Debiased Recommendation via Wasserstein Causal Balancing

HAO WANG and ZHICHAO CHEN, Zhejiang University, Hangzhou, China
HONGLEI ZHANG, Beijing Jiaotong University, Beijing, China
ZHENGNAN LI, Communication University of China, Beijing, China
LICHENG PAN, Zhejiang University, Hangzhou, China
HAOXUAN LI, Center for Data Science, Peking University, Beijing, China
MINGMING GONG, The University of Melbourne, Melbourne, Australia

Recommendation systems are pivotal in improving user experience on various digital platforms. However, observational training data in recommendation systems introduce selection bias, which leads to a distributional discrepancy between training data and real-world scenarios, resulting in suboptimal performance. Current causal debiasing methods such as inverse propensity score and doubly robust rely on accurately estimated propensity scores, typically optimized through negative log-likelihood (NLL) minimization. However, recent studies have highlighted the limitations of this approach, as perfect NLL minimization may not adequately correct for selection bias. To address this issue, we propose Wasserstein Balancing Metric (WBM), a novel metric that measures and enhances the balancing capacity of propensity scores in causal debiasing methods by minimizing the Wasserstein discrepancy between reweighted populations. On the basis, we introduce IPS-WBM and DR-WBM, incorporating WBM as a regularizer in standard inverse propensity score and doubly robust estimators, which enhances causal balancing capacity without introducing additional bias. Extensive experiments on three real-world recommendation datasets demonstrate that our methods improve the causal balancing capability of learned propensities and enhance debiasing performance.

CCS Concepts: • **Information systems** → **Information retrieval**; • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Electronic commerce**;

Additional Key Words and Phrases: Recommender System, Causal Inference, Selection Bias, Entire Space, Selection Bias, Multi-Task Learning, Post-click Conversion Rate Estimation

ACM Reference format:
Hao Wang, Zhichao Chen, Honglei Zhang, Zhengnan Li, Licheng Pan, Haoxuan Li, and Mingming Gong. 2025. Debiased Recommendation via Wasserstein Causal Balancing. *ACM Trans. Inf. Syst.* 43, 6, Article 145 (September 2025), 24 pages.
https://doi.org/10.1145/3725731

This work was supported by National Natural Science Foundation of China (623B2002), ARC DE210101624, and ARC DP240102088.

Authors' Contact Information: Hao Wang, Zhejiang University, Hangzhou, China; e-mail: haohaow@zju.edu.cn; Zhichao Chen, Zhejiang University, Hangzhou, China; e-mail: 12032042@zju.edu.cn; Honglei Zhang, Beijing Jiaotong University, Beijing, China; e-mail: honglei.zhang@bjtu.edu.cn; Zhengnan Li, Communication University of China, Beijing, China; e-mail: lzhengnan389@gmail.com; Licheng Pan, Zhejiang University, Hangzhou, China; e-mail: 22132045@zju.edu.cn; Haoxuan Li (corresponding author), Center for Data Science, Peking University, Beijing, China; e-mail: hxli@stu.pku.edu.cn; Mingming Gong (corresponding author), The University of Melbourne, Melbourne, Australia; e-mail: mingming.gong@unimelb.edu.au.

ACM 1558-2868/2025/9-ART145
https://doi.org/10.1145/3725731

1 Introduction

Recommendation systems have become integral components of modern digital platforms. By analyzing user historical interactions and behaviors, personalized suggestions are generated to enhance user satisfaction and engagement. For instance, e-commerce giants like Alibaba [36, 53] utilize recommendation systems to suggest products that match a user's browsing history and purchase patterns, thereby increasing the conversion rate and transaction volume. Streaming services such as Kuaishou [15, 16] and TikTok [51] rely on content recommendation engines to curate movies, TV shows, and music playlists tailored to individual tastes, improving user immersion duration and video completion rate. In social media, platforms like X (formerly known as Twitter) [50] employ recommendation mechanisms to present content that resonates with user interests, enhancing user retention on the platform. The pivotal role of recommendation systems in driving user experience and business value underscores their significance in the digital era [21, 28, 64].

A unique challenge in recommendation systems is the reliance of training on observational data collected from user interactions, which introduces selection bias [8, 55]. Specifically, the data collected are missing not at random [67], a common yet challenging pattern of missing data [5, 56], which causes the observed data to under-represent the broader user population and all possible item interactions [4, 59]. For example, in rating prediction tasks, users are more likely to rate items they have a strong opinion about—either positively or negatively—leading to a dataset skewed toward extreme ratings and under-representing neutral or unrated items [25, 27]. In **click-through rate (CTR)** prediction, users are predominantly exposed to items that the current recommendation system deems relevant, causing future data to be biased toward items with historically higher clicks [39]. This selection bias manifests as a distributional discrepancy between the training data and the test scenario, where the objective is to provide online service to all users and items. Such discrepancies can adversely affect the generalization ability of the model, leading to suboptimal online service performance where recommendations do not align with true user preferences [63, 65, 67].

To mitigate the effects of selection bias, causal methods play a crucial role in debiased recommendation by constructing unbiased estimators of the learning objective using biased observational data [4, 46, 61]. The central idea of causal methods is to adjust the weights of the samples from the biased population to approximate the distribution of an ideal unbiased population. An important approach is the **inverse propensity score (IPS)** [43, 45], which reweights each sample by the inverse of its propensity score, the probability of the sample being treated under the current policy. By doing so, IPS aims to create a pseudo-population where the expectation of observed samples mirrors that of a randomized experiment. Another prominent method is the **doubly robust (DR)** estimator [11, 17, 29], which combines IPS with outcome modeling to correct for bias and reduce variance. DR leverages both the estimated propensity scores and imputed outcomes to achieve unbiased learning against misspecification in either component, providing more reliable estimates even when the estimated propensity scores are imperfect [24, 29].

The effectiveness of IPS and DR estimators critically relies on the accurate estimation of propensity scores [26]. Propensity scores represent the likelihood of observing or treating each sample under the current recommendation policy and are central to correcting selection bias. Current methodologies primarily optimize the propensity estimator by minimizing the **negative log-likelihood (NLL)** based on observed data [53, 67]. However, recent studies have doubted the rationale of this approach [26]: *a propensity estimator that minimizes NLL simplifies the IPS and DR to the naive average [53] and the Error Imputation-Based (EIB) estimator [37], respectively.* These simplified estimators fail to account for selection bias, leading to biased estimates of the learning objective. Therefore, relying solely on NLL minimization, despite its empirical efficacy in many practices [11, 45, 53, 67],

is insufficient to train propensity estimators in debiased recommendation. Consequently, there is a pressing need for more tailored metrics or learning objectives to enhance the balancing capability of learned propensity scores, ensuring the effectiveness of prevailing debiased estimators in recommendation systems [26, 30].

Recognizing that the primary role of propensity scores is to reweight biased samples toward a balanced representation of the population, it is intuitive to learn propensity scores by directly minimizing the discrepancy between the reweighted distributions. Building upon this insight, we introduce the **Wasserstein balancing metric (WBM)**, an innovative metric for measuring and enhancing the balancing capacity of propensity scores. WBM is calculated with the Wasserstein discrepancy [9], a mathematical tool from **optimal transport (OT)** theory, to quantify the discrepancy between the reweighted populations. Minimizing WBM during the training of propensity estimators can effectively enhance the capability of the learned propensity scores to balance biased populations. Moving forward, we propose the IPS-WBM and DR-WBM estimators, which integrate WBM as a regularization term into the standard IPS and DR estimators. This integration enhances the balancing capability of the propensity scores without introducing additional bias into the estimators, thereby achieving more effective debiasing. Finally, we validate the effectiveness of our proposed methods through comprehensive experiments on three real-world recommendation datasets.

