

# ESCM<sup>2</sup>: Entire Space Counterfactual Multi-Task Model for Post-Click Conversion Rate Estimation

Hao Wang  
Ant Group  
Hangzhou, China  
zhiqiang.wh@antgroup.com

Tai-Wei Chang  
Ant Group  
Hangzhou, China  
taiwei.twc@antgroup.com

Tianqiao Liu  
Ant Group  
Hangzhou, China  
liutianqiao.ltq@antgroup.com

Jianmin Huang  
Ant Group  
Hangzhou, China  
caiqiong@antgroup.com

Zhichao Chen  
Ant Group  
Hangzhou, China  
czc347886@antgroup.com

Chao Yu  
Ant Group  
Hangzhou, China  
tianjing.yc@antgroup.com

Ruopeng Li  
Ant Group  
Hangzhou, China  
ruopeng.lrp@antfin.com

Wei Chu  
Ant Group  
Hangzhou, China  
weichu.cw@antgroup.com

## ABSTRACT

Accurate estimation of post-click conversion rate is critical for building recommender systems, which has long been confronted with sample selection bias and data sparsity issues. Methods in the Entire Space Multi-task Model (ESMM) family leverage the sequential pattern of user actions, *i.e.*, *impression*  $\rightarrow$  *click*  $\rightarrow$  *conversion* to address data sparsity issue. However, they still fail to ensure the unbiasedness of CVR estimates. In this paper, we theoretically demonstrate that ESMM suffers from the following two problems: (1) Inherent Estimation Bias (IEB), where the estimated CVR of ESMM is inherently higher than the ground truth; (2) Potential Independence Priority (PIP) for CTCVR estimation, where there is a risk that the ESMM overlooks the causality from click to conversion. To this end, we devise a principled approach named Entire Space Counterfactual Multi-task Modelling (ESCM<sup>2</sup>), which employs a counterfactual risk minimizer as a regularizer in ESMM to address both IEB and PIP issues simultaneously. Extensive experiments on offline datasets and online environments demonstrate that our proposed ESCM<sup>2</sup> can largely mitigate the inherent IEB and PIP issues and achieve better performance than baseline models.

## CCS CONCEPTS

• **Computer systems organization**  $\rightarrow$  **Neural networks**; • **Information systems**  $\rightarrow$  **Recommender systems**.

## KEYWORDS

Recommender System, Entire Space Multi-task Learning, Selection Bias, Propensity Score, Post-click Conversion Rate Estimation

### ACM Reference Format:

Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. 2022. ESCM<sup>2</sup>: Entire Space Counterfactual Multi-Task Model for Post-Click Conversion Rate Estimation. In *Woodstock '22: ACM Symposium on Neural Gaze Detection*, June 03–05, 2022, Woodstock, NY. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

Recommender system aims to deliver valuable items from a large body of candidates to users [7, 12], which has been the main driving force of user growth in e-commerce [14], social media [4], and advertising [26]. Figure 1 exhibits the two-stage pipeline to build an effective recommender system in an industrial setting. During the offline stage, the ranking model is trained with user-profile features, item features and user-item-interaction features parsed from user logs. During the online stage, we rely on a number of ranking metrics, including but not limited to click-through rate (CTR), post-click conversion rate (CVR), and click-through&conversion rate (CTCVR) to expose to the user the items that might capture his/her interest. The typical user behavior path in e-commerce setting can be summarized as *impression*  $\rightarrow$  *click*  $\rightarrow$  *conversion* [14, 16]. Feedback generated from users is utilized during the offline training stage to optimize the performance of the recommender system.

As shown in Figure 2, CVR denotes the transition probability from the click space to the conversion space. In general, CVR estimation is similar to CTR estimation besides that conversion feedback is unavailable for unclicked events, because not clicking on an item does not necessarily indicate that the user is not interested in purchasing it [25]. Hence, a common approach to building CVR estimator is to include only the clicked items in the training space while overlooking the unclicked items, as indicated in Figure 2.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Woodstock '22, June 03–05, 2022, Woodstock, NY

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00

<https://doi.org/10.1145/1122445.1122456>

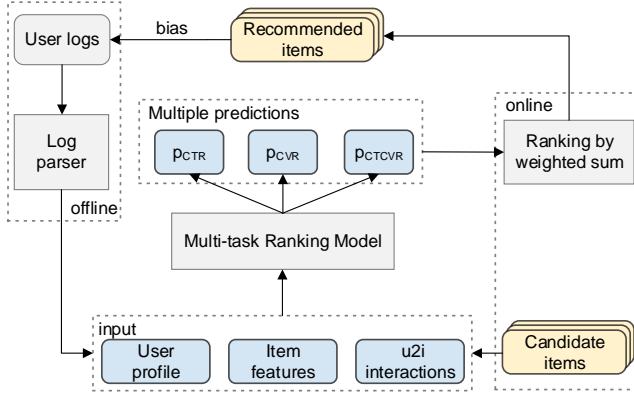


Figure 1: A diagram of recommender system in e-commerce.

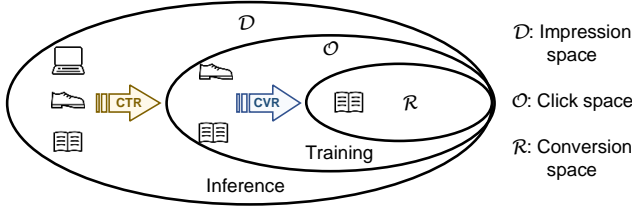


Figure 2: Illustration of sample selection bias and data sparsity in CVR estimation task, where the training space only consists of clicked samples, while the inference space is the entire space for all impression samples.

Two critical issues that challenge naïve CVR estimators have been reported by Ma et al. [14] and Zhang et al. [25]. The first problem is **sample selection bias** due to the training space composing solely of clicked items. Specifically, items with lower CVR are less likely to be clicked, *i.e.*, to be included in the training space, and vice versa [15], which makes the training space Missing Not At Random (MNAR). Therefore, there is a distribution shift between the training space  $\mathcal{O}$  and the inference space  $\mathcal{D}$ . Another problem stems from the **data sparsity** of clicked samples (in our case, we have a CTR of 3.8% on the production dataset and 4% on the public dataset). As CVR models are trained over click space, this problem severely hinders their generalization beyond click space.

Ma et al. proposed Entire Space Multi-task Model (ESMM) to address sample selection bias and data sparsity. It has mitigated data sparsity problem effectively via multi-task learning [14, 16, 22]; however, unbiasedness of its CVR estimate is still not guaranteed. Emerging empirical evidence [25] suggests that ESMM’s CVR estimation is biased, raising concerns of this approach.

In this paper, we report two critical issues of ESMM.

- **Inherent Estimation Bias (IEB):** We rigorously demonstrate that the estimated CVR of ESMM is higher than the ground truth even under very relaxed conditions. Subtle experiments are designed to support this claim.
- **Potential Independence Priority (PIP):** We demonstrate that there is a risk that ESMM estimates of CTR and CVR are conditionally independent, which is contrary to the reality.

Leveraging causality methodology, we propose Entire Space Counterfactual Multi-task Model (ESCM<sup>2</sup>), a model that incorporates counterfactual risk minimizer (CRM), *i.e.*, the inverse propensity score (IPS) and the doubly robust (DR) mechanism into ESMM to regularize CVR estimation. As we will discuss in Section 4, the introduced CRM regularizer makes CVR estimates unbiased and CTCVR estimates free from the priority of independence.

The contributions of this paper are summarized as follows.

- This is the first work that rigorously demonstrates the inherent bias of ESMM’s CVR estimates. Mathematical proofs and experiment results are provided to support this claim.
- We show that the ESMM’s CTCVR estimates are subjected to potential independence priority. We have designed experiments to back up this claim.
- We propose ESCM<sup>2</sup>, the first work that improves ESMM from a causal perspective. ESCM<sup>2</sup> effectively eliminates IEB and PIP in ESMM. Extensive experimental results and mathematical proofs are provided to verify our claims.

## 2 PRELIMINARIES

### 2.1 Notations

We use uppercase letter, *e.g.*,  $\mathcal{O}$  to denote a random variable and lowercase letter, *e.g.*,  $o$  to denote an associated specific value. Letters in calligraphic font, *e.g.*,  $\mathcal{O}$  represent the sample space of the corresponding random variable, and  $\mathbb{P}()$  represents the probability distribution of the random variable, *e.g.*,  $\mathbb{P}(\mathcal{O})$ .

