR}}$ as $\mathcal{R}_{\text{DR}}^{\text{err}} + \mathcal{R}_{\text{DR}}^{\text{imp}}$ in Eq.(12).
end if
Calculate $\mathcal{L}_{\text{CTR}}$ and $\mathcal{L}_{\text{CTCVR}}$ in Eq.(4).
Calculate $\mathcal{L}_{\text{ESCM}^2}$ in Eq.(13).
```

TABLE I
DATASET DESCRIPTION

Name	# Train	# Valid	# Test	# User	# Click	# Conversion
Industry	61.58M	0.39M	24.28M	37.73M	3.73M	0.32M
Ali-CCP	33.12M	3.67M	37.64M	0.25M	1.42M	7.92K

RQ3: Does ESMM suffer from false independence prior in CTCVR estimation? Does ESCM^2 mitigate this problem?

RQ4: How to tune the weights of learning objectives? Is the performance of ESCM^2 sensitive to it?

A. Setup

1) Dataset: Experiments are conducted using two datasets, in Table I. The **Industry** dataset is constructed using our industrial recommendation logs over 90 days, segmented chronologically into training, validation, and test sets. Negative samples are downsampled in the training phase to maintain an approximate exposure:click:conversion ratio of 100:10:1. The **Ali-CCP** dataset is incorporated for reproducibility.[3] Only single-valued categorical fields are used following Xi et al. [24], and 10% of the training set is reserved for validation.

2) Baselines: Given that Multi-Task Learning (MTL) significantly enhances the performance of recommender systems [7], single-task CVR estimation approaches [10], [14] are excluded from our baselines to provide a fair comparison. We commence with three prevalent methods that co-train CTR and CVR estimators and share embeddings between them:

- **Naïve**[4] [23] optimizes the CTR estimator in the exposure space and the CVR estimator in the click space using the biased approach described in (3).
- **MTL-IMP** [1] extends **Naïve** by including unclicked samples as negative samples to train the CVR estimator.
- **ESMM**[4] [1] employs a multitask approach to optimize two independent learning objectives for CTCVR and CTR [1], and implicitly optimizes the CVR estimator.

Moreover, we incorporate debiased methods as follows:

- **MTL-EIB** [25] imputes the CVR estimation error for all samples and corrects its imputation with clicked samples to achieve a theoretically unbiased CVR estimation.
- **MTL-IPS**[5] and **MTL-DR** [7] integrates the IPS and DR [10] into a multitask learning framework, respectively, providing a theoretically unbiased CVR estimator.

3) Training Protocol: For all methods in comparison, the multitask estimator is implemented as a standard MMoE model [23], beginning with a shared embedding layer. The embedding dimension is uniformly set to 5, with other model settings consistent with standard MMoE. The learning rate and weight decay are set to $1e^{-4}$ and $1e^{-3}$, respectively. Other optimizer settings are consistent with Adam optimizer [26]. Notably, due to the one-epoch saturation phenomenon observed in industrial recommenders [27], where model performance tends to degrade after more than one epoch of training, each model is trained for a single epoch with batch size 512. The weighting factors, λ_g and λ_c, are determined based on the outcomes of a hyperparameter study presented in Section V-F, with values set to 1 and 0.1. All experiments are conducted on K8S clusters in Ant Group with Intel Xeon Platinum CPUs.

4) Evaluation Protocol: We mainly use the area under the receiver operating characteristic curve (AUC) metric to assess the ranking performance of the models. While AUC is a robust measure for evaluating the average ranking performance across all possible thresholds, it does not provide detailed insights into performance at a specific threshold. Therefore, we supply the KS, recall and F1 metrics at the best thresholds. Performance is evaluated every 1 thousand iterations on the validation dataset, where the model with the highest AUC is selected for further evaluation on the test dataset.

B. Overall Performance

1) Performance Evaluation: We compare ESCM^2 against baselines for CVR estimation in Table II. Key observations are summarized as follows:

- Debiased baselines generally outperform biased methods. For example, MTL-IPS attains the highest AUC and F1 scores, improving ESMM's AUC by 0.39% and KS by 1.04%. This highlights the potential of integrating unbiased estimators to improve ESMM's CVR estimation.
- ESCM^2 significantly enhances performance beyond the best baseline models, attributed to its effective mitigation of the IEB and FIP defects with ESMM, and the benefits of ESMM structure to address data sparsity.

In industry, CTCVR is a more prevalent metric as it encapsulates both clicks and conversions. Notable findings of CTCVR estimation in Table II, are summarized below:

- ESMM demonstrates competitive CTCVR performance, achieving advanced recall and F1 scores on the Ali-CCP dataset and leading recall and AUC scores on the industry dataset. This success is attributed to the inclusion of CTCVR estimation in ESMM's learning objectives.