Contributions. Our contributions can be summarized as follows.

- —We propose the WBM, a novel metric that refines NLL for measuring and enhancing the balancing capacity of propensity scores in recommendation systems.
- —We develop the IPS-WBM and DR-WBM estimators, incorporating WBM as a regularizer into the standard IPS and DR estimators, which improves the balancing capability of propensity scores without introducing additional bias. We provide unbiasedness conditions of the estimators based on the minimization of WBM.
- —We perform experiments on three real-world recommendation datasets. The results showcase that WBM effectively boosts the balancing capability of the learned propensities and significantly improves debiasing performance.

Organizations. The remaining sections are structured as follows: Section 2 encapsulates technical background to understand the methodology and contributions in this work; Section 3 formulates WBM and theoretically justifies its balancing property; On this basis, Section 4 constructs counterfactual estimators, formalizes the unbiasdness conditions, and delineates the computational workflow for debiased recommendation; Section 5 presents experiments to verify the efficacy of the proposed methods; Section 6 offers a brief literature review regarding debiased recommendation and propensity estimation; Section 7 discusses conclusions, limitations, and future works.

2 Preliminaries

In this section, we first formalize the debiased recommendation task in the potential outcome framework. Subsequently, we give a brief introduction to OT, a mathematical tool for quantifying distribution discrepancy. The important notations are summarized in Table 1.

2.1 Potential Outcomes Formalization of Recommendation

We denote by $\mathcal{U} = \{u\}$ the set of users and by $\mathcal{I} = \{i\}$ the set of items. In the context of a recommendation system, using the potential outcome framework necessitates the definition of several key elements. (1) *Covariate* $x_{u,i}$: the features of user u and item i, such as user profiles (age, gender, historical interests, etc.) and item characteristics (price, category, etc.); (2) *Treatment* $o_{u,i}$: the indicator noting whether the feedback $r_{u,i}$ is observed ($o_{u,i} = 1$) or missing ($o_{u,i} = 0$);

Table 1. List of Important Notations

Notation	Description
Notations for Debiased Recommendation	
$x_{u,i}$	The covariate indexed by user u and item i.
$o_{u,i}$	The treatment indicator indexed by user u and item i.
$r_{u,i}$	The outcome indexed by user u and item i.
$r_{u,i}^{(o)}$	The counterfactual outcome if we make a treatment o on user u and item i, with its estimate $\hat{r}_{u,i}$.
$q_{u,i}$	The propensity scores indexed by user u and item i, with its estimate $\hat{q}_{u,i}$.
$e_{u,i}$	The error of outcome prediction indexed by user u and item i, with its imputed value $\hat{e}_{u,i}$.
Notations for OT	
α, β	The treated and untreated populations.
α', β'	The treated and untreated populations after reweighting.
n, m	The numbers of samples in α and β, respectively.
$\mathbf{D}$	The pairwise Euclidean distance.
$\mathbf{P}$	The transport matrix.
$\mathbf{a}, \mathbf{b}$	The mass vector recording the mass of units in α and β.
$\mathbf{a}', \mathbf{b}'$	The mass vector recording the mass of units in α and β.
$\mathcal{W}$	The Wasserstein discrepancy.
Π	The constraint set of OT.
$\langle \cdot, \cdot \rangle$	The operation of inner product.
ϵ	The strength of entropic regularization.

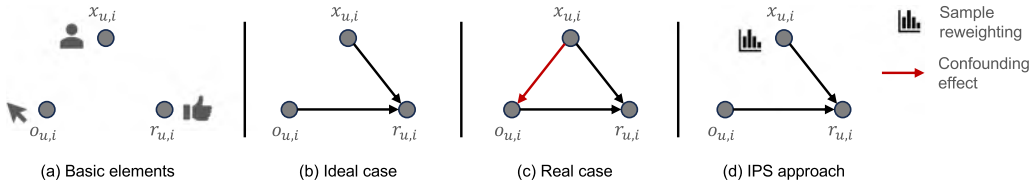


Fig. 1. Causal graphs in the context of debiased recommendation.

(3) *Outcome* $r_{u,i}$: the feedback from user u to item i, such as a rating, click, or conversion; (4) *Potential Outcome* $r_{u,i}^{(o)}$: the hypothetical feedback if $o_{u,i}$ is set to o; (5) *Ideal Population* $\mathcal{D} = \mathcal{U} \times \mathcal{I}$: the complete set of user-item pairs; (6) *Treated Population* $\mathcal{O}$: The subset of $\mathcal{D}$ where $o_{u,i} = 1$. Based on these elements, debiased recommendation can be represented as the causal graphs in Figure 1.

A widely adopted counterfactual query in debiased recommendation is: "What would the feedback be if an intervention were made on a user?" This translates to estimating the causal estimand $\mathbb{E}[r_{u,i}^{(1)} | x_{u,i}]$, which predicts the potential outcome using covariates. Such modeling can reinterpret recommendation tasks as causal problems as follows [59, 61]:

(1) *Rating Prediction*: Predict the rating $r_{u,i}^{(1)}$ if $o_{u,i}$ is forcibly set to 1 [28]. The treatment $o_{u,i} = \{0, 1\}$ indicates whether the user u rates the item i. The potential outcome $r_{u,i}^{(1)}$ denotes the true rating of the user u for the item i if we force $o_{u,i} = 1$ [25, 27, 28, 30].

(2) *Post-View CTR Prediction*: Estimate the CTR $r_{u,i}^{(1)}$ assuming exposure of the item i to the user u. The treatment $o_{u,i} = \{0, 1\}$ indicates whether the item i has been exposed to the user u. The potential outcome $r_{u,i}^{(1)}$ denotes the CTR if the item i is deliberately exposed to the user u [39, 71].

(3) *Post-Click Conversion Rate (CVR) Prediction*: Compute the CVR $r_{u,i}^{(1)}$ assuming that user u has clicked the item i. The treatment $o_{u,i} = \{0, 1\}$ indicates whether the item i has been clicked by the user. The potential outcome $r_{u,i}^{(1)}$ denotes the conversion rate assuming the user u has clicked the item i. Immediately, CVR can be represented as $\mathbb{E}[r_{u,i}^{(1)}|x_{u,i}]$ [49, 55, 68].

To estimate $r_{u,i}^{(1)}$ using $x_{u,i}$, we define the prediction model $\hat{r}_{u,i} = f(x_{u,i})$. The ideal learning objective, given by

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

where $e_{u,i}$ is an error measure like cross-entropy. However, the outcomes are only observed in the treated population ($\mathcal{O}$), which makes $\mathcal{L}_{\text{ideal}}$ incomputable. A naive but common shortcut is to estimate the learning objective over $\mathcal{O}$:

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

where $|\mathcal{D}|$ denotes the number of samples in the ideal population. Nonetheless, $r_{u,i}$ is missing not at random. For instance, in rating prediction, individuals are more inclined to rate items that capture their interest; similarly, in CTR prediction, users are more likely to be exposed to items they are predisposed to click. Therefore, there is a distribution discrepancy between $\mathcal{O}$ and $\mathcal{D}$, which stems from the confounding effects of $x_{u,i}$. Therefore, the naive approach based solely on the treated population $\mathcal{O}$ is biased compared to the ideal loss $\mathcal{L}_{\text{ideal}}$, resulting in suboptimal performance.