### 2.2 Problem Formulation

Let  $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$  be the set of  $m$  users,  $\mathcal{I} = \{i_1, i_2, \dots, i_n\}$  be the set of  $n$  items, and  $\mathcal{D} = \mathcal{U} \times \mathcal{I}$  be the set of all exposed user-item pairs. Denote  $\mathcal{O}$  as the click matrix where each entry  $o_{u,i} \in \{0, 1\}$  indicates whether clicking takes place for user-item pair  $(u, i)$ ,  $\mathbf{R} \in \{0, 1\}^{m \times n}$  as the observed conversion label where each entry  $r_{u,i} \in \{0, 1\}$  indicates whether conversion takes place.

If  $\mathbf{R}$  is fully observed, the ideal loss function is formulated as:

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

where  $\delta$  denotes the prediction error of  $r_{u,i}$ ,  $\hat{r}_{u,i}$  denotes the predicted  $r_{u,i}$ . For  $\delta$ , we use the binary cross entropy loss:

$$\delta(r_{u,i}, \hat{r}_{u,i}) = -r_{u,i} \log \hat{r}_{u,i} - (1 - r_{u,i}) \log(1 - \hat{r}_{u,i}). \quad (2)$$

However,  $r_{u,i}$  can only be observed for user-item pairs in the click space  $\mathcal{O}$ . As such, a naïve approach [17, 20] estimates the ideal loss with the expectation over  $\mathcal{O}$ :

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

where  $|\mathcal{O}| = \sum_{(u,i) \in \mathcal{D}} (o_{u,i})$ . It is widely adopted by many existing methods, but leads to biased estimation, *i.e.*,  $\mathbb{E}_{\mathcal{O}}[\mathcal{L}_{\text{naive}}] \neq \mathcal{P}$ .

### 2.3 Entire Space Multitask Model Approach

The Entire Space Multitask Model Approach (ESMM) [14] uses chain rule to obtain CVR estimate indirectly:

$$\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1) = \frac{\mathbb{P}(r_{u,i} = 1, o_{u,i} = 1)}{\mathbb{P}(o_{u,i} = 1)}. \quad (4)$$

Two towers are used to predict CTR (*i.e.*,  $\mathbb{P}(o_{u,i} = 1)$ ) and CVR (*i.e.*,  $\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1)$ ) in ESMM. The product of these two towers gives CTCVR estimation (*i.e.*,  $\mathbb{P}(r_{u,i} = 1, o_{u,i} = 1)$ ). During the training stage, ESMM minimizes the empirical risk of CTR and CTCVR estimation over the entire impression space  $\mathcal{D}$ :

$$\begin{aligned}\mathcal{L}_{\text{CTR}} &= \mathbb{E}_{(u,i) \in \mathcal{D}} [\delta(o_{u,i}, \hat{o}_{u,i})] \\ \mathcal{L}_{\text{CTCVR}} &= \mathbb{E}_{(u,i) \in \mathcal{D}} [\delta(o_{u,i} * r_{u,i}, \hat{o}_{u,i} * \hat{r}_{u,i})].\end{aligned}\quad (5)$$

During the inference stage, ESMM uses the output of the CVR tower as its predicted CVR. This approach circumvents the sample selection bias problem by not modeling CVR over the click space. However, as illustrated in Section 4.1, such approach yields inherent overestimation of CVR. To that end, we seek to develop an unbiased CVR estimator to address sample selection bias.

## 2.4 Propensity Score Based Approach

The inverse propensity score (IPS) estimator [17] weights each clicked event with  $1/q_{u,i}$ , where the propensity  $q_{u,i} = \mathbb{P}(o_{u,i} = 1)$  denotes the probability of user  $u$  clicking item  $i$ , *i.e.*, CTR in the CVR estimation task. The IPS estimator can be formulated as:

$$\mathcal{L}_{\text{IPS}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i} \delta_{u,i}}{q_{u,i}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i} \delta_{u,i}}{\hat{q}_{u,i}}. \quad (6)$$

As the real value  $q_{u,i}$  is always unavailable, an auxiliary classifier is introduced to estimate the propensity  $q_{u,i}$  with  $\hat{q}_{u,i}$ . The IPS estimator renders an unbiased estimate of the ideal loss function, *i.e.*,  $\mathbb{E}_{\mathcal{O}}(\mathcal{L}_{\text{IPS}}) = \mathcal{P}$ , given that the estimated  $\hat{q}_{u,i}$  is accurate [17].

However, propensity score in the IPS estimator might suffer from severely high variance, for which the doubly robust (DR) estimator is introduced [20]. In particular, DR introduces imputed error  $\hat{\delta}_{u,i}$  to model the prediction error for all events in  $\mathcal{D}$ , and corrects the error deviation  $\hat{e}_{u,i} = \delta_{u,i} - \hat{\delta}_{u,i}$  for clicked events:

$$\mathcal{L}_{\text{DR}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \hat{\delta}_{u,i} + \frac{o_{u,i} \hat{e}_{u,i}}{\hat{q}_{u,i}}, \quad (7)$$

where  $\hat{q}_{u,i}$  seeks to eliminate the MNAR effect for  $e_{u,i}$ . Double robustness stems from that the estimator is unbiased if either the imputed error  $\hat{\delta}_{u,i}$  or the predicted propensity  $\hat{q}_{u,i}$  is accurate. The accuracy of  $\hat{\delta}_{u,i}$  and  $\hat{q}_{u,i}$  is usually ensured by auxiliary tasks.

## 3 DISCUSSION ON ESMM

### 3.1 Is ESMM an Unbiased CVR Estimator?

The existence of **inherent estimation bias (IEB)** in ESMM has been perceived by researchers [25]; however, to the best of our knowledge, a theoretical proof of its CVR estimation bias remains lacking<sup>1</sup>. In this paper, IEB is formulated and proved in Theorem 1.

**THEOREM 1.** *Let random variables  $O, R, C$  be the indicator of click, post-click conversion and click & conversion, and  $o_{u,i}, r_{u,i}, c_{u,i}$  be the corresponding value of  $O, R, C$  given user-item pairs, and  $\hat{o}_{u,i}, \hat{r}_{u,i}, \hat{c}_{u,i}$  be the predicted value of  $o_{u,i}, r_{u,i}, c_{u,i}$ . ESMM's CVR estimation is always larger than the ground-truth.*

<sup>1</sup>Zhang et al. [25] provided an example to illustrate the existence of bias in ESMM, while the theoretical proof was not solid. We provide a complete proof using probability inequalities in the Theorem 1.

**PROOF.** Following the loss function in Eq. (5), ESMM ensures:

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}[O - \hat{O}] &= \int (o_{u,i} - \hat{o}_{u,i}) d(u, i) = 0 \\ \mathbb{E}_{\mathcal{D}}[C - \hat{C}] &= \int (c_{u,i} - \hat{c}_{u,i}) d(u, i) = 0.\end{aligned}\quad (8)$$

Let  $\mathbb{E}_{\mathcal{D}}[R]$  and  $\mathbb{E}_{\mathcal{D}}[\hat{R}]$  be the expectation of CVR ground truth and estimate, respectively. The CVR estimation bias is

$$\begin{aligned}\text{Bias}^{\text{ESMM}} &\stackrel{(1)}{=} \mathbb{E}_{\mathcal{D}}[\hat{R}] - \mathbb{E}_{\mathcal{D}}[R] \\ &\stackrel{(2)}{>} \mathbb{E}_{\mathcal{D}}[\hat{R}] - \mathbb{E}_{\mathcal{O}}[R] \\ &\stackrel{(3)}{=} \mathbb{E}_{\mathcal{D}}[\hat{R}] - \frac{\mathbb{E}_{\mathcal{D}}[C]}{\mathbb{E}_{\mathcal{D}}[O]} \\ &\stackrel{(4)}{=} \mathbb{E}_{\mathcal{D}}\left[\frac{\hat{C}}{\hat{O}}\right] - \frac{\mathbb{E}_{\mathcal{D}}[C]}{\mathbb{E}_{\mathcal{D}}[O]},\end{aligned}\quad (9)$$

where (1) is the CVR estimation bias over the exposure space  $\mathcal{D}$ , (2) uses the fact that the expectation of CVR label over  $\mathcal{O}$  is large than that over  $\mathcal{D}$ , *i.e.*,  $\mathbb{E}_{\mathcal{O}}[R] > \mathbb{E}_{\mathcal{D}}[R]^2$ , (3) estimates the ground-truth CVR expectation over  $\mathcal{D}$  following the label distribution regime in Fig 2, (4) stems from the probability decomposition of ESMM.