[3] https://tianchi.aliyun.com/datalab/dataSet.html?dataId=408

[4] https://github.com/PaddlePaddle/PaddleRec/tree/master/models/multitask

[5] https://github.com/DongHande/AutoDebias/tree/main/baselines

TABLE II
OVERALL PERFORMANCE COMPARISON OF CVR AND CTCVR ESTIMATION (MEAN±STD)

Dataset	Model	CVR task				CTCVR task			
		Recall	F1	KS	AUC	Recall	F1	KS	AUC
Industry	Naïve	$0.5789_{\pm 0.0071}$	$0.3344_{\pm 0.0052}$	$0.3872_{\pm 0.0043}$	$0.7515_{\pm 0.0164}$	$0.6602_{\pm 0.0764}$	$1.0048_{\pm 0.1751}$	$0.4631_{\pm 0.0054}$	$0.7954_{\pm 0.0091}$
	ESMM	$0.5742_{\pm 0.0055}$	$0.6330_{\pm 0.0074}$	$0.3856_{\pm 0.0051}$	$0.7547_{\pm 0.0183}$	$\underline{0.6819}_{\pm 0.0672}$	$1.1062_{\pm 0.0527}$	$0.4827_{\pm 0.0042}$	$\underline{0.8153}_{\pm 0.0094}$
	MTL-EIB	$0.5121_{\pm 0.0072}$	$0.5808_{\pm 0.0048}$	$0.3371_{\pm 0.0051}$	$0.7272_{\pm 0.0140}$	$0.5975_{\pm 0.0590}$	$0.8458_{\pm 0.0097}$	$0.4220_{\pm 0.0043}$	$0.7912_{\pm 0.0108}$
	MTL-IMP	$\underline{0.5841}_{\pm 0.0092}$	$0.1272_{\pm 0.0040}$	$0.3974_{\pm 0.0047}$	$0.7563_{\pm 0.0114}$	$0.5880_{\pm 0.0765}$	$\underline{\mathbf{1.3393}}_{\pm 0.0094}$	$0.4126_{\pm 0.0047}$	$0.7752_{\pm 0.0084}$
	MTL-IPS	$0.5651_{\pm 0.0068}$	$\underline{0.6810}_{\pm 0.0042}$	$0.3960_{\pm 0.0048}$	$\underline{0.7586}_{\pm 0.0112}$	$0.6716_{\pm 0.0502}$	$1.0653_{\pm 0.0081}$	$0.4840_{\pm 0.0063}$	$0.8044_{\pm 0.0067}$
	MTL-DR	$0.5137_{\pm 0.0081}$	$0.6804_{\pm 0.0042}$	$\underline{0.4016}_{\pm 0.0046}$	$0.7579_{\pm 0.0135}$	$0.6684_{\pm 0.0733}$	$1.1707_{\pm 0.0061}$	$\underline{0.4844}_{\pm 0.0044}$	$0.8106_{\pm 0.0107}$
	ESCM^2-IPS	$0.5932_{\pm 0.0094}$	$\mathbf{0.7161}^{*}_{\pm 0.0089}$	$\mathbf{0.4144}^{*}_{\pm 0.0051}$	$\mathbf{0.7730}_{\pm 0.0150}$	$0.6804_{\pm 0.0771}$	$1.1753_{\pm 0.0259}$	$0.8198_{\pm 0.0042}$	$0.7730_{\pm 0.0062}$
	ESCM^2-DR	$\mathbf{0.5986}^{*}_{\pm 0.0068}$	$0.6884_{\pm 0.0052}$	$0.4119_{\pm 0.0050}$	$0.7679_{\pm 0.0113}$	$\mathbf{0.7013}^{*}_{\pm 0.0852}$	$1.2842_{\pm 0.0070}$	$\mathbf{0.5134}^{*}_{\pm 0.0038}$	$\mathbf{0.8265}_{\pm 0.0090}$
Ali-CCP	Naïve	$0.2854_{\pm 0.0050}$	$0.0991_{\pm 0.0053}$	$0.1123_{\pm 0.0056}$	$0.5987_{\pm 0.0139}$	$0.2921_{\pm 0.0044}$	$0.0978_{\pm 0.0036}$	$0.1192_{\pm 0.0040}$	$0.6003_{\pm 0.0133}$
	ESMM	$0.2968_{\pm 0.0036}$	$0.1157_{\pm 0.0084}$	$\underline{0.1267}_{\pm 0.0043}$	$0.6071_{\pm 0.0133}$	$\underline{0.3027}_{\pm 0.0032}$	$\underline{0.1139}_{\pm 0.0050}$	$0.1292_{\pm 0.0036}$	$0.6081_{\pm 0.0125}$
	MTL-EIB	$0.2372_{\pm 0.0043}$	$0.0825_{\pm 0.0051}$	$0.0717_{\pm 0.0057}$	$0.5603_{\pm 0.0135}$	$0.2542_{\pm 0.0046}$	$0.0959_{\pm 0.0039}$	$0.0697_{\pm 0.0040}$	$0.5699_{\pm 0.0140}$
	MTL-IMP	$0.2962_{\pm 0.0056}$	$0.1135_{\pm 0.0055}$	$0.1163_{\pm 0.0043}$	$\underline{0.6114}_{\pm 0.0137}$	$0.2973_{\pm 0.0045}$	$0.1110_{\pm 0.0039}$	$0.1264_{\pm 0.0038}$	$0.6087_{\pm 0.0155}$
	MTL-IPS	$\underline{0.2975}_{\pm 0.0070}$	$0.0941_{\pm 0.2163}$	$0.1177_{\pm 0.0063}$	$0.6091_{\pm 0.0123}$	$0.2991_{\pm 0.0050}$	$0.1044_{\pm 0.0066}$	$0.1302_{\pm 0.0028}$	$\underline{0.6138}_{\pm 0.0155}$