To counteract the confounding bias and construct an unbiased estimator of $\mathcal{L}_{\text{ideal}}$, the recommendation community has focused on causal-inspired techniques [3, 29, 53]. The central idea is to adjust the weights of samples in $\mathcal{O}$ to approximate the ideal population $\mathcal{D}$. A notable technique is the IPS method [46], which uses the propensity score to inversely weight observed events, defined by:

$$\mathcal{L}_{\text{IPS}} := \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}e_{u,i}}{q_{u,i}}\right], \tag{3}$$

where $q_{u,i}$ represents the propensity score. The IPS method assigns higher weights to less likely observed events to neutralize selection bias. Theoretically, $\mathcal{L}_{\text{IPS}}$ is an unbiased estimator of $\mathcal{L}_{\text{ideal}}$ if the propensity scores are accurately estimated. However, a significant defect with IPS is its dependency on precise propensity score estimation, which can be challenging to achieve in practice. To address this issue, the DR estimator [42, 46] introduces an error imputation technique, defined by:

$$\mathcal{L}_{\text{DR}} := \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{q_{u,i}}\right], \tag{4}$$

where $\hat{e}_{u,i}$ represents the imputed error and $(e_{u,i} - \hat{e}_{u,i})$ is the residual error after imputation. This design makes $\mathcal{L}_{\text{DR}}$ doubly robust, providing an unbiased estimate if either the imputed errors $\hat{e}_{u,i}$ or the propensity scores $q_{u,i}$ are accurately determined.

2.2 OT

OT is a mathematical framework designed to quantify the discrepancy between two distributions by identifying the minimum cost required to transform one distribution into the other. Originally proposed by Monge [38], the formulation involved finding an optimal mapping between two continuous distributions, which posed challenges related to the existence and uniqueness of solutions.

Addressing these issues, Kantorovich [18] proposed a more computationally feasible approach, defined as follows:

Definition 2.1 (Wasserstein Discrepancy). Consider empirical distributions $\alpha = \alpha_{1:n}$ and $\beta = \beta_{1:m}$, each with n and m samples, respectively. The Kantorovich problem seeks a feasible plan $\mathbf{P} \in \mathbb{R}_+^{n\times m}$ to transport α to β at the minimum possible cost:

$$\mathcal{W}(\alpha,\beta) := \min_{\mathbf{P}\in\Pi(\alpha,\beta)} \langle \mathbf{D}, \mathbf{P}\rangle,$$
$$\Pi(\alpha,\beta) := \begin{cases} \mathbf{P}_{i,1} + \ldots + \mathbf{P}_{i,m} = \mathbf{a}_i, i = 1, \ldots, n, \\ \mathbf{P}_{1,j} + \ldots + \mathbf{P}_{n,j} = \mathbf{b}_j, j = 1, \ldots, m, \\ \mathbf{P}_{i,j} \geq 0, i = 1, \ldots, n, j = 1, \ldots, m, \end{cases} \tag{5}$$

where $\mathcal{W}(\alpha,\beta)$ denotes the minimum transport cost, also known as the Wasserstein discrepancy [14, 41]; $\mathbf{D} \in \mathbb{R}_+^{n\times m}$ represents the pairwise distances calculated as $\mathbf{D}_{i,j} = |\alpha_i - \beta_j|_2^2$; $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ and $\mathbf{b} = [\mathbf{b}_1, \ldots, \mathbf{b}_m]$ are the masses of samples in α and β, respectively; Π defines the set of constraints.

The formulation above is a linear programming problem solvable via convex optimization techniques [12]. Moreover, the solution process can be accelerated through the Sinkhorn algorithm [2] which merely consists of tensor multiplications compatible with graph processing unit backends. The Wasserstein discrepancy has advantageous properties, which makes it versatile in diverse applications, such as sampling algorithm design [6, 32, 52], domain adaptation [9, 73], data imputation [57, 58], cross-domain recommendation [33–35], and causal inference [47, 54].

3 WBM for Propensity Score Estimation

3.1 The Central Role of Propensity Score in Causal Methods for Debiased Recommendations

Propensity score estimation plays a critical role in debiased recommendations, which refers to the probability of a sample being treated. The unbiasedness of the IPS-based estimators depends on the accuracy of the learned propensity scores. Therefore, to enhance the accuracy of propensity estimation, diverse techniques have been developed. Wang et al. [53] propose a multitask learning approach to train the propensity model and potential outcome prediction model simultaneously. Li et al. [26] propose a balancing penalty to facilitate the training of the propensity model. Li et al. [24] propose a DR learning approach that incorporates the estimated propensity scores to the training process of the imputed errors. Chen et al. [3] and Zheng [69] suggested incorporating a small unbiased dataset to enhance propensity estimation that is resilient to unobserved confounding effect.

Some might argue that the DR estimator is unbiased when the error imputation is accurate, regardless of the accuracy of the learned propensity scores. However, the training of the imputed errors relies heavily on accurately learned propensity scores [11, 26]. Specifically, the error imputation model $e_{u,i}$ is typically trained by minimizing

$$\mathcal{L}_{\text{imp}} = \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}(\hat{e}_{u,i} - e_{u,i})^2}{\hat{q}_{u,i}},$$

where $\hat{q}_{u,i}$ is the estimated propensity. Therefore, if the learned propensity scores are less accurate, the imputed errors are likely to be inaccurate, resulting in biased DR estimates and even bias amplification.

Given the widespread and important role of propensity scores in debiased recommendations, we aim to establish a unified propensity training standard. Importantly, there are several questions that need to be answered [26]. How to learn propensity that is more helpful for debiasing performance? Is

it plausible to merely pursue the accuracy of $o_{u,i}$ as accurately as possible? Which metric reasonably measures the quality of the learned propensity scores?

3.2 Analysis on Prevailing Propensity Score Estimators

In this section, we examine prevalent methods for propensity estimation and analyze their limitations for debiased recommendations. Early techniques estimate propensity based on item popularity [44, 46]. While these methods are easy to implement and understand, they frequently lead to biased recommendations due to their oversimplified assumptions.

More recent studies have shifted toward parametric models for propensity estimation [11, 53], optimized with the NLL defined as:

$$\mathcal{L}_p = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[-o_{u,i}\log(\hat{q}_{u,i}) - (1-o_{u,i})\log(1-\hat{q}_{u,i})\right],$$

where $\hat{q}_{u,i}$ is the estimated propensity. However, minimizing NLL does not necessarily guarantee improved debiasing performance [26]. For instance, an ideal estimator minimizing NLL (where $\hat{q}_{u,i}=1$ for treated and $\hat{q}_{u,i}=0$ for untreated samples) could reduce the IPS estimator to a naive estimator:

$$\mathcal{L}_{\text{naive}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{O}} e_{u,i}, \tag{6}$$

which merely averages the outcomes over the treated data, thereby producing biased estimates against $\mathcal{L}_{\text{ideal}}$ defined in the target population. Similarly, the ideal propensity estimator degrades the DR estimator to an EIB estimator [37, 67]:

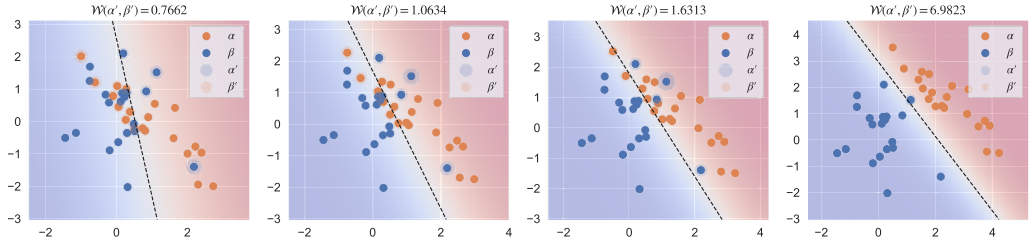
$$\mathcal{L}_{\text{eib}} = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} o_{u,i}e_{u,i} + (1-o_{u,i})\hat{e}_{u,i},$$

which loses the doubly robustness since its unbiasedness necessitates the unbiasedness of the error imputation model $\hat{e}_{u,i}$. Therefore, while parametric models represent an advance for propensity estimation, the minimization of NLL does not suffice to train propensity for debiased recommendation. It underlines the need for further research into more tailored metric to train propensity estimator that makes the reweighted distributions balanced almost surely.

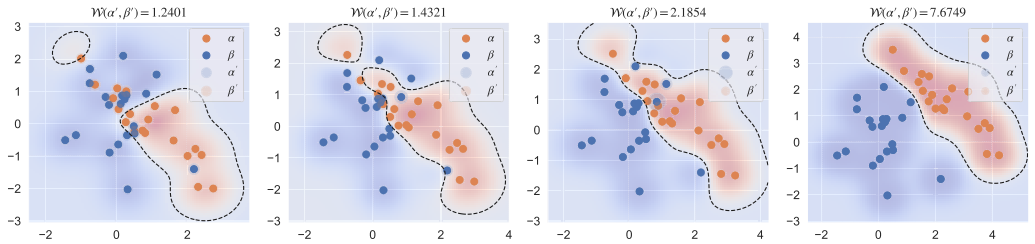
Case Study. To support the assertion that minimizing NLL alone is inadequate for debiased recommendation training, a case study was conducted in Figure 2. This toy dataset includes treated populations (α) and untreated populations (β), with increasing discrepancies across the panels from left to right. Reweighted sets α' and β' are created by inversely weighting α and β with $q_{u,i}$ and $1-q_{u,i}$, respectively. Key observations are summarized as follows.

— *Propensity Learned by Minimizing NLL Facilitates Balancing Distributions.* The estimated propensity tends to be lower for less likely treated samples. Therefore, the weighting mechanism in IPS assigns higher weights to less likely treated samples, which facilitates reducing selection bias and approximating the ideal population $\mathcal{D}$.

— *Enhanced Accuracy in Propensity Estimation Does Not Guarantee More Balanced Distributions Almost Surely.* As shown in Figure 2(a), despite increasing propensity estimation accuracy from 0.675 to 0.975, the discrepancy between reweighted populations rises from 0.78 to 6.98. Moreover, comparing the propensity models of logistic regression in Figure 2(a) and support vector machine in Figure 2(b), the support vector machine effectively enhances propensity estimation accuracy yet fails to enhance the balance of reweighted populations.



(a) Results with logistic regression, with the propensity estimation accuracy being 0.675, 0.75, 0.85 and 0.975 from left to right.



(b) Results with support vector machine (RBF kernel), with the propensity estimation accuracy being 0.775, 0.875, 0.925 and 1.0 from left to right.

Fig. 2. The balancing capability of the balancing scores calculated by propensity estimators: the logistic regression model in (a) and the support vector machine in (b). The dark area indicates the (uniform) weight of samples in α and β; the light area indicates the weight of samples in the reweighted sets α' and β'. The dashed black line indicates the decision boundary. $\mathcal{W}(\alpha',\beta')$ indicates the discrepancy between the reweighted sets.

3.3 WBM

The role of propensity score is to reweight biased populations toward a balanced state. This objective suggests that propensity score estimation could effectively be achieved by minimizing the discrepancy between the reweighted populations. To quantify distribution discrepancies, Wasserstein discrepancy provides a natural and effective approach, which has been widely employed in various contexts such as vision [9], natural language processing [62], and generative models [7, 40]. The Wasserstein discrepancy can be calculated by computing the OT problem in Definition 2.1.

Definition 3.1 (WBM). Consider empirical distributions $\alpha = \alpha_{1:n}$ and $\beta = \beta_{1:m}$, $\mathbf{q} \in \mathbb{R}^{n+m}$ represents the propensity score under optimization. WBM is defined as the Wasserstein discrepancy between the reweighted populations α' and β':

$$\begin{aligned} \mathrm{WBM}_\phi(\mathbf{q}) &:= \min_{\mathbf{P}\in\Pi(\alpha',\beta')} \left\langle \mathbf{D}^{(\phi)}, \mathbf{P} \right\rangle, \\ \Pi(\alpha',\beta') &:= \left\{ \mathbf{P}\mathbf{1}_m = \mathbf{a}', \quad \mathbf{P}^\top \mathbf{1}_n = \mathbf{b}', \quad \mathbf{P} \geq 0 \right\}, \end{aligned} \tag{7}$$

where $\phi(\cdot)$ denotes the representation function of a sample, $\mathbf{D}^{(\phi)} \in \mathbb{R}_+^{n\times m}$ represents the pairwise distances calculated as $\mathbf{D}^{(\phi)}_{i,j} = \|\phi(\alpha_i) - \phi(\beta_j)\|_2^2$, $\mathbf{a}' = \mathrm{Softmax}([1/q_1, \ldots, 1/q_n])$ and $\mathbf{b}' = \mathrm{Softmax}([1/(1-q_1), \ldots, 1/(1-q_m)])$ are the reweighted mass vectors of the samples; Softmax is used to normalize the mass vectors.

On this basis, we introduce the WBM, a novel metric designed to enhance the balancing capability of the propensity score estimates. Formally defined in Definition 3.1, WBM quantifies the Wasserstein discrepancy between the reweighted populations, α' and β', which are obtained by applying the IPSs to α and β. This metric directly evaluates how well the propensity scores balance the two populations. Furthermore, in Theorem 1, we theoretically demonstrate that the WBM

reaches zero if and only if the populations weighted by learned propensity scores are equivalent, which reflects the efficacy of the WBM to enhance the balancing property of the learned propensity scores.

Theorem 1 (Balancing Property). *The distributions reweighted by propensity scores* $\mathbf{q}$ *are equivalent, i.e.,* $\alpha' = \beta'$ *if and only if the WBM is minimized to zero, i.e.,* $\mathrm{WBM}_\phi(\mathbf{q}) = 0$.

Proof. Let α' and β' be the reweighted populations in Definition 3.1. To establish the theorem, it suffices to prove: (1) $\alpha' = \beta' \Rightarrow \mathcal{W}(\alpha', \beta') = 0$; (2) $\mathcal{W}(\alpha', \beta') = 0 \Rightarrow \alpha' = \beta'$. Without loss of generality, we consider the case $n = m$ for simplicity; the argument extends similarly when $n \neq m$ by appropriately padding with zeros.