Let  $\mathbb{P}(\hat{c}, \hat{o})$  be the joint probability of  $\hat{C} = \hat{c}$  and  $\hat{O} = \hat{o}$  over exposure space. The first term in Eq. (9) is formulated as

$$\begin{aligned}\mathbb{E}_{\mathcal{D}}\left[\frac{\hat{C}}{\hat{O}}\right] &\stackrel{(1)}{=} \int \frac{\hat{c}}{\hat{o}} \mathbb{P}(\hat{c}, \hat{o}) d(\hat{c}, \hat{o}) \\ &\stackrel{(2)}{\geq} \int \frac{\hat{c}}{\hat{o}} \mathbb{P}(\hat{c}) \mathbb{P}(\hat{o}) d(\hat{c}, \hat{o}) \\ &\stackrel{(3)}{=} \int \hat{c} \mathbb{P}(\hat{c}) d\hat{c} \int \frac{1}{\hat{o}} \mathbb{P}(\hat{o}) d\hat{o} \\ &\stackrel{(4)}{=} \mathbb{E}_{\mathcal{D}}[\hat{C}] \mathbb{E}_{\mathcal{D}}\left[\frac{1}{\hat{O}}\right] \\ &\stackrel{(5)}{\geq} \frac{\mathbb{E}_{\mathcal{D}}[\hat{C}]}{\mathbb{E}_{\mathcal{D}}[\hat{O}]} \stackrel{(6)}{=} \frac{\mathbb{E}_{\mathcal{D}}[C]}{\mathbb{E}_{\mathcal{D}}[O]}.\end{aligned}\quad (10)$$

Below is some explanation for the derivation:

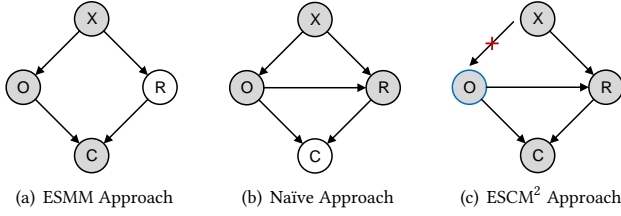
- (1) is the expectation of the random variable  $\hat{C}/\hat{O}$ .
- (2) holds because we have  $P(\hat{c}, \hat{o}) = P(\hat{c})P(\hat{o}|\hat{c})$ , and  $P(\hat{o}|\hat{c}) \geq P(\hat{o})$ . The equality holds only when  $\hat{O} \perp \hat{C}$ .
- (3) breaks down the integrals of the product into the product of the integrals, and get the expectations in (4).
- (5) holds because we have Jensen's inequality  $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$  for convex function  $f(X) = 1/X$ . The equality holds only when the variance of  $X$  equals to zero.
- (6) holds given a fully trained ESMM in Eq. (8).

As such, ESMM's CVR estimation is always larger than the ground-truth, *i.e.*,  $\text{Bias}^{\text{ESMM}} > 0$ . The proof is completed.  $\square$

Theorem 1 demonstrates that ESMM suffers from IEB even under relaxed conditions, which makes its CVR estimation sub-optimal.

It inspires us to propose a CVR estimator trained over the click space directly and address the sample selection bias problem directly, which will be formulated in the Section 4.

<sup>2</sup>The inequality stems from the fact that users who clicked are more likely to convert than those did not [17].



**Figure 3: Causal graphs depicting training of ESMM [14], Naive [25] and ESCM<sup>2</sup>, where X, O, R, C represent the user-item pair, click, conversion and click & conversion, respectively. Hollow circles indicate latent variables, and shaded circles indicate observed variables. The intervention on O, represented as a blue circle in (c), eliminates the dependence between X and O, and thus removes the arrow  $X \rightarrow O$ .**

### 3.2 Is ESMM a well-defined CTCVR Estimator?

The underlying causal graph of ESMM based on Eq. (4) and Eq. (5) is illustrated in Figure 3 (a). Specifically, ESMM uses two towers to predict CTR and CVR, and multiplies them to predict CTCVR:

$$\mathbb{P}(o_{u,i} = 1, r_{u,i} = 1) = \mathbb{P}(o_{u,i} = 1) * \mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1), \quad (11)$$

where the CVR estimate depends on the click event. Specifically, conversion only happens after the click event takes place. As such, there should be a causality link from  $O$  to  $R$ . However, in Figure 3 (a), the causality link is omitted by the entire space modeling strategy [14], as indicated by the missing arrow from  $O \rightarrow R$ . The missing arrow poses the risk that, for CTCVR task, ESMM models CVR as  $\mathbb{P}(r_{u,i} = 1)$  as shown in Figure 3 (a), rather than the desired  $\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1)$ . We formulate this risk as the potential independence priority problem since it replaces  $\mathbb{P}(r_{u,i} = 1 | o_{u,i} = 1)$  with  $\mathbb{P}(r_{u,i} = 1)$ , and thus introduces false independent prior.

The naïve approach [25] in Figure 3 (b) trains the CVR model within the click space, and thus there is an arrow from node  $O$  to node  $R$ <sup>3</sup>. However, the sample selection bias, indicated by the link  $X \rightarrow O$ , makes the naïve approach sub-optimal [14, 25]. Specifically, the dependency of node  $O$  on  $X$  makes the conversion feedback MNAR, which makes training and evaluation biased.

One approach to eliminate the selection bias is to model CVR as  $\mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1))$  instead. The "do" denotes the do-calculus, represented by the blue circle in the Figure 3 (c). For clicked samples, it is consistent with the conventional CVR definition, but for unclicked samples, it models a counterfactual problem: how likely is it that a user would be converted if he/she clicked the item? This method is called backdoor adjustment [9], which truncates the dependency from  $X$  to  $O$ . Finally, the adjusted CTCVR is

$$\mathbb{P}(o_{u,i} = 1, r_{u,i} = 1) = \mathbb{P}(o_{u,i} = 1) * \mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1)), \quad (12)$$

which addresses PIP and selection bias simultaneously. The obtained CTCVR estimate is thus valid for the entire inference space.

<sup>3</sup>This causal graph is consistent with Figure 1 by Gu et al. and Figure 1 (b) by Bareinboim and Pearl.

## 4 PROPOSED METHOD

### 4.1 Counterfactual Risk Regularizer

ESCM<sup>2</sup> leverages counterfactual risk minimizers based on inverse propensity score [17, 25] and doubly robust [8, 20] to regularize ESMM. We further illustrate that these regularizers mitigate both IEB and PIP in ESMM.

#### 4.1.1 Inverse Propensity Score Regularizer.

Following Eq. (6), IPS regularizer inversely weights the CVR loss with  $\hat{q}_{u,i}$ , where  $\hat{q}_{u,i}$  is usually modeled via a logistic CTR prediction model. As shown in Figure 4, we use the output of the CTR tower  $\hat{o}_{u,i}$  in ESMM to estimate the  $q_{u,i}$  in Eq. (6). As such, the regularizer [25] can be formulated as:

$$\begin{aligned} \mathcal{R}_{\text{IPS}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}) &= \mathbb{E}_{(u,i) \in \mathcal{D}} \left[ \frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i}(x_{u,i}; \phi_{\text{CVR}}))}{\hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}})} \right] \\ &= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i}(x_{u,i}; \phi_{\text{CVR}}))}{\hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}})}, \end{aligned} \quad (13)$$

where  $x_{u,i}$  denotes the concatenated embedding given the user-item pair, and  $\delta_{u,i}$  represents the estimation error of CVR.  $\phi_{\text{CTR}}$  and  $\phi_{\text{CVR}}$  are parameters of the CTR and the CVR tower in ESMM.

Inspired by Zhang et al., we first show that the IPS-based counterfactual risk regularizer mitigates inherent estimation bias (IEB) of CVR estimation in ESMM. Theorem 2 illustrates that  $\mathcal{R}_{\text{IPS}}$  converges to  $\mathcal{P}$  over the exposure space, which gives us unbiased CVR estimation and hence mitigate IEB.