	MTL-DR	$0.2953_{\pm 0.0178}$	$\underline{0.1159}_{\pm 0.0084}$	$0.1255_{\pm 0.0141}$	$0.6065_{\pm 0.0172}$	$0.2980_{\pm 0.0168}$	$0.1096_{\pm 0.0075}$	$\underline{0.1360}_{\pm 0.0164}$	$0.6130_{\pm 0.0192}$
	ESCM^2-IPS	$0.3061_{\pm 0.0059}$	$0.1180_{\pm 0.0047}$	$0.1312_{\pm 0.0060}$	$\mathbf{0.6163}_{\pm 0.0151}$	$\mathbf{0.3184}^{*}_{\pm 0.0051}$	$\mathbf{0.1207}^{*}_{\pm 0.0038}$	$0.1436_{\pm 0.0038}$	$0.6189_{\pm 0.0118}$
	ESCM^2-DR	$\mathbf{0.3095}^{*}_{\pm 0.0054}$	$\mathbf{0.1315}^{*}_{\pm 0.0053}$	$\mathbf{0.1393}^{*}_{\pm 0.0042}$	$0.6142_{\pm 0.0133}$	$0.3117_{\pm 0.0044}$	$0.1180_{\pm 0.0040}$	$\mathbf{0.1494}^{*}_{\pm 0.0038}$	$\mathbf{0.6245}_{\pm 0.0123}$

1. Bold indicates the best performance. The underline marks the best performance across all baseline models except ESCM^2.
2. "*" marks the significant improvement of the bold metrics relative to underlined metrics, with p-value < 0.01 in the paired samples t-test.
3. Mean and standard deviation are reported over 10 random seeds. F1 is reported in percentages to highlight the performance differences.

TABLE III
ONLINE A/B TEST RESULTS IN 3 SCENARIOS

Scenario	# UV	# PV	# Order	# Premium	CVR	CTCVR
1	2.2M	3.1M	+2.84%	+10.85%	+5.64%	+3.92%
2	3.4M	4.9M	+4.26%	+3.88%	+0.43%	+1.75%
3	125K	136K	+40.55%	+12.69%	-	-

TABLE IV
COMPREHENSIVE ONLINE A/B TEST RESULTS IN SCENARIO 1

Metrics	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
# Order	−9.76%	−1.85%	−1.43%	**+9.07%**	**+0.73%**	**+6.26%**
# Premium	**+64.53%**	**+37.47%**	**+22.09%**	−12.49%	**+4.26%**	**+11.10%**
UV−CVR	**+7.25%**	−1.66%	**+9.39%**	**+8.58%**	**+2.51%**	**+8.62%**
UV−CTCVR	**+0.20%**	−3.50%	**+2.50%**	**+9.48%**	**+2.75%**	**+6.64%**

- ESCM^2 outperforms all competitors in CTCVR estimation, albeit with a smaller margin compared to CVR estimation. The superiority primarily stems from two factors: (1) incorporating CTCVR learning objective improves the estimation of CTCVR. Secondly, mitigating the IEB and FIP defects through counterfactual regularizers offers more accurate CVR estimate and thereby facilitating CTCVR estimation.

C. Online A/B Test

To further demonstrate the advantage of ESCM^2 over ESMM, online experiments are conducted on the industrial recommendation systems in Alipay. We first implement ESMM and ESCM^2[6] using our C++ based machine learning engine and open data processing service. Then, unique visitors (UV) are assigned to either ESMM or ESCM^2, and performance is compared using four metrics: UV-CVR, UV-CTCVR, order quantity (# Order) and total premium (# Premium). Experiments across three large-scale scenarios, summarized in Table III, yielded the following results:

- **Scenario 1.** This scenario is the insurance recommendation from Alipay. Over six days with 3.1M page views (PVs) and 2.2M UVs, ESCM^2 increased the total premium by 10.85%, order quantity by 2.84% and UV-CVR by 5.64%. Daily comparisons are performed in Table IV, where ESCM^2 performs best in most metrics.
- **Scenario 2.** This scenario is a renovation of the scenario above, where ESCM^2 boosted the total premium by 3.88%, order quantity by 4.26%, UV-CVR by 0.43%, and UV-CTCVR by 1.75%.