First, we prove $\alpha' = \beta' \Rightarrow \mathcal{W}(\alpha', \beta') = 0$. Assume that $\alpha' = \beta'$. In this case, we can construct a transport plan $\mathbf{P}$ where each $\mathbf{P}_{i,j}$ is zero for all $i \neq j$, and $\mathbf{P}_{i,i} = \mathbf{a}'_i$ for all $i = 1, 2, ..., n$. This transport plan naturally satisfies the marginal constraints since for each i and j:

$$\sum_{j=1}^{m} \mathbf{P}_{i,j} = \mathbf{P}_{i,i} = \mathbf{a}'_i, \quad i = 1, 2, ..., n,$$
$$\sum_{i=1}^{n} \mathbf{P}_{i,j} = \mathbf{P}_{j,j} = \mathbf{b}'_j = \mathbf{a}_j, \quad j = 1, 2, ..., m.$$

Evaluating the transportation cost under this plan yields

$$\langle \mathbf{D}, \mathbf{P} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{D}_{i,j} \mathbf{P}_{i,j} = \sum_{i=1}^{n} \mathbf{D}_{i,i} \mathbf{a}'_i = 0,$$

which immediately follows from the fact that transporting mass to itself incurs no cost, i.e., $\mathbf{D}_{i,i} = 0$ for all $i = 1, 2, ..., n$.

Furthermore, since $\mathcal{W}$ is defined by the minimum over all feasible plan, we have

$$\mathcal{W}(\alpha', \beta') = \min_{\mathbf{P} \in \Pi(\alpha', \beta')} \langle \mathbf{D}, \mathbf{P} \rangle \leq \langle \mathbf{D}, \mathbf{P} \rangle = 0.$$

Given that both $\mathbf{D}$ and $\mathbf{P}$ are non-negative, we have $\mathcal{W}(\alpha', \beta') \geq 0$. Therefore, we have $\mathcal{W}(\alpha', \beta') = 0$ if $\alpha' = \beta'$.

Second, we prove $\mathcal{W}(\alpha', \beta') = 0 \Rightarrow \alpha' = \beta'$. Assume now that $\mathcal{W}(\alpha', \beta') = 0$. By definition, there exists a feasible transport plan $\mathbf{P}^* \in \Pi(\alpha', \beta')$ such that $\langle \mathbf{D}, \mathbf{P}^* \rangle = 0$. Given that both $\mathbf{D}$ and $\mathbf{P}$ are non-negative, it immediately follows that for all i, j, either $\mathbf{D}_{i,j} = 0$ or $\mathbf{P}^*_{i,j} = 0$. Since $\mathbf{D}$ represents the transportation cost and assuming that $\mathbf{D}_{i,j} > 0$ for $i \neq j$, it immediately follows that $\mathbf{P}^*_{i,j} = 0$ whenever $i \neq j$. Consequently, the transport plan $\mathbf{P}^*$ must be diagonal; that is, $\mathbf{P}^*_{i,j} = 0$ for all $i \neq j^*$.

Since $\mathbf{P}^*$ is a valid transport plan, it must satisfy the marginal constraints:

$$\begin{aligned} \sum_{j=1}^{m} \mathbf{P}_{i,j} &= \mathbf{P}_{i,i} = \mathbf{a}'_i, \quad i = 1, 2, ..., n, \\ \sum_{i=1}^{n} \mathbf{P}_{i,j} &= \mathbf{P}_{j,j} = \mathbf{b}'_j, \quad j = 1, 2, ..., m. \end{aligned} \tag{8}$$

Given that $\mathbf{P}^*$ is diagonal, these constraints simplify to $\mathbf{a}'_i = \mathbf{b}'_i$ for all i. Therefore, the distributions α' and β' must be identical. This completes the proof. □

Compared with other distributional discrepancy measures, Wasserstein discrepancy offers distinct advantages in terms of numerical stability and interpretability. For instance, when compared to Kullback-Leibler divergence, Wasserstein discrepancy remains applicable given distributions with disjoint supports; when compared to maximum mean discrepancy, Wasserstein discrepancy

Algorithm 1: The Computational Procedure of WBM

Input: $\mathbf{D}^{(\phi)}$: the pair-wise distance matrix, $\mathbf{q}$: the propensity scores to optimize, $\mathbf{o}$ the treatment indicators.
Parameter: ϵ: the strength of entropic regularization; τ: the convergence threshold; $\ell_{\max}$: the maximum iterations.
Output: $\mathrm{WBM}_{\phi}(\mathbf{q})$: the WBM metric of the propensity scores.

$\mathbf{a}', \mathbf{b}' \leftarrow \mathrm{Initialize}(\mathbf{q}, \mathbf{o})$
$\mathbf{K} \leftarrow \exp(-\mathbf{D}^{(\phi)}/\epsilon)$
$\ell \leftarrow 1$, $\mathbf{u}^{\ell} \leftarrow \mathbf{1}_n$, $\mathbf{v}^{\ell} \leftarrow \mathbf{1}_m$
while $\ell < \ell_{\max}$ **do**
  $\ell \leftarrow \ell + 1$
  $\mathbf{u}^{\ell} \leftarrow \mathbf{a}'/(\mathbf{K}\mathbf{v}^{\ell-1})$
  $\mathbf{v}^{\ell} \leftarrow \mathbf{b}'/(\mathbf{K}^{\mathrm{T}}\mathbf{u}^{\ell})$
  if $\|\mathbf{u}^{\ell} - \mathbf{u}^{\ell-1}\|_2^2 + \|\mathbf{v}^{\ell} - \mathbf{v}^{\ell-1}\|_2^2 < \tau$ **then**
    Break
$\mathbf{P}^* \leftarrow \mathrm{diag}(\mathbf{u}^{\ell})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{\ell})$
$\mathrm{WBM}_{\phi}(\mathbf{q}) \leftarrow \left\langle \mathbf{D}^{(\phi)}, \mathbf{P}^* \right\rangle$

offers a more interpretable method for quantifying discrepancies by considering the "cost" of transporting mass from one distribution to another. These advantages motivate our use of Wasserstein discrepancy in this work.

3.4 Computation of WBM

In this section, we discuss the computation of the WBM in Definition 3.1. The WBM formulation involves solving a linear programming problem, which can be addressed using the simplex algorithm [1, 12]. However, the simplex algorithm introduces non-differentiability, which prevents the direct optimization of propensity estimators. To overcome this limitation, we employ the Sinkhorn algorithm [2, 10, 13], which makes the WBM differentiable with respect to the propensity scores. The detailed steps of this process are outlined in Algorithm 1 and are justified as follows.