**THEOREM 2.** *Let  $o_{u,i}$  be the indicator of whether user  $u$  clicks on item  $i$ ,  $\hat{o}_{u,i}$  be the predicted CTR by ESMM with IPS regularizer. Given that the predicted CTR gives an accurate estimate for the propensity score, i.e.,  $\hat{q}_{u,i} = q_{u,i}$ , the IPS-based counterfactual risk regularizer  $\mathcal{R}_{\text{IPS}}$  gives an unbiased CVR estimation:*

$$\mathbb{E}_{(u,i) \in \mathcal{D}} [\mathcal{R}_{\text{IPS}}] - \mathcal{P} = 0. \quad (14)$$

PROOF.

$$\begin{aligned} \mathbb{E}_{(u,i) \in \mathcal{D}} [\mathcal{R}_{\text{IPS}}] &= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \mathbb{E}_{\mathcal{D}} \left[ \frac{o_{u,i} \delta(r_{u,i}, \hat{r}_{u,i})}{\hat{q}_{u,i}} \right] \\ &= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \frac{q_{u,i} \delta(r_{u,i}, \hat{r}_{u,i})}{\hat{q}_{u,i}} \\ &= \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} \delta(r_{u,i}, \hat{r}_{u,i}) = \mathcal{P}. \end{aligned} \quad (15)$$

□

Also, we show that the IPS-based counterfactual risk regularizer models CVR as  $\mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1))$ , as shown in Eq. (12). It concurs with the causal dependency of the conversion event on the click event, and thus mitigates potential independence priority (PIP) of the CTCVR estimation in ESMM.

**THEOREM 3.** *Let  $\mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1))$  be the counterfactual conversion rate given the click event happens,  $\hat{r}_{u,i}^{\text{IPS}}$  be the predicted CVR of ESMM with IPS regularizer. For all user-item pairs in the exposure space, the IPS regularizer encourages:*

$$\hat{r}_{u,i}^{\text{IPS}} \rightarrow \mathbb{P}(r_{u,i} = 1 | do(o_{u,i} = 1)). \quad (16)$$

PROOF. Partition the user-item pairs into  $K$  boxes  $L_1, L_2, \dots, L_K$  such that conditional exchangeability holds in each box  $L_k$ . That is, within the group  $L_k$ , the observed distribution of click events  $O$  is independent of the counterfactual error of CVR estimation:

$$\{\delta_k^{(0)}, \delta_k^{(1)}\} \perp\!\!\!\perp O_k, \quad (17)$$

where  $\delta_k^{(1)}$  is the counterfactual distribution of CVR estimation error if  $o_{u,i} = 1$  for all  $(u, i) \in L_k$ ,  $\delta_k^{(0)}$  is the counterfactual distribution of error if  $o_{u,i} = 0$  for all  $(u, i) \in L_k$ .

Leveraging the law of iterated expectations:

$$\mathcal{R}_{\text{IPS}} = \mathbb{E}_{(u,i) \in \mathcal{D}} \left[ \frac{o_{u,i} \delta_{u,i}}{\hat{q}_{u,i}} \right] = \mathbb{E}_k \left\{ \mathbb{E}_{(u,i) \in L_k} \left[ \frac{o_{u,i}}{\hat{q}_{u,i}} \delta_{u,i} \right] \right\}, \quad (18)$$

and within the group  $L_k$ , we have:

$$\begin{aligned} \mathbb{E}_{(u,i) \in L_k} \left[ \frac{o_{u,i}}{\hat{q}_{u,i}} \delta_{u,i} \right] &= \mathbb{E}_{(u,i) \in L_k} \left[ \frac{o_{u,i}}{\hat{q}_{u,i}} \delta_{u,i}^{(1)} \right] \\ &= \mathbb{E}_{(u,i) \in L_k} \left[ \frac{o_{u,i}}{\hat{q}_{u,i}} \right] \mathbb{E}_{(u,i) \in L_k} [\delta_{u,i}^{(1)}] \\ &= \mathbb{E}_{(u,i) \in L_k} [\delta_{u,i}^{(1)}] \triangleq \Delta_k, \end{aligned} \quad (19)$$

where  $\Delta_k$  denotes the expectation of  $\delta_k^{(1)}$ . Therefore,  $\mathcal{R}_{\text{IPS}}$  can be reformulated as:

$$\mathcal{R}_{\text{IPS}} = \mathbb{E}_k [\Delta_k]. \quad (20)$$

As such, minimizing  $\mathcal{R}_{\text{IPS}}$  is equivalent to minimizing  $\Delta_k$  for the groups  $k = 1, 2, \dots, K$ . Particularly in the group  $L_k$ , minimizing  $\Delta_k$  makes the  $\hat{r}_{u,i}^{\text{IPS}}$  converge to the ground-truth CVR given click happens for all samples in this group.  $\square$

So far we have proved that the IPS regularizer is able to alleviate IEB and PIP in ESMM. However, it still suffers from high variance problem [17, 25] that makes the training process unstable. Therefore, we extend our scope to a doubly robust estimator, which gives us lower variance and double robustness [8, 20].

#### 4.1.2 Doubly Robust Regularizer.

Following Eq. (7), the DR regularizer inversely weights the error deviation  $\hat{e}_{u,i} = \delta_{u,i} - \hat{\delta}_{u,i}$ , instead of  $\delta_{u,i}$ , with  $\hat{o}_{u,i}$ . As shown in Figure 4, we utilize the output of the CTR tower  $\hat{o}_{u,i}$  in ESMM to estimate the propensity of click behavior, and the output of imputation tower  $\hat{\delta}_{u,i}$  to predict the estimation error of CVR  $\delta_{u,i}$ . As such, the DR regularizer can be formulated as:

$$\begin{aligned} \mathcal{R}_{\text{DR}}^{\text{err}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}, \phi_{\text{IMP}}) \\ = \mathbb{E}_{(u,i) \in \mathcal{D}} \left[ \hat{\delta}_{u,i}(x_{u,i}; \phi_{\text{IMP}}) + \frac{o_{u,i} \hat{e}_{u,i}(x_{u,i}; \phi_{\text{CVR}}, \phi_{\text{IMP}})}{\hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}})} \right], \end{aligned} \quad (21)$$

where  $x_{u,i}$  denotes the concatenated embedding given the user-item pair,  $\delta_{u,i}$  the estimation error of CVR,  $\phi_{\text{CTR}}$  and  $\phi_{\text{CVR}}$  the parameters of the CTR and the CVR tower in ESMM, and  $\phi_{\text{IMP}}$  the parameter of the imputation tower.

The double robustness derives from the fact that the estimator is unbiased if either  $\hat{\delta}_{u,i}$  or  $\hat{o}_{u,i}$  is accurate. The accuracy of  $\hat{o}_{u,i}$  is guaranteed by a properly trained CTR estimator. To ensure the

accuracy of  $\hat{\delta}_{u,i}$ , an auxiliary imputation loss is introduced:

$$\begin{aligned} \mathcal{R}_{\text{DR}}^{\text{imp}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}, \phi_{\text{IMP}}) \\ = \mathbb{E}_{(u,i) \in \mathcal{D}} \left[ \frac{o_{u,i} \hat{e}_{u,i}^2(x_{u,i}; \phi_{\text{CVR}}, \phi_{\text{IMP}})}{\hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}})} \right], \end{aligned} \quad (22)$$

and the doubly robust regularizer is formulated as:

$$\begin{aligned} \mathcal{R}_{\text{DR}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}, \phi_{\text{IMP}}) \\ = \mathcal{R}_{\text{DR}}^{\text{err}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}, \phi_{\text{IMP}}) + \mathcal{R}_{\text{DR}}^{\text{imp}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}, \phi_{\text{IMP}}). \end{aligned} \quad (23)$$

The DR regularizer also mitigates IEB and PIP in ESMM, since Theorems 2-3 also hold for the DR estimator. This proof is simple: we just need to replace the CVR estimation error  $\delta$  in IPS with the error deviation  $e$ . We omit the exact proof for brevity.