- **Scenario 3.** The third scenario is the Wufu campaign in Alipay, where ESCM^2 achieved a 40.55% increase in order quantity and a 12.69% rise in premium.

D. Additional Study on Intrinsic Estimation Bias

In this section, we examine the IEB defect in ESMM and assess the effectiveness of ESCM^2 in addressing it. We compare the average CVR labels ($\bar{r}$) with the model estimates ($\tilde{r}$) as shown in Table V. Since conversion labels are only available within the click space, $\bar{r}$ serves as an upper bound for the actual average CVR across the entire exposure space. Thus, any deviation from $\bar{r}$ provides a lower-bound estimate of the CVR estimation bias.

Table V indicates that ESMM consistently overestimates CVR values, validating the theoretical result in Theorem 1 and confirming the presence of IEB. Conversely, ESCM^2 significantly mitigates CVR estimation bias. On the industry dataset, ESCM^2-IPS achieves a bias reduction of 48.95% and 28.55% in the training and test sets, respectively, while ESCM^2-DR achieves reductions of 58.58% and 36.81%.Similar improvements are observed on the Ali-CCP dataset, with ESCM^2-IPS and ESCM^2-DR reducing bias by comparable percentages, highlighting the general effectiveness of ESCM^2 in addressing IEB. This efficacy is attributed to the counterfactual regularizers in ESCM^2, which ensure unbiased CVR estimates, as demonstrated in Theorem 4

E. Additional Study on False Independence Prior

In this analysis, we investigate the FIP defect in ESMM and assess how ESCM^2 addresses it. FIP arises since the model

[6] We use the IPS regularizer for advantageous training efficiency.

TABLE V
OVERVIEW OF THE IEB DEFECT WITH ESMM AND THE EFFICACY OF ESCM^2 TO MITIGATE IEB

Dataset	Subset	$\bar{r}$	ESMM		ESCM^2-IPS			ESCM^2-DR		
			$\tilde{r}$	$\lvert\tilde{r}-\bar{r}\rvert$	$\tilde{r}$	$\lvert\tilde{r}-\bar{r}\rvert$	$\Delta(\%)$	$\tilde{r}$	$\lvert\tilde{r}-\bar{r}\rvert$	$\Delta(\%)$
Ali-CCP	Train	0.0055	$0.0101_{\pm 0.0011}$	$0.0046_{\pm 0.0011}$	$0.0059_{\pm 0.0005}$	$0.0004_{\pm 0.0005}$	91.30%↓	$0.0076_{\pm 0.0018}$	$0.0021_{\pm 0.0018}$	54.34%↓
Ali-CCP	Test	0.0056	$0.0113_{\pm 0.0008}$	$0.0057_{\pm 0.0008}$	$0.0060_{\pm 0.0009}$	$0.0004_{\pm 0.0009}$	92.98%↓	$0.0059_{\pm 0.0009}$	$0.0003_{\pm 0.0009}$	94.73%↓
Industry	Train	0.0953	$0.1588_{\pm 0.0107}$	$0.0635_{\pm 0.0107}$	$0.1277_{\pm 0.0053}$	$0.0324_{\pm 0.0053}$	48.97%↓	$0.1216_{\pm 0.0093}$	$0.0263_{\pm 0.0093}$	58.58%↓
Industry	Test	0.0407	$0.1643_{\pm 0.0095}$	$0.1236_{\pm 0.0095}$	$0.1290_{\pm 0.0092}$	$0.0883_{\pm 0.0092}$	28.55%↓	$0.1188_{\pm 0.0022}$	$0.0781_{\pm 0.0022}$	36.81%↓

1. $\bar{r}$ indicates the average of the CVR label in the click space, calculated as the quotient of the number of samples: $|\mathcal{O}|/|\mathcal{D}|$.
2. $\tilde{r}$ indicates the average of the model's CVR estimates for all samples in the dataset. Its bias from the mean CVR ground truth is denoted as $|\tilde{r}-\bar{r}|$.
3. Δ denotes the relative reduction in the model's bias compared to the ESMM.

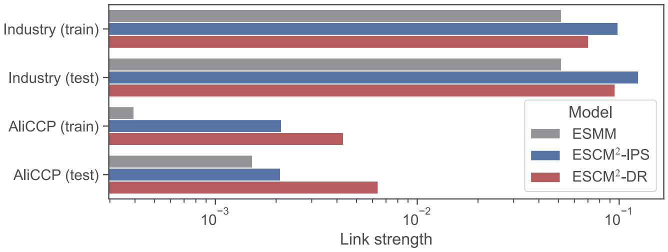


Fig. 5. Comparative study of the causal link strength $O \rightarrow R$ with and without counterfactual regularizers.