The initial step involves setting up the mass vectors as $\mathbf{a}' = \mathrm{Softmax}([1/q_1, \ldots, 1/q_n])$ and $\mathbf{b}' = \mathrm{Softmax}([1/(1-q_1), \ldots, 1/(1-q_m)])$ (step 1). The softmax operation ensures that the mass vectors sum to 1, transforming them into valid probability vectors, which are necessary for defining the OT problem. Afterward, the WBM formulation is refined by introducing an entropic regularizer following Cuturi [10]:

$$\mathbf{P}^* := \arg\min_{\mathbf{P} \in \Pi(\alpha', \beta')} \left\langle \mathbf{D}^{(\phi)}, \mathbf{P} \right\rangle - \epsilon \mathrm{H}(\mathbf{P}), \tag{9}$$

where $\mathrm{H}(\mathbf{P}) := -\sum_{i,j} \mathbf{P}_{ij}(\log(\mathbf{P}_{ij}) - 1)$ is the entropy term, and ϵ controls the strength of the entropic regularization. This regularization makes the problem ϵ-convex and computationally tractable using the Sinkhorn algorithm. To solve this, define $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$ as the Lagrangian multipliers. The Lagrangian of the problem is:

$$\Phi(\mathbf{P}, \mathbf{f}, \mathbf{g}) = \langle \mathbf{D}^{(\phi)}, \mathbf{P} \rangle - \epsilon \mathrm{H}(\mathbf{P}) - \langle \mathbf{f}, \mathbf{P}\mathbf{1}_n - \mathbf{a} \rangle - \left\langle \mathbf{g}, \mathbf{P}^{\mathrm{T}}\mathbf{1}_m - \mathbf{b} \right\rangle.$$

According to the first-order condition of constrained optimization problem, the OT matrix $\mathbf{P}^*$ satisfies:

$$\mathbf{P}^*_{ij} = \exp\left(\frac{\mathbf{f}_i}{\epsilon}\right) * \exp\left(-\frac{\mathbf{D}^{(\phi)}{}_{ij}}{\epsilon}\right) * \exp\left(\frac{\mathbf{g}_j}{\epsilon}\right), \tag{10}$$

which immediately follows from $\frac{\partial \Phi(\mathbf{P},\mathbf{f},\mathbf{g})}{\partial \mathbf{P}_{ij}}\Big|_{\mathbf{P}=\mathbf{P}*} = \mathbf{D}^{(\phi)}{}_{ij} + \varepsilon \log\left(\mathbf{P}^*_{ij}\right) - \mathbf{f}_i - \mathbf{g}_j = 0$. For convenience, we introduce the variables $\mathbf{u}_i := \exp\left(\frac{\mathbf{f}_i}{\epsilon}\right)$, $\mathbf{v}j := \exp\left(\frac{\mathbf{g}j}{\epsilon}\right)$, and $\mathbf{K}ij := \exp\left(-\frac{\mathbf{D}^{(\phi)}{}ij}{\epsilon}\right)$. The OT matrix

can be written as:

$$\mathbf{P}^* = \operatorname{diag}(\mathbf{u})\mathbf{K}\operatorname{diag}(\mathbf{v}).$$

Meanwhile, the transport matrix must satisfy the mass-preserving constraints:

$$\mathbf{P}^*\mathbf{1}_m = \operatorname{diag}(\mathbf{u})\mathbf{K}\operatorname{diag}(\mathbf{v})\mathbf{1}_m = \mathbf{a}, \quad \mathbf{P}^{\top}\mathbf{1}_n = \operatorname{diag}(\mathbf{v})\mathbf{K}\operatorname{diag}(\mathbf{u})\mathbf{1}_n = \mathbf{b},$$

or equivalently, let $\odot$ be the entry-wise multiplication of vectors, we have:

$$\mathbf{u} \odot (\mathbf{K}\mathbf{v}) = \mathbf{a} \quad \text{and} \quad \mathbf{v} \odot \left(\mathbf{K}^{\mathrm{T}}\mathbf{u}\right) = \mathbf{b},$$

which is known as the matrix scaling problem. An intuitive approach to solving these constraints is through iterative updates, as outlined in steps 3–7 of Algorithm 1:

$$\mathbf{u}^{\ell+1} = \frac{\mathbf{a}}{\mathbf{K}\mathbf{v}^{\ell}} \quad \text{and} \quad \mathbf{v}^{\ell+1} = \frac{\mathbf{b}}{\mathbf{K}^{\mathrm{T}}\mathbf{u}^{\ell+1}}. \tag{11}$$

These updates are repeated until the convergence criterion is met. Once convergence is achieved, the OT matrix $\mathbf{P}^*$ is computed using (10). The WBM is then evaluated as the inner product of $\mathbf{P}^*$ and $\mathbf{D}^{(\phi)}$ (steps 10 and 11 in Algorithm 1). Since $\mathbf{P}^*$ is derived through tensor multiplication involving the mass vectors $\mathbf{a}'$ and $\mathbf{b}'$, the resulting WBM is differentiable with respect to the propensity score estimate $\mathbf{q}$. This property allows for the optimization of the propensity score estimator using gradient-descent-based approaches.

4 WBM-Enhanced Debiased Recommendation

4.1 WBM-Enhanced Counterfactual Estimators

The WBM is designed to assess and improve the balancing capability of learned propensity scores, rather than directly training an unbiased prediction model. In this section, we introduce the IPS-WBM and DR-WBM estimators, which incorporate WBM as a regularizer to enhance the standard IPS and DR estimators. This integration enhances the balancing capability of propensity scores without introducing additional bias. The IPS-WBM estimator is defined as

$$\mathcal{L}_{\text{IPS-WBM}} = \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}e_{u,i}}{\kappa \cdot \hat{q}_{u,i}}\right] + \lambda_{\text{b}} \cdot \text{WBM}_{\phi}(\hat{\mathbf{q}}), \tag{12}$$

where $\kappa = |\mathcal{O}|/|\mathcal{D}|$ is a constant incorporated for deriving unbiasedness, λ_{b} is the weight of WBM. Similarly, the DR-WBM estimator is defined as

$$\mathcal{L}_{\text{DR-WBM}} = \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i} - \hat{e}_{u,i})}{\kappa \cdot \hat{q}_{u,i}}\right] + \lambda_{\text{b}} \cdot \text{WBM}_{\phi}(\hat{\mathbf{q}}), \tag{13}$$

which extends $\mathcal{L}_{\text{IPS-WBM}}$ by incorporating an imputation arm that aims to accurately estimate the error $e_{u,i}$ in $\mathcal{D}$. The output of this arm, denoted $\hat{e}_{u,i}$, is corrected by $e_{u,i} - \hat{e}_{u,i}$ in $\mathcal{O}$, where the actual $e_{u,i}$ is observed. The accuracy of $\hat{e}_{u,i}$ is improved through an auxiliary learning task defined as:

$$\mathcal{L}_{\text{imp}} = \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}\hat{e}_{u,i}^2}{\hat{\mathbf{q}}_{u,i}}\right]. \tag{14}$$

Well-optimized propensity scores ensure the unbiasedness of IPS-WBM and DR-WBM. According to Theorem 2, both proposed estimators are unbiased if the WBM is effectively minimized to zero. Therefore, incorporating WBM does not introduce bias to IPS and DR estimators and can enhance the balancing property of the learned propensity scores.