## 4.2 Architecture of ESCM<sup>2</sup>

As shown in Figure 4, we employ the multi-task learning technique to train ESCM<sup>2</sup>, which is effective in alleviating data sparsity [14, 16, 22]. Specifically, the learning process is composed of three tasks: the empirical risk minimizer for CTR estimation, the global risk minimizer for CTCVR estimation, and the counterfactual risk minimizer for CVR estimation. The empirical risk minimizer aims to minimize the empirical risk of CTR estimation over the entire impression space  $\mathcal{D}$ :

$$\mathcal{L}_{\text{CTR}}(\phi_{\text{CTR}}) = \mathbb{E}_{(u,i) \in \mathcal{D}} [\delta(o_{u,i}, \hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}}))], \quad (24)$$

where  $\delta$  is binary cross-entropy in our case. Afterwards, the global risk minimizer optimizes the risk of CTCVR estimation over  $\mathcal{D}$ :

$$\begin{aligned} \mathcal{L}_{\text{CTCVR}}(\phi_{\text{CTR}}, \phi_{\text{CVR}}) \\ = \mathbb{E}_{(u,i) \in \mathcal{D}} [\delta(o_{u,i} * r_{u,i}, \hat{o}_{u,i}(x_{u,i}; \phi_{\text{CTR}}) * \hat{r}_{u,i}(x_{u,i}; \phi_{\text{CVR}}))]. \end{aligned} \quad (25)$$

However, this approach suffers from inherent estimation bias (IEB) and potential independent priority (PIP). As such, we introduce the counterfactual risk minimizers in section 4.1 and derive the final learning objective of ESCM<sup>2</sup>:

$$\mathcal{L}_{\text{ESCM}^2} = \mathcal{L}_{\text{CTR}} + \lambda_c \mathcal{L}_{\text{CVR}} + \lambda_g \mathcal{L}_{\text{CTCVR}}, \quad (26)$$

where  $\lambda_c$  and  $\lambda_g$  are hyperparameters to control the weights of counterfactual risk and global risk, respectively.

Based on the formulation of  $\mathcal{L}_{\text{CVR}}$ , we devise two specific implementations of ESCM<sup>2</sup>, i.e., ESCM<sup>2</sup>-IPS and ESCM<sup>2</sup>-DR:

- ESCM<sup>2</sup>-IPS: It computes the counterfactual risk based on  $\mathcal{R}_{\text{IPS}}$  in Eq. (13). Specifically, it calculates the empirical risk of its CVR estimation over the click space, and inversely weights the risk with the output from its CTR tower.
- ESCM<sup>2</sup>-DR: It augments ESCM<sup>2</sup>-IPS with an imputation tower, and models the counterfactual risk of CVR estimation based on  $\mathcal{R}_{\text{DR}}$  in Eq. (21).

As shown in Figure (4), the architecture of ESCM<sup>2</sup> exploits the sequential pattern of user actions, i.e., *impression*  $\rightarrow$  *click*  $\rightarrow$  *conversion* in e-commerce scenario. The amount of training data for CTR task is significantly greater than that of CVR and CTCVR tasks by 1-2 order of magnitudes. Therefore, we share the feature representations learned in the CTR task with the CVR and the CTCVR tasks through a common look-up embedding table, which is effective in alleviating data sparsity.

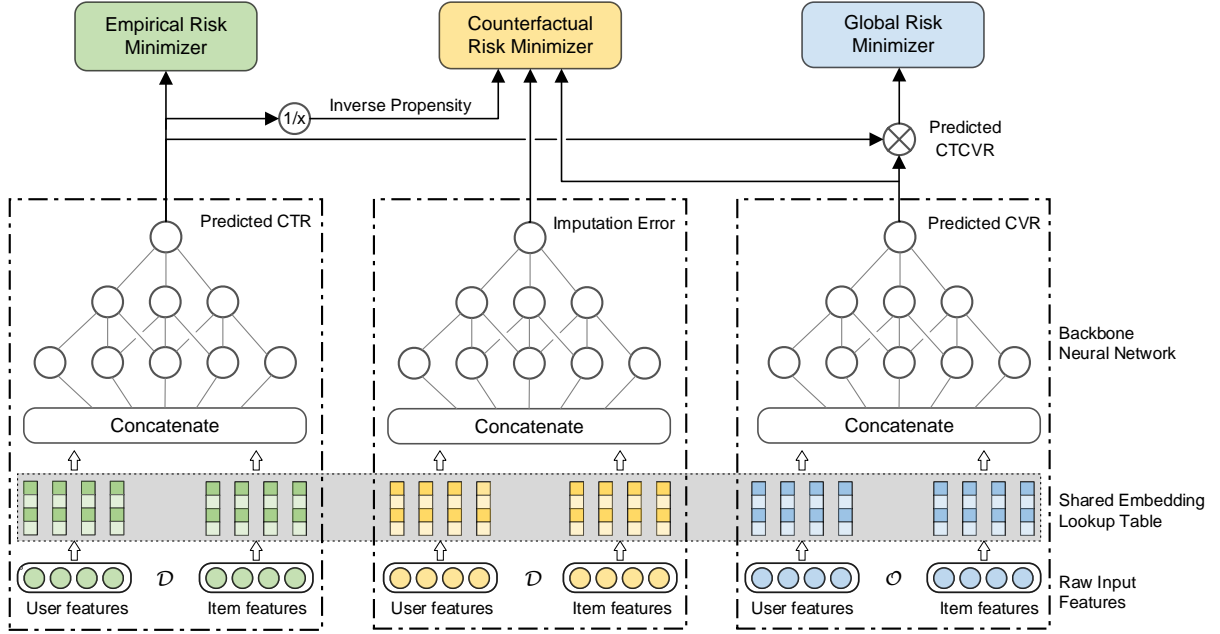


Figure 4: Architecture of ESCM<sup>2</sup>-IPS and ESCM<sup>2</sup>-DR, composed of the empirical risk minimizer for CTR estimation, the global risk minimizer for CTCVR estimation and the counterfactual risk minimizer for CVR estimation. All of the three optimizers share a common embedding lookup table. The ESCM<sup>2</sup>-DR augments the ESCM<sup>2</sup>-IPS with an imputation error tower.

## 5 EXPERIMENTS

We conduct experiments to evaluate the performance of ESCM<sup>2</sup> and answer the following research questions:

**RQ1:** Does ESCM<sup>2</sup> outperform SOTA CVR and CTCVR estimators?  
**RQ2:** Does ESMM inject bias on CVR estimation? To what extent does ESCM<sup>2</sup> reduce bias?

**RQ3:** Does ESMM inject independence prior when estimating CTCVR? To what extent does ESCM<sup>2</sup> alleviate this problem?

**RQ4:** How do critical components (*i.e.*,  $\lambda_c, \lambda_g$ ) affect the performance of ESCM<sup>2</sup>? Is ESCM<sup>2</sup> sensitive to these components?

### 5.1 Experimental Setup

#### 5.1.1 Datasets.

- **Industrial dataset:**<sup>4</sup> The industrial dataset comes from 90-day offline log of our business’s recommender system, divided for training, validation and test in a chronological order. Afterwards, we downsample the negative samples of the training set to keep the ratio of exposure:click:conversion to be 100:10:1, approximately.
- **Public dataset:**<sup>5</sup> The public dataset Ali-CCP (Alibaba Click and Conversion Prediction) is used to benchmark the performance of relevant methods for reproducibility consideration. Following [23], all single-valued categorical features are fed

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

Table 1: Dataset description.

into the models, and we randomly held out 10% of the training dataset as our validation dataset.

#### 5.1.2 Competitors.

Multi-task learning (MTL) has been shown to significantly improve recommender system’s performance in many previous works [16, 25]. For fairness consideration, methods based on single-task learning (*e.g.*, the vanilla Inverse Propensity Score [17] and Doubly Robust [20] approach) are not considered. There are two common methods which co-learn CTR and CVR simultaneously:

- **Naïve**<sup>6</sup> [13]: It directly models CVR as per Eq. (3), and shares embeddings across CTR and CVR tasks.
- **MTL-IMP**<sup>7</sup> [19]: It is implemented in the same way as the Naïve method, except that the unclicked samples are used directly as negative samples for CVR task.

The methods above have been known to induce selection bias for CVR estimation [14], and hence deteriorates our estimation for CTCVR. It is therefore necessary to include debiased approaches as our benchmark for fairness consideration.

<sup>4</sup>The desensitized and encrypted data set does not contain any Personal Identifiable Information (PII). Adequate data protection was carried out during experiment to prevent the risk of data copy leakage, and the data set was destroyed after the experiment. The data set does not represent any business situation, only used for academic research

<sup>5</sup><https://tianchi.aliyun.com/datalab/dataSet.html?dataId=408>

<sup>6</sup><https://github.com/drawbridge/keras-mmoe>

<sup>7</sup><https://github.com/shenweichen/DeepCTR/blob/>



- **ESMM**<sup>8</sup> [14, 22]: It leverages the multi-task learning to avoid the selection bias of CVR in an heuristic way.
- **MTL-EIB** [18]: The Error Imputation based approach leverages the imputed error for unclicked events and the prediction error for clicked events to get unbiased pcvr estimation.
- **MTL-IPS**<sup>9</sup> [25]: It implements the IPS estimator [17] via multi-task learning approach, which is theoretical unbiased.
- **MTL-DR**<sup>9</sup> [25]: It is a multi-task learning version of DR estimator [20], which is theoretical unbiased and more robust.