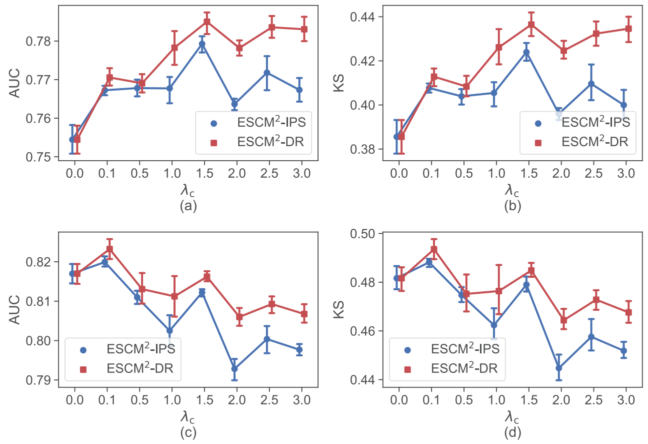


Fig. 6. Performance of CVR (a-b) and CTCVR estimation (c-d) with varying counterfactual risk weight λ_c.

fails to capture the causal link between clicks and conversions, as illustrated by the missing $O \rightarrow R$ in Fig. 3 (a). To quantify FIP, we measure the causation strength from clicks to conversions using propensity score matching, which eliminates the confounding effect of X in Fig. 3, CVR and CTR estimates are treated as outcomes and propensities, respectively, to study the causal link strength. Following [28], we divide samples into clicked and unclicked groups and pair clicked samples with unclicked ones that have the most similar propensity scores. The causal link strength $O \rightarrow R$ is estimated using the causal risk ratio (CRR) [29], where a CRR close to 1 indicates weak causation. Thus, the causation strength is quantified as $|\text{CRR} - 1|$.

Fig. 5 shows that ESMM exhibits minimal CRR, approximately 0.05 and 0.001 on the industry and Ali-CCP datasets, respectively, confirming the presence of the FIP defect. In contrast, ESCM^2 significantly increases causation strength, with ESCM^2-IPS achieving over 0.12 and 0.002 on the respective datasets. This improvement is attributed to ESCM^2's counterfactual regularizers, which estimate CVR as $\mathbb{P}(r_{u,i} = 1 \mid do(o_{u,i} = 1))$ according to Theorem 5, explicitly accounting for the causal effect of clicks on conversions and mitigating the FIP defect.

F. Hyper-Parameter Tuning and Ablation Study

Two crucial hyperparameters of ESCM^2 are the weighting factors (*i.e.,*λ_c and λ_g) in the learning objective (13). In this section, they are tuned within the range [0, 3] to investigate the impact of causal regularization and global risk minimization on the performance of CVR and CTCVR estimation.

- The weighting factor λ_c is investigated in Fig. 6. Evidently, increasing λ_c consistently benefits CVR estimation, which showcases the effectiveness of causal regularization. The AUC of ESCM^2-DR, for instance, grows from 0.755 at $\lambda_c = 0$ where the causal regularization is not applied, to around 0.785 at $\lambda_c = 1.5$. Additionally, causal regularization also benefits CTCVR estimates. For example, the AUC of ESCM^2-IPS boosts from 0.817 at $\lambda_c = 0$ to 0.821 at $\lambda_c = 0.1$. Nevertheless, the overemphasis on CVR risk has a negative impact on CTCVR estimation. For example, a drop in AUC by 0.012 is observed for ESCM^2-IPS from $\lambda_c = 0.1$ to $\lambda_c = 3.0$. This phenomenon is attributed to the seesaw effect in multitask learning [30], *i.e.,* overemphasis on the CVR risk misleads the optimizer to ignore the CTR risk, reducing CTR and CTCVR estimation performance. Therefore, we suggest tuning λ_c within the range [0, 0.1].
- The weighting factor λ_g is studied in Fig. 7. Overall, increasing λ_g within the range [0, 3] is beneficial for both CTR and CTCVR estimation. In the CVR estimation task, for instance, increasing λ_g from 0 to 2.5, the KS of ESCM^2-DR climbs from 0.385 to about 0.434; the AUC of ESCM^2-IPS grows by 0.023, significantly. These observations verify the effectiveness of entire space multitask modelling paradigm, exploiting the sequential user behavior track as per Fig. 2.

VI. RELATED WORK

Post-click conversion rate (CVR) estimation is a crucial task in recommendation, which aims to predict the likelihood of a user completing a transaction after clicking [1], [31], [32].

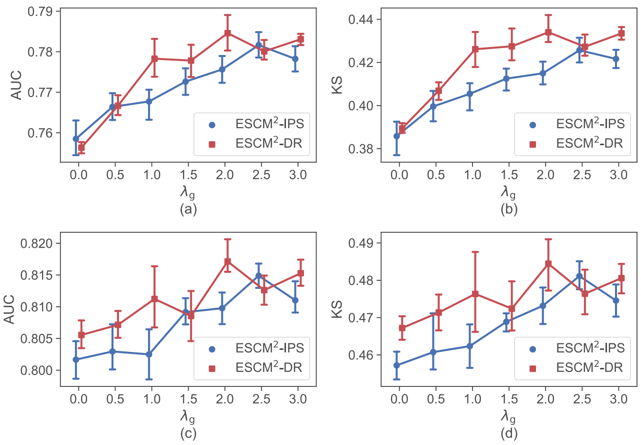


Fig. 7. Performance of CVR (a-b) and CTCVR estimation (c-d) with varying CTCVR risk weight λ_g.