Theorem 2 (Unbiasedness Property). *Given the WBM is minimized, i.e.,* $\mathrm{WBM}_\phi(\hat{\mathbf{q}}) = 0$, *we have:*

(a) $\mathcal{L}_{\mathrm{IPS-WBM}}$ *is an unbiased estimator of* $\mathcal{L}_{\mathrm{ideal}}$*;*

(b) $\mathcal{L}_{\mathrm{DR-WBM}}$ *is an unbiased estimator of* $\mathcal{L}_{\mathrm{ideal}}$*, whether the imputed errors are accurate or not.*

Proof. Suppose $\hat{q}_{u,i}$ is the learned balancing score that makes $\mathrm{WBM}_\phi(\hat{\mathbf{q}}) = 0$. That is, the wasserstein distance between the treated population $\mathcal{O}$ inversely weighted with $\hat{p}$ and the ideal population is zero. For $\mathcal{L}_{\mathrm{IPS-WBM}}$, we have

$$\begin{aligned}
\mathbb{E}_{(u,i)\in\mathcal{D}}[\mathcal{L}_{\mathrm{IPS-WBM}}] &= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}e_{u,i}}{\kappa\cdot\hat{q}_{u,i}}\right] + \lambda_{\mathrm{b}}\cdot\mathrm{WBM}_\phi(\hat{\mathbf{q}}) \\
&= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}e_{u,i}}{\kappa\cdot\hat{q}_{u,i}}\right] && (15a)\\
&= \mathbb{E}_{(u,i)\in\mathcal{O}}\left[\frac{e_{u,i}}{\hat{q}_{u,i}}\right] && (15b)\\
&= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[e_{u,i}\right] = \mathcal{L}_{\mathrm{ideal}}, && (15c)
\end{aligned}$$

where (15a) follows immediately from the balancing penalty $\mathrm{WBM}_\phi(\hat{\mathbf{q}}) = 0$, and the estimator degrades to the canonical IPS estimator; (15b) follows from $\kappa = |\mathcal{O}|/|\mathcal{D}|$ and $o_{u,i} = 1$ only for $(u,i) \in \mathcal{O}$; (15c) holds since the expectation on the treated population $\mathcal{O}$ that is inversely weighted by the balancing score $\hat{q}$ is equivalent to that on the target population $\mathcal{D}$, given the balancing score satisfies $\mathrm{WBM}_\phi(\hat{\mathbf{q}}) = 0$.

Similarly, if the balancing property holds, the $\mathcal{L}_{\mathrm{DR-WBM}}$ degrades to the canonical DR estimator, and the expectation over $\mathcal{O}$ approximates the expectation over $\mathcal{D}$ almost surely:

$$\begin{aligned}
\mathbb{E}_{(u,i)\in\mathcal{D}}[\mathcal{L}_{\mathrm{DR-WBM}}] &= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\kappa\cdot\hat{q}_{u,i}}\right] + \lambda_{\mathrm{b}}\cdot\mathrm{WBM}_\phi(\hat{\mathbf{q}}) \\
&= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i} + \frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\kappa\cdot\hat{q}_{u,i}}\right] && (16a)\\
&= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i}\right] + \mathbb{E}_{(u,i)\in\mathcal{O}}\left[\frac{e_{u,i}-\hat{e}_{u,i}}{\hat{q}_{u,i}}\right] && (16b)\\
&= \mathbb{E}_{(u,i)\in\mathcal{D}}\left[e_{u,i}\right] = \mathcal{L}_{\mathrm{ideal}}. && (16c)
\end{aligned}$$

□

4.2 Overall Workflow

In this section, we detail the procedure for employing the proposed estimators to achieve debiased recommendation. The steps are formally outlined in Algorithm 2. The process begins with retrieving the embeddings of user u and item i from the embedding table and concatenating them. It follows by acquiring estimates for the treatment ($\hat{q}_{u,i}$), potential outcome ($\hat{r}_{u,i}$), and imputation error ($\hat{e}_{u,i}$) through parametric models, denoted as f_{prop}, f_{pred}, and f_{imp} in steps 1–4. We then compute the actual error $e_{u,i}$ within the treated population $\mathcal{O}$ in step 5.

The specification of the prediction error, denoted as $\mathcal{L}_{\mathrm{pred}}$, depends on the selected counterfactual estimator, to mitigate the selection bias and achieve the unbiased estimation of the ideal learning objective $\mathcal{L}_{\mathrm{ideal}}$:

— For the IPS estimator, $\mathcal{L}_{\mathrm{pred}}$ is defined according to (12). Importantly, the gradient of $e_{u,i}$ with respect to $\hat{q}_{u,i}$ is stopped to prevent unintended impacts of $e_{u,i}$ on the learning of $\hat{q}_{u,i}$.

Algorithm 2: The Computational Procedure for IPS-WBM and DR-WBM

Input: $(u, i) \in \mathcal{D}$: the user-item pairs; $o_{u,i}$: the treatment label; $r_{u,i}$: the outcome label for $(u, i) \in \mathcal{O}$.
Parameter: λ_b: the weight of WBM.
Output: $\mathcal{L}_{\text{pred}}$: the learning objective.

$x_{u,i} \leftarrow \text{Embedding}(u, i)$.
$\hat{q}_{u,i} \leftarrow f_{\text{prop}}(x_{u,i})$.
$\hat{r}_{u,i} \leftarrow f_{\text{pred}}(x_{u,i})$.
$\hat{e}_{u,i} \leftarrow f_{\text{imp}}(x_{u,i})$.
$e_{u,i} \leftarrow -r_{u,i} \log \hat{r}_{u,i} - (1 - r_{u,i}) \log(1 - \hat{r}_{u,i})$.
$\tilde{q}_{u,i} \leftarrow \text{StopGradient}(\hat{q}_{u,i})$.
if model is IPS-WBM **then**
    $\mathcal{L}_{\text{pred}} \leftarrow \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\tilde{q}_{u,i}} + \lambda_b \cdot \text{WBM}_\phi(\hat{\mathbf{q}})$.
else if model is DR-WBM **then**
    $\hat{e}_{u,i} \leftarrow e_{u,i} - \hat{e}_{u,i}$.
    $\mathcal{L}_{\text{DR}} \leftarrow \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \hat{e}_{u,i} + \frac{o_{u,i} \hat{e}_{u,i}}{\tilde{q}_{u,i}} + \lambda_b \cdot \text{WBM}_\phi(\hat{\mathbf{q}})$.
    $\mathcal{L}_{\text{imp}} \leftarrow \mathbb{E}_{(u,i)\in\mathcal{D}} \left[\frac{o_{u,i} \hat{e}^2_{u,i}}{\tilde{q}_{u,i}} \right]$
    $\mathcal{L}_{\text{pred}} \leftarrow \mathcal{L}_{\text{DR}} + \mathcal{L}_{\text{imp}}$.

Table 2. Description of Employed Datasets

Dataset	# User	# Item	# Interaction	Density	Domain	Threshold
Music	14,877	1,000	15,400	0.812%	Music	4
Coat	290	300	6,960	5.333%	Clothing	4
KuaiRec	1,411	3,327	4,676,570	0.996%	Video	2

—For the DR estimator, $\mathcal{L}_{\text{pred}}$ is defined according to (13). This approach includes refining the accuracy of the imputation model, encapsulated by the imputation loss $\mathcal{L}_{\text{imp}}$.

5 Experiments

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

—*Performance. How does WBM perform compared to the prevalent unbiased estimators in recommendation?* Section 5.2 compares IPS-WBM and DR-WBM against state-of-the-art debiased recommendation baselines using public datasets.
—*Balance. Does WBM enhance the capability of balancing populations?* Section 5.3 offers a case study to compare the balancing capability of NLL and WBM.
—*Efficacy. Does WBM enhance the performance of debiased recommendation?* Section 5.4 incorporates a comparative study to showcase the advantage of WBM over NLL with varying weights.
—*Sensitivity. Is WBM sensitive to hyperparameters?* Section 5.5 presents a sensitivity analysis of the key hyperparameters to provide a comprehensive understanding of the proposed method.
—*Complexity. Is WBM computationally intensive?* Section 5.6 compares the computational complexity given different experimental settings.

5.1 Setup

5.1.1 Dataset. There are three datasets involved in this study, and their details are summarized in Table 2 and detailed as follows.