### 5.1.3 Training Protocol.

MMoE has been selected as the feature extractor for ESCM<sup>2</sup> and its competitors. It would also make sense to replace it with more powerful models such as AITM [23] and GemNN [3]. For fairness consideration, all models are trained for 300k iterations with Adam [10] optimizer and the same set of hyperparameters to make results comparable. We set the learning rate to be  $1e^{-4}$ , and the weight decay to be  $1e^{-3}$ . Other hyperparameters of the optimizer follow the literature [10]. Following [23], the embedding dimension is set to 5. The objective weights  $\lambda_g$  and  $\lambda_c$  are set to 1 and 0.1, respectively, based on the parameter study in Section 5.5. We checkpoint the performance over the validation set every 1k iterations and export the best model to evaluate its performance on the test set.

### 5.1.4 Evaluation Protocol.

For the offline experiments, following existing works, AUC (Area Under ROC) is primarily used as the main ranking metric to gauge performance. However, AUC only evaluates the average ranking performance of the model at all thresholds. To better understand how each model performs, we also report Kolmogorov-Smirnov (KS) score at the best threshold on ROC-curve and Recall and F1 score at the best threshold on PR-curve, respectively.

## 5.2 Performance Comparison

### 5.2.1 Offline results.

We compare the offline performance of ESCM<sup>2</sup> and its competitors on both the industrial dataset and the public dataset, with mean and standard deviation reported over ten runs with different random seeds. We summarize the performance of the CVR estimation in Table 2, and get the following observations<sup>10</sup>:

- Biased estimators achieve competitive performance on the CVR estimation task. Specifically, ESMM achieves an AUC of 0.754 on the industrial dataset, while the Naïve and the MTL-IMP estimators achieve AUCs of 0.751 and 0.756, respectively. Available empirical evidence [14, 25] only demonstrates that ESMM outperforms biased estimators on single-task learning. Therefore, ESMM's better performance in our case might be attributed to its multi-task learning paradigm.
- Unbiased baseline estimators at large generally outperform the biased ones across these two datasets. For example, MTL-IPS achieves best performance over two metrics on the industrial dataset, which improves the KS and AUC of ESMM by 1.04% and 0.39%, respectively. It is therefore promising to

regularize ESMM to get unbiased CVR estimation and better ranking performance.

- The proposed ESCM<sup>2</sup> achieves significant improvement compared with various state-of-the-art baselines. Combined with aforementioned comparisons, We attribute its performance to unbiasedness of its estimation due to the counterfactual risk minimizers and the efficiency of the entire-space methodology without IEB problem.

CTCVR is usually used as the primary ranking metric in industrial scenarios due to its empirical effectiveness. By eliminating both the IEB and the PIP problem, we expect better CTCVR estimation from ESCM<sup>2</sup> and hence better business metrics. We evaluate the performance of CTCVR estimation in Table 3 and obtain the following observations<sup>10</sup>:

- Across all multi-task learning baselines, ESMM showcases competitive performance in the CTCVR estimation task. Specifically, ESMM achieves the highest AUC and Recall on the industrial dataset and the highest F1 and Recall on the Ali-CCP dataset. We attribute this to the fact that ESMM incorporates CTCVR explicitly into its learning objective. Benefiting from its superiority on CTCVR estimation, ESMM is widely deployed in numerous business scenarios.
- Although less significantly than on the CVR estimation task, ESCM<sup>2</sup> still beats competitors over most metrics. Specifically, ESCM<sup>2</sup>-IPS achieves the best performance on F1 and Recall on Ali-CCP and ESCM<sup>2</sup>-DR achieves the best performance on AUC and KS on the industrial dataset. We attribute ESCM<sup>2</sup>'s better performance to the fact that it incorporates CTCVR into its learning objective directly and alleviates IEB and PIP through counterfactual risk regularization.

### 5.2.2 Online A/B results.

We conducted extensive online experiments to further showcase the superiority of ESCM<sup>2</sup> over ESMM. Specifically, we implemented ESMM and ESCM<sup>2</sup> with our internal C++ based deep learning framework, where ESCM<sup>2</sup> was built with the IPS regularizer for its competitive offline performance and training efficiency. We then randomly assigned users into several buckets, and observed each model's performance in respective bucket using our online A/B testing platform. Live Experiments were conducted in three large-scale scenarios. We are mainly interested in the models' performance on UV-CVR, UV-CTCVR and two additional business metrics: the number of orders and the actual premium.

**Scenario 1:** This experiment was deployed in our primary production environment. It lasted 6 days and covered 2.2 million unique visitors (UVs) and 3.1 million page views (PVs). Overall, ESCM<sup>2</sup> improved the number of orders by 2.84% and the premium by 10.85%. We also observed an significant 5.64% and 3.92% increase in UV-CVR and UV-CTCVR, respectively. We further report daily experimental results in Table 4, with ESCM<sup>2</sup> consistently outperforming ESMM on four metrics.

**Scenario 2:** This experiment was deployed in a newly revamped production environment for six days and covered 3.4 million UVs and 4.9 million PVs. Overall, ESCM<sup>2</sup> increased the number of orders by 4.26% and the premium by 3.88%. UV-CVR and UV-CTCVR also increased by 0.43% and 1.75% respectively.

<sup>8</sup><https://github.com/PaddlePaddle/PaddleRec/blob/master/models/multitask/esmm>

<sup>9</sup>The MTL-based implementation is not available. IPS and DR implementations locates <https://github.com/DongHande/AutoDebias/tree/main/baselines> for reference.

<sup>10</sup>To make the results more distinguishable, we report the F1 score in percentages.

Dataset Model	Industrial Dataset				Ali-CCP Dataset			
	AUC	KS	F1	Recall	AUC	KS	F1	Recall
Naive	0.7515±0.0164	0.3872±0.0043	0.3344±0.0052	0.5789±0.0071	0.5987±0.0139	0.1123±0.0056	0.0991±0.0053	0.2854±0.0050
ESMM	0.7547±0.0183	0.3856±0.0051	0.6330±0.0074	0.5742±0.0055	0.6071±0.0133	0.1267±0.0043	0.1157±0.0084	0.2968±0.0036
MTL-EIB	0.7272±0.0140	0.3371±0.0051	0.5808±0.0048	0.5121±0.0072	0.5603±0.0135	0.0717±0.0057	0.0825±0.0051	0.2372±0.0043
MTL-IMP	0.7563±0.0114	0.3974±0.0047	0.1272±0.0040	0.5841±0.0092	0.6114±0.0137	0.1163±0.0043	0.1135±0.0055	0.2962±0.0056
MTL-IPS	<u>0.7586±0.0112</u>	0.3960±0.0048	<u>0.6810±0.0042</u>	0.5651±0.0068	0.6091±0.0123	0.1177±0.0063	0.0941±0.2163	<u>0.2975±0.0070</u>
MTL-DR	0.7579±0.0135	0.4016±0.0046	0.6804±0.0042	0.5137±0.0081	0.6065±0.0172	0.1255±0.0141	0.1159±0.0084	0.2953±0.0178
ESCM <sup>2</sup> -IPS	<b>0.7730±0.0150</b>	<b>0.4144±0.0051*</b>	<b>0.7161±0.0089*</b>	0.5932±0.0094	<b>0.6163±0.0151</b>	0.1312±0.0060	0.1180±0.0047	0.3061±0.0059
ESCM <sup>2</sup> -DR	0.7679±0.0113	0.4119±0.0050	0.6884±0.0052	<b>0.5986±0.0068*</b>	0.6142±0.0133	<b>0.1393±0.0042*</b>	<b>0.1315±0.0053*</b>	<b>0.3095±0.0054*</b>

**Table 2: Offline performance (mean±std) on the CVR estimation task. Underlined results indicate the best baselines over each metric. "\*" marks the methods that improve the best baselines significantly at p-value < 0.01 over paired samples t-test.**