Accurate CVR estimation not only helps to drive transactions, but also improves traffic resource allocation, resulting in increased revenue. However, CVR estimation faces significant challenges, namely sample selection bias and data sparsity [1], [7]. To address these challenges, existing methods generally fall into two paradigms: the entire space multitask paradigm and the causal recommendation paradigm.

A. Entire Space Multitask Paradigm

The entire space multitask paradigm, pioneered with ESMM [1], innovatively avoids training CVR estimator directly. Instead, it employs a multitask approach to optimize two independent learning objectives for CTCVR and CTR [1]. This strategy leverages all exposure samples for training both CTCVR and CTR objectives, thereby alleviating data sparsity and enhancing practical performance. On the basis of ESMM, a line of work initiated in [13] advocates for including additional conversion-related actions, such as adding to favorites and cart. These actions enrich the user behavior track and introduce auxiliary tasks, providing more granular supervision labels to further mitigate data sparsity [33], [34], [35]. Subsequently, graph models have been explored as alternatives to the Markov chain approach for capturing increasingly complex interactions among user behaviors [36], [37]. In another line of work, arious plug-ins, such as language models [38] and the delayed feedback calibrator [39], have been integrated into the training paradigm to improve recommendation quality.

This paradigm is particularly prevalent in industrial recommendation systems where accurate CVR estimation is crucial. The data sparsity issue is effectively mitigated by this paradigm through the incorporation of auxiliary tasks with abundant data; however, sample selection bias persists due to the absence of unbiasedness guarantee [7] and overly simplistic dependency assumptions. These concerns are formulated as two defects with ESMM in this study, namely the intrinsic estimation bias and the false independence prior, which renders ESMM's CVR estimation biased and leaves room for further improvement.

B. Causal Recommendation Paradigm

The causal recommendation paradigm, initiated with [10], estimates the ideal learning objective by adjusting the biased dataset using propensity scores. On this basis, a line of work focuses on enhancing the estimation of propensity scores, which is crucial for ensuring the unbiasedness of the adjusted learning objective [16]. Early methods estimated the propensity score based on heuristic item popularity [40], [41], and later progressed to parametric models like logistic regression [7], [42]. Subsequent advancements have incorporated various learning techniques, such as feature selection [43], joint optimization [7], [44], alternative training [45] and kernel balancing [46], for enhanced identifiability and estimation quality. Another line of works advocates for more complex causal adjustment approaches to reduce estimation variance [47], improve training stability [16], resist noisy labels [48] and model mis-specification [49]. Some recent works innovatively incorporate a small subset of unbiased data during training [9], [50], [51], which effectively addresses missing confounders while minimizing additional data collection efforts.

While this paradigm is widely acknowledged in academic research, its application in industrial recommendation scenarios remains limited. Although the sample selection bias is effectively addressed through causal adjustment with solid theoretical guarantees; the training only involves treated (*i.e.,* clicked) samples which are sparse in real-world applications. As a result, the issue of data sparsity is ignored by this paradigm, which severely compromises the performance of CVR estimators in industrial practice.

In conclusion, the entire space multitask paradigm and the causal recommendation excel handling data sparsity and sample selection bias, respectively, but fall short of tackling both challenges concurrently. We innovatively synthesize the two paradigms and construct ESCM^2, which incorporates a counterfactual risk regularizer within the ESMM framework to regularize CVR estimation. It leverages the advantage of ESMM for mitigating data sparsity, while addressing sample selection bias by estimating unbiased CVR through counterfactual regularizers.

VII. Conclusion

This study demonstrates that the ESMM method, while effective, secretly suffers from the IEB and PIP issues. To address these challenges, the research introduces a counterfactual risk regularizer within the ESMM framework. This integration not only preserves ESMM's ability to mitigate data sparsity but also effectively addresses sample selection bias through the use of counterfactual regularization techniques. Empirical evidence from real-world experiments validates the efficacy of the proposed method, demonstrating its capability to alleviate both the IEB and PIP issues, thereby enhancing the CVR estimation performance relative to the original ESMM approach.

Limitation & future works. In this work, we focus on the sequential user behavior track illustrated in Fig. 2. However, in industrial scenarios, there are diverse user behaviors whose dependencies excess the expressive capabilities of Markov chains. Although some studies [13], [36] extend ESMM to describe such complex dependencies via decomposition, they also inevitably suffer from both IEB and FIP defects. Leveraging the counterfactual regularization techniques in ESCM^2, these methods could be further enhanced to effectively address both IEB and FIP defects.

Acknowledgment

The authors would like to thank the Ant Group for providing their expertise in data preparation and online A/B test.

References

[1] X. Ma et al., "Entire space multi-task model: An effective approach for estimating post-click conversion rate," in *Proc. SIGIR*, 2018, pp. 1137–1140.

[2] G. Zhou et al., "Deep interest network for click-through rate prediction," in *Proc. SIGKDD*, Jul. 2018, pp. 1059–1068.

[3] C. Gao, T.-H. Lin, N. Li, D. Jin, and Y. Li, "Cross-platform item recommendation for online social e-commerce," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 2, pp. 1351–1364, Feb. 2023.

[4] Y. Lai, Y. Zhu, W. Fan, X. Zhang, and K. Zhou, "Toward adversarially robust recommendation from adaptive fraudster detection," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 907–919, 2024.

[5] X. Zhang, J. Chen, R. Zhang, C. Wang, and L. Liu, "Attacking recommender systems with plausible profile," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4788–4800, 2021.

[6] V. Kumar, D. Khattar, S. Gairola, Y. Kumar Lal, and V. Varma, "Identifying clickbait: A multi-strategy approach using neural networks," in *Proc. SIGIR*, Jun. 2018, pp. 1225–1228.

[7] W. Zhang et al., "Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning," in *Proc. WWW*, Apr. 2020, pp. 2775–2781.

[8] H. Wang et al., "ESCM2: Entire space counterfactual multi-task model for post-click conversion rate estimation," in *Proc. SIGIR*, Jul. 2022, pp. 363–372.

[9] H. Li et al., "Removing hidden confounding in recommendation: A unified multi-task learning approach," in *Proc. NeurIPS*, vol. 36, 2024, pp. 54614–54626.

[10] T. Schnabel, A. Swaminathan, A. Singh, N. Chandak, and T. Joachims, "Recommendations as treatments: Debiasing learning and evaluation," in *Proc. ICML*, 2016, pp. 1670–1679.

[11] B. Marlin, R. S. Zemel, S. Roweis, and M. Slaney, "Collaborative filtering and the missing at random assumption," 2012, *arXiv:1206.5267*.

[12] H. Wang et al., "Optimal transport for treatment effect estimation," in *Proc. NeurIPS*, vol. 36, 2024, pp. 5404–5418.

[13] H. Wen et al., "Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction," in *Proc. SIGIR*, Jul. 2020, pp. 2377–2386.

[14] X. Wang, R. Zhang, Y. Sun, and J. Qi, "Doubly robust joint learning for recommendation on data missing not at random," in *Proc. ICML*, May 2019, pp. 6638–6647.

[15] H. Li et al., "Multiple robust learning for recommendation," in *Proc. AAAI*, Jun. 2023, pp. 4417–4425.

[16] H. Li, C. Zheng, X.-H. Zhou, and P. Wu, "Stabilized doubly robust learning for recommendation on data missing not at random," in *Proc. ICLR*, 2023, pp. 1–9.

[17] T. Gu et al., "Estimating true post-click conversion via group-stratified counterfactual inference," in *Proc. ADKDD*, 2021, pp. 1–6.

[18] E. Bareinboim and J. Pearl, "Controlling selection bias in causal inference," in *Proc. AISTATS*, 2012, pp. 100–108.

[19] J. Pearl, *Causality*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[20] E. L. Ionides, "Truncated importance sampling," *J. Comput. Graph. Statist.*, vol. 17, no. 2, pp. 295–311, 2008.

[21] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. Amer. Stat. Assoc.*, vol. 95, no. 449, pp. 135–143, Mar. 2000.

[22] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *Stat. Sci.*, vol. 34, no. 1, pp. 129–155, Feb. 2019.

[23] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *Proc. SIGKDD*, 2018, pp. 1930–1939.

[24] D. Xi et al., "Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising," in *Proc. SIGKDD*, Aug. 2021, pp. 3745–3755.

[25] H. Steck, "Training and testing of recommender systems on data missing not at random," in *Proc. SIGKDD*, Jul. 2010, pp. 713–722.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–9.

[27] Z.-Y. Zhang et al., "Towards understanding the overfitting phenomenon of deep click-through rate models," in *Proc. CIKM*, Oct. 2022, pp. 2671–2680.

[28] M. M. Garrido et al., "Methods for constructing and assessing propensity scores," *Health Services Res.*, vol. 49, no. 5, pp. 1701–1720,2014.

[29] R. J. Hernán, *Causal Inference: What If*. Boca Raton, FL, USA: CRC Press, 2020.

[30] Y. Huang, E. H. Wang, Z. Liu, L. Pan, H. Li, and X. Liu, "Modeling task relationships in multivariate soft sensor with balanced mixture-of-experts," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6556–6564, May 2023.

[31] Z. Chan et al., "Capturing conversion rate fluctuation during sales promotions: A novel historical data reuse approach," in *Proc. SIGKDD*, Aug. 2023, pp. 3774–3784.