Dataset Model	Industrial Dataset				Ali-CCP Dataset			
	AUC	KS	F1	Recall	AUC	KS	F1	Recall
Naive	0.7954	0.4631	1.0048	0.6602	0.6003	0.1192	0.0978	0.2921
ESMM	0.8153	0.4827	1.1062	0.6819	0.6081	0.1292	0.1139	0.3027
MTL-EIB	0.7912	0.4220	0.8458	0.5975	0.5699	0.0697	0.0959	0.2542
MTL-IMP	0.7752	0.4126	<b>1.3393</b>	0.5880	0.6087	0.1264	0.1110	0.2973
MTL-IPS	0.8044	0.4840	1.0653	0.6716	0.6138	0.1302	0.1044	0.2911
MTL-DR	0.8106	0.4844	1.1707	0.6684	0.6130	0.1360	0.1096	0.2980
ESCM <sup>2</sup> -IPS	0.8198	0.4991	1.1753	0.6804	0.6189	0.1436	<b>0.1207*</b>	<b>0.3184*</b>
ESCM <sup>2</sup> -DR	<b>0.8265</b>	<b>0.5134*</b>	1.2842	<b>0.7013*</b>	<b>0.6245</b>	<b>0.1494</b>	0.1180	0.3117

**Table 3: Offline performance on the CTCVR estimation task. "\*" marks the methods that improve the best baselines significantly at p-value < 0.01 over paired samples t-test.**

Metrics	Day1	Day2	Day3	Day4	Day5	Day6
# Order	-20.59%	<b>+11.46%</b>	<b>+7.26%</b>	<b>+0.70%</b>	-9.52%	<b>+7.49%</b>
# Premium	<b>+64.53%</b>	<b>+37.47%</b>	<b>+22.09%</b>	-12.49%	<b>+4.26%</b>	<b>+11.10%</b>
UV-CVR	<b>+7.25%</b>	-1.66%	<b>+9.39%</b>	<b>+8.58%</b>	<b>+2.51%</b>	<b>+8.62%</b>
UV-CTCVR	<b>+0.20%</b>	-3.50%	<b>+2.50%</b>	<b>+9.48%</b>	<b>+2.75%</b>	<b>+6.64%</b>

**Table 4: Results of Online A/B test**

**Scenario 3:** This experiment was deployed in a production environment for a promotional campaign which covered 120 thousand UVs and 136 thousand PVs. Overall, ESCM<sup>2</sup> significantly increased the number of orders by 40.5% and the premium by 12.9%.

### 5.3 Discussion on Inherent Estimation Bias

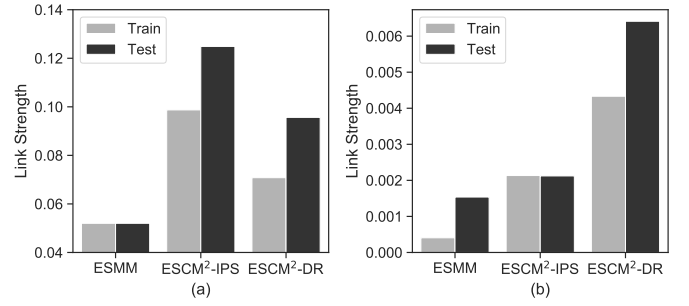
We provide experimental results to support our claim that ESMM's CVR estimation suffers from IEB. Specifically the "Label" field in Table 5 records the ground-truth conversion rate within the click space, which can be reasonably considered as an upper bound for the expected ground-truth CVR over the entire exposure space. Afterwards, we compare it with the mean of different estimators' CVR estimates over the exposure space.

The first observation from Table 5 is that the ESMM's CVR estimates are consistently greater than the ground truth. For example, on the Ali-CCP trainset, the ground-truth CVR expectation does not exceed 0.0056, while ESMM reaches an expectation of 0.01. On the test set, ESMM overestimates CVR by 0.0057 on average. It concurs with our theorem on IEB in Section 3.1.

The second observation is that counterfactual regularization significantly alleviates IEB, which supports our claims in Section 4.1. Specifically, on the industrial dataset, ESCM<sup>2</sup>-IPS reduces the error of expectation by 0.031 on the training set and 0.035 on the

Dataset	Subset	Label	ESMM	ESCM <sup>2</sup> -IPS	ESCM <sup>2</sup> -DR
Ali-CCP	Train	0.0055	0.0101±0.0011	0.0059±0.0005	0.0076±0.0018
Ali-CCP	Test	0.0056	0.0113±0.0008	0.0060±0.0009	0.0059±0.0009
Industry	Train	0.0953	0.1588±0.0107	0.1277±0.0053	0.1216±0.0093
Industry	Test	0.0407	0.1643±0.0095	0.1290±0.0092	0.1188±0.0022

**Table 5: Expectation of CVR estimates, where "Label" indicates the expectation of observed conversion labels.**



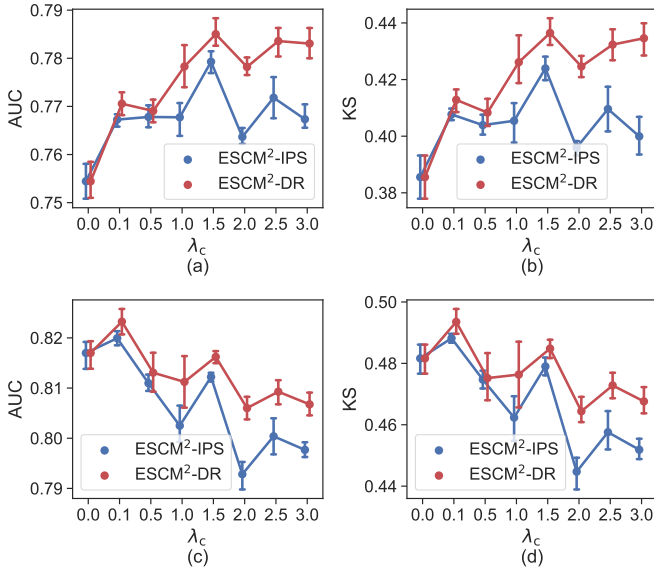
**Figure 5: The strength of the causal relationship between O and R, measured by the absolute error of the causal risk ratio to 1 on (a) industrial dataset and (b) Ali-CCP.**

test set; ESCM<sup>2</sup>-DR significantly reduces the error by 0.037 and 0.045 on the training and test sets, respectively.

### 5.4 Discussion on Potential Independence Priority

PIP stems from the model failing to capture the intrinsic causality from clicking to conversion, as indicated by the missing edge O→R in Figure 3 (a). We identify the strength of the causality from click to conversion to showcase the existence of PIP for ESMM's CTCVR estimation. Due to the presence of confounder in Figure 3, causality cannot be quantified using statistical estimators, such as the Pearson's correlation coefficient. Therefore, we preprocess data with propensity score matching (PSM) to eliminate confounding effect, following the settings in the tutorial by Garrido [5]. Specifically, we consider the CTR and CVR estimates as the propensity scores and outcomes, respectively. We then divide samples into two groups according to the click indicator O. Each clicked sample is matched to some unclicked samples with similar propensity scores.





**Figure 6: Parameter study of the counterfactual risk weight  $\lambda_c$  on (a-b): CVR estimation; (c-d): CTCVR estimation.**

PSM eliminates the confounding effects, which makes it possible to estimate causality effect through statistical estimators. Causal risk ratio (CRR) is an important metric to estimate the causality effect [9]. In short, the closer the CRR is to 1, the weaker the causality is, and vice versa. As such, we model the causality strength as the absolute error between CRR and 1.

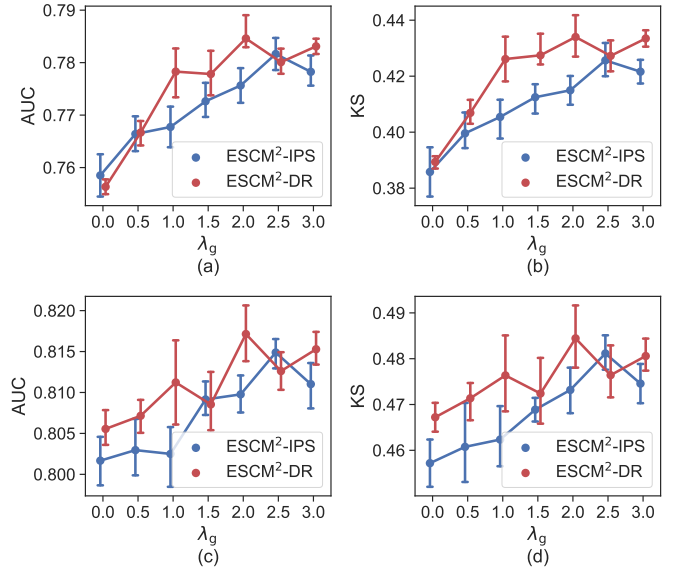
Figure 5 reports the causality strength from the CTR estimates to the CVR estimates. On both datasets, the strengths of ESM are negligible, which backs up our claim regarding PIP. ESCM<sup>2</sup> augments the link strength significantly by modeling the causal dependency directly, which supports our arguments in Theorem 3.