[32] Z. Chen, R. Xiao, C. Li, G. Ye, H. Sun, and H. Deng, "ESAM: Discriminative domain adaptation with non-displayed items to improve long-tail performance," in *Proc. SIGIR*, Jul. 2020, pp. 579–588.

[33] Y. Yang, C. Huang, L. Xia, Y. Liang, Y. Yu, and C. Li, "Multi-behavior hypergraph-enhanced transformer for sequential recommendation," in *Proc. SIGKDD*, Aug. 2022, pp. 2263–2274.

[34] Y. Shen, B. Ou, and R. Li, "MBN: Towards multi-behavior sequence modeling for next basket recommendation," *ACM Trans. Knowl. Discovery Data*, vol. 16, no. 5, pp. 1–23, Oct. 2022.

[35] J. Jin et al., "Multi-scale user behavior network for entire space multi-task learning," in *Proc. CIKM*, Oct. 2022, pp. 874–883.

[36] W. Bao, H. Wen, S. Li, X.-Y. Liu, Q. Lin, and K. Yang, "GMCM: Graph-based micro-behavior conversion model for post-click conversion rate estimation," in *Proc. SIGIR*, Jul. 2020, pp. 2201–2210.

[37] H. Wen, J. Zhang, F. Lv, W. Bao, T. Wang, and Z. Chen, "Hierarchically modeling micro and macro behaviors via multi-task learning for conversion rate prediction," in *Proc. SIGIR*, Jul. 2021, pp. 2187–2191.

[38] X. Wu, A. Magnani, S. Chaidaroon, A. Puthenputhussery, C. Liao, and Y. Fang, "A multi-task learning framework for product ranking with BERT," in *Proc. WWW*, Apr. 2022, pp. 493–501.

[39] Y. Zhao et al., "Entire space cascade delayed feedback modeling for effective conversion rate prediction," in *Proc. CIKM*, Oct. 2023, pp. 4981–4987.

[40] Y. Saito, S. Yaginuma, Y. Nishino, H. Sakata, and K. Nakata, "Unbiased recommender learning from missing-not-at-random implicit feedback," in *Proc. WSDM*, Jan. 2020, pp. 501–509.

[41] T. Joachims, A. Swaminathan, and T. Schnabel, "Unbiased learning-to-rank with biased feedback," in *Proc. WSDM*, 2017, pp. 781–789.

[42] J.-W. Lee, S. Park, and J. Lee, "Dual unbiased recommender learning for implicit feedback," in *Proc. SIGIR*, Jul. 2021, pp. 1647–1651.

[43] S. M. Shortreed and A. Ertefaie, "Outcome-adaptive lasso: Variable selection for causal inference," *Biometrics*, vol. 73, no. 4, pp. 1111–1122, Dec. 2017.

[44] Y. Zhang et al., "User retention: A causal approach with triple task modeling," in *Proc. IJCAI*, Aug. 2021, pp. 3399–3405.

[45] Z. Zhu, Y. He, Y. Zhang, and J. Caverlee, "Unbiased implicit recommendation and propensity estimation via combinational joint learning," in *Proc. RecSys*, Sep. 2020, pp. 551–556.

[46] H. Li et al., "Debiased collaborative filtering with kernel-based causal balancing," in *Proc. ICLR*, Apr. 2024, pp. 1–9.

[47] S. Guo et al., "Enhanced doubly robust learning for debiasing post-click conversion rate estimation," in *Proc. SIGIR*, Jul. 2021, pp. 275–284.

[48] H. Li, C. Zheng, W. Wang, H. Wang, F. Feng, and X.-H. Zhou, "Debiased recommendation with noisy feedback," in *Proc. SIGKDD*, Aug. 2024, pp. 1576–1586.

[49] H. Li et al., "Relaxing the accurate imputation assumption in doubly robust learning for debiased collaborative filtering," in *Proc. ICML*, vol. 235, 2024, pp. 29448–29460.

[50] J. Chen et al., "AutoDebias: Learning to debias for recommendation," in *Proc. SIGIR*, Jul. 2021, pp. 21–30.

[51] H. Li, Y. Xiao, C. Zheng, and P. Wu, "Balancing unobserved confounding with a few unbiased ratings in debiased recommendations," in *Proc. WWW*, Apr. 2023, pp. 1305–1313.

[52] Y. Saito, "Asymmetric tri-training for debiasing missing-not-at-random explicit feedback," in *Proc. SIGIR*, 2020, pp. 309–318.

[53] Z. Wang, X. Chen, R. Wen, S.-L. Huang, E. Kuruoglu, and Y. Zheng, "Information theoretic counterfactual learning from missing-not-at-random feedback," in *Proc. NeurIPS*, vol. 33, 2020, pp. 1854–1864.

[54] Q. Dai et al., "A generalized doubly robust learning framework for debiasing post-click conversion rate prediction," in *Proc. SIGKDD*, Aug. 2022, pp. 252–262.