### 5.5 Parameter Sensitivity and Ablation Study

We discuss two critical hyperparameters in ESCM<sup>2</sup>, *i.e.*,  $\lambda_c$  and  $\lambda_g$  in Eq. (26), which are the weights in the learning objective and influence the final performance significantly.

In Figure 6, we vary  $\lambda_c$  in the range [0,3] to investigate the influence of causal regularization. Specifically, increasing  $\lambda_c$  consistently improves the ranking performance of CVR estimates. For example, the AUC of ESCM<sup>2</sup>-IPS increases from 0.755 at  $\lambda_c = 0$  to approximately 0.78 at  $\lambda_c = 1.5$ . Another observation is that causal regularization also benefits CTCVR estimates, with AUC increasing from 0.817 at  $\lambda_c = 0$  to 0.821 at  $\lambda_c = 0.1$ . However, assigning large weight to CVR risk is detrimental to CTCVR estimates. We speculate that placing a higher focus on CVR risk in a multitask learning framework leads to difficulties in CTR learning, which in turn leads to sub-optimal CTCVR estimation. It might also be due to high variance of IPS and sparsity of clicking data within batches. As such, we generally finetune  $\lambda_c$  within [0,0.1].

In Figure 7, we vary  $\lambda_g$  to investigate the influence of global risk minimizer. Specifically, within the range [0,3], increasing  $\lambda_g$  consistently improves both CVR and CTCVR estimation. For example, the



**Figure 7: Parameter study of the global risk weight  $\lambda_g$  on (a-b): CVR estimation; (c-d): CTCVR estimation.**

AUC of ESCM<sup>2</sup>-IPS increases from 0.758 at  $\lambda_g = 0$  to approximately 0.781 at  $\lambda_g = 2.5$ , and the KS of ESCM<sup>2</sup>-DR increases from 0.385 at  $\lambda_g = 0$  to approximately 0.434 at  $\lambda_g = 2.5$ .

## 6 RELATED WORK

Multi-task learning with task dependency is a challenging subject in recommender system because dependencies between tasks make feedback labels MNAR [17, 24]. Current works in this community can be broadly divided into two groups with respect to the methodology used to address the MNAR problem.

Methods in the first group leverage the sequential pattern of user behaviors to conduct probability decomposition and circumvent the MNAR problem. For example, ESM [14] models CTCVR as the product of CTR and CVR and learns CVR indirectly. ESM<sup>2</sup> [22] utilizes additional purchase-related post-click macro behaviors when conducting probability decomposition. GMCM [1] encodes user micro behaviors as graphs and utilizes graph convolutional networks to model interactions. HM<sup>3</sup> [21] models both micro and macro user behaviors explicitly in a unified deep learning framework.

Methods in the second group aim to build unbiased estimators on biased datasets directly. Schnabel et al. [17] proposes to conduct unbiased training and evaluation leveraging causality methodology, which inspires many subsequent research works. Wang et al. [20] analyzes the unbiasedness and effectiveness of doubly robust estimator in recommendation system, and Guo et al. [8] extends it to a more robust implementation. Zhang et al. [25] proposes to co-learn CTR, CVR and conducted debiasing within a multi-task learning framework. Recently, Lee et al. [11] proposes a dual learning framework that simultaneously eliminates the confounding effect in clicked and unclicked data.

## 7 CONCLUSION AND FUTURE WORK

Due to the effectiveness of modeling associations between tasks, ESMM dominates many large-scale business scenarios. However, it still suffers from inherent estimation bias for its CVR estimation and potential independence priority for its CTCVR estimation. A principled approach named ESCM<sup>2</sup> is devised to augment ESMM with counterfactual regularization. Extensive experiments demonstrate that ESCM<sup>2</sup> can largely mitigate IEB and PIP, and achieve better performance compared with other baseline models.

In e-commerce scenarios, inherent dependencies between multiple tasks are frequently observed [23]. Existing approaches [1, 22], in accordance with ESMM, leverage probabilistic decomposition to model dependency between tasks; meanwhile, they also introduce IEB and PIP inadvertently. A promising direction is to conduct cascaded counterfactual regularization over these models, which we leave for future work.

## REFERENCES

- [1] Wentian Bao, Hong Wen, Sha Li, Xiao-Yang Liu, Quan Lin, and Keping Yang. 2020. GMCM: Graph-based Micro-behavior Conversion Model for Post-click Conversion Rate Estimation. In *SIGIR*. 2201–2210.
- [2] Elias Bareinboim and Judea Pearl. 2012. Controlling selection bias in causal inference. In *Artificial Intelligence and Statistics*. 100–108.
- [3] Hongliang Fei, Jingyuan Zhang, Xingxuan Zhou, Junhao Zhao, Xinyang Qi, and Ping Li. 2021. GemNN: Gating-enhanced Multi-task Neural Networks with Feature Interaction Learning for CTR Prediction. In *SIGIR*. 2166–2171.
- [4] Chen Gao, Tzu-Heng Lin, Nian Li, Depeng Jin, and Yong Li. 2021. Cross-platform Item Recommendation for Online Social E-Commerce. *TKDE* (2021).
- [5] Garrido. 2014. Methods for constructing and assessing propensity scores. , 1701–1720 pages.
- [6] Tiangkai Gu, Kun Kuang, Hong Zhu, Jingjie Li, Zhenhua Dong, Wenjie Hu, Zhen-guo Li, Xiuqiang He, and Yue Liu. 2021. Estimating True Post-Click Conversion via Group-stratified Counterfactual Inference. In *ADKDD*.
- [7] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. In *IJCAI*. 1725–1731.
- [8] 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 *SIGIR*. 275–284.
- [9] Robins JM Hernán MA. 2020. *Causal Inference: What If*. oca Raton: Chapman Hall/CRC.
- [10] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.
- [11] Jae-woong Lee, Seongmin Park, and Jongwuk Lee. 2021. Dual Unbiased Recommender Learning for Implicit Feedback. In *SIGIR*. 1647–1651.
- [12] Tianqiao Liu, Zhiwei Wang, Jiliang Tang, Songfan Yang, Gale Yan Huang, and Zitao Liu. 2019. Recommender Systems with Heterogeneous Side Information. In *WWW*. 3027–3033.
- [13] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts. In *SIGKDD*. 1930–1939.
- [14] 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 *SIGIR*. 1137–1140.
- [15] Benjamin Marlin, Richard S Zemel, Sam Roweis, and Malcolm Slaney. 2012. Collaborative filtering and the missing at random assumption. *arXiv preprint arXiv:1206.5267* (2012).
- [16] Conor O'Brien, Kin Sum Liu, James Neufeld, Rafael Barreto, and Jonathan J. Hunt. 2021. An Analysis Of Entire Space Multi-Task Models For Post-Click Conversion Prediction. In *RecSys*. 613–619.
- [17] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *ICML*. 1670–1679.
- [18] Harald Steck. 2010. Training and testing of recommender systems on data missing not at random. In *SIGKDD*. 713–722.
- [19] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. 2020. Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. In *RecSys*. 269–278.
- [20] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2019. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *ICML*. 6638–6647.
- [21] Hong Wen, Jing Zhang, Fuyu Lv, Wentian Bao, Tianyi Wang, and Zulong Chen. 2021. *SIGIR*. 2187–2191.
- [22] Hong Wen, Jing Zhang, Yuan Wang, Fuyu Lv, Wentian Bao, Quan Lin, and Keping Yang. 2020. Entire Space Multi-Task Modeling via Post-Click Behavior Decomposition for Conversion Rate Prediction. In *SIGIR*. 2377–2386.
- [23] Dongbo Xi, Zhen Chen, Peng Yan, Yinger Zhang, Yongchun Zhu, Fuzhen Zhuang, and Yu Chen. 2021. Modeling the Sequential Dependence among Audience Multi-step Conversions with Multi-task Learning in Targeted Display Advertising. In *SIGKDD*. 3745–3755.
- [24] Mengyue Yang, Quanyu Dai, Zhenhua Dong, Xu Chen, Xiuqiang He, and Jun Wang. 2021. Top-N Recommendation with Counterfactual User Preference Simulation. In *CIKM*. 2342–2351.
- [25] 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 *WWW*. 2775–2781.
- [26] 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 *SIGKDD*. 1059–1068.