— *Music* [46] contains 311,704 biased ratings for training, which involve 15,400 users and 1,000 items. Additionally, 5,400 users rate 10 randomly selected items, yielding 54,000 unbiased ratings for evaluation. The user-item pairs with ratings greater than 4 are seen as positive, and others are viewed as negative.
— *Coat* [46] is a public dataset which consists of 290 users and 300 items; each user subjectively selects 24 items to rate based on their preference, yielding 6,960 biased ratings in the training set. Additionally, each user rates 16 items that are randomly selected, yielding 4,640 unbiased ratings for model evaluation. The user-item pairs with rating greater than 4 are seen as positive, and others are viewed as negative.
— *KuaiRec* [15] is a public large-scale industrial dataset, which consists of 4,676,570 video watching ratio records from 1,411 users for 3,327 videos. The user-item pairs with ratings less than two are viewed as negative, and those with ratings of 2 or higher are viewed as positive.

5.1.2 Baselines. We take the **matrix factorization (MF)** [20] as the base model and compare the proposed IPS-WBM and DR-WBM with the following training methodologies:

— *Naïve* [20, 36] calculates the prediction loss in the treated population, using the naive learning objective in (6).
— *Multi-IMP* [36] mirrors the Naïve approach but includes untreated samples as negative samples for outcome estimator training.
— *ESMM* [36] employs a multitask approach which implicitly optimizes the outcome estimator.
— *IPS* and *DR* [46] calculate the prediction error using the standard IPS and DR estimator.
— *AS-IPS* [45] incorporates a clipping strategy to DR to balance bias and variance.
— *Multi-EIB* [36] imputes the prediction error in $\mathcal{D}$ and corrects its imputation in $\mathcal{O}$.
— *Multi-IPS* and *Multi-DR* [67] enhance IPS and DR by learning the propensity in a multitask learning manner.
— *ESCM*^{2}*-IPS* and *ESCM*^{2}*-DR* [53] fuse IPS and DR with the ESMM learning paradigm.
— *SDR* [29] bounds the bias and variance of DR under small propensity scores.
— *DR-JL* [60] enhances DR by jointly learning the prediction model and the error imputation model.
— *MRDR-JL* [17] reduces the variance of DR while retaining its double robustness.
— *DR-MSE* [11] enables to balance the bias and variance of DR flexibly for better generalization performance.

5.1.3 Training Protocol. We employ MF as the propensity model. All experimental procedures are executed using the PyTorch framework, utilizing the Adam optimizer [19] for its adaptive learning rate capabilities and efficient convergence properties. The experiments are conducted on a hardware platform comprising two Intel Xeon Platinum 8383C CPUs operating at 2.70 GHz and an NVIDIA GeForce RTX 4090 GPU. Hyperparameter optimization is systematically performed following the standard protocol [25, 27] to enhance model performance. The balancing weight λ_b is set to 1 for overall performance comparison and investigated specially in Section 5.4. The representation function ϕ is acquired by concatenating the user and item embeddings. The learning rate is tuned within {0.001, 0.005, 0.01, 0.05, 0.1}; the batch size is tuned within {32, 64, 128, 256} for the Coat dataset and {1,024, 2,048, 4,096} for the Music and KuaiRec datasets; the embedding size in the MF model is tuned over {2, 4, 8, 16, 32, 64} for Coat and {16, 32, 64, 128, 256, 512} for Music and KuaiRec. For baselines where the results are available in related works [27, 55], we use the reported results.

5.1.4 Evaluation Protocol. We primarily use the **area under the receiver operating characteristic curve (AUC)**, which measures the model's ability to distinguish between positive and negative

Table 3. Comparative Study on Debiased Recommendation on Real-World Datasets

Method	COAT			MUSIC			KuaiRec		
	AUC	NDCG@5	F1@5	AUC	NDCG@5	F1@5	AUC	NDCG@50	F1@50
Naive	$0.680_{\pm 0.006}$	$0.616_{\pm 0.011}$	$0.470_{\pm 0.006}$	$0.651_{\pm 0.005}$	$0.626_{\pm 0.001}$	$0.300_{\pm 0.001}$	$0.741_{\pm 0.003}$	$0.724_{\pm 0.003}$	$0.566_{\pm 0.002}$
ESMM	$0.686_{\pm 0.004}$	$0.638_{\pm 0.005}$	$0.485_{\pm 0.008}$	$0.601_{\pm 0.002}$	$0.665_{\pm 0.001}$	$0.328_{\pm 0.001}$	$0.721_{\pm 0.006}$	$0.764_{\pm 0.003}$	$0.576_{\pm 0.004}$
Multi-EIB	$0.604_{\pm 0.012}$	$0.540_{\pm 0.007}$	$0.419_{\pm 0.008}$	$0.664_{\pm 0.002}$	$0.637_{\pm 0.003}$	$0.319_{\pm 0.001}$	$0.660_{\pm 0.002}$	$0.624_{\pm 0.003}$	$0.529_{\pm 0.002}$
Multi-IMP	$0.713_{\pm 0.003}$	$0.613_{\pm 0.011}$	$0.462_{\pm 0.004}$	$0.627_{\pm 0.003}$	$0.661_{\pm 0.003}$	$0.328_{\pm 0.002}$	$0.735_{\pm 0.006}$	$0.719_{\pm 0.007}$	$0.573_{\pm 0.006}$
IPS	$0.710_{\pm 0.003}$	$0.603_{\pm 0.009}$	$0.450_{\pm 0.008}$	$0.656_{\pm 0.002}$	$0.633_{\pm 0.002}$	$0.308_{\pm 0.002}$	$0.750_{\pm 0.003}$	$0.734_{\pm 0.003}$	$0.572_{\pm 0.002}$
Multi-IPS	$0.711_{\pm 0.005}$	$0.604_{\pm 0.008}$	$0.463_{\pm 0.009}$	$0.651_{\pm 0.002}$	$\underline{0.667}_{\pm 0.001}$	$\underline{0.331}_{\pm 0.002}$	$0.748_{\pm 0.003}$	$0.738_{\pm 0.008}$	$0.579_{\pm 0.003}$
CVIB	$0.718_{\pm 0.004}$	$0.640_{\pm 0.007}$	$0.486_{\pm 0.008}$	$\underline{0.685}_{\pm 0.001}$	$0.645_{\pm 0.003}$	$0.315_{\pm 0.001}$	$0.758_{\pm 0.001}$	$0.752_{\pm 0.001}$	$0.575_{\pm 0.001}$
AS-IPS	$0.712_{\pm 0.008}$	$0.627_{\pm 0.010}$	$0.470_{\pm 0.007}$	$0.661_{\pm 0.003}$	$0.641_{\pm 0.004}$	$0.322_{\pm 0.003}$	$0.746_{\pm 0.009}$	$0.733_{\pm 0.004}$	$0.585_{\pm 0.006}$
ESCM^2-IPS	$\underline{0.721}_{\pm 0.005}$	$\underline{0.645}_{\pm 0.009}$	$\mathbf{0.490}_{\pm 0.005}$	$0.653_{\pm 0.003}$	$0.653_{\pm 0.002}$	$0.322_{\pm 0.002}$	$\underline{0.779}_{\pm 0.001}$	$\underline{0.767}_{\pm 0.003}$	$\underline{0.592}_{\pm 0.002}$
IPS-WBM	$\mathbf{0.740}^{*}_{\pm 0.001}$	$\mathbf{0.686}^{*}_{\pm 0.002}$	$\underline{0.481}_{\pm 0.006}$	$\mathbf{0.705}^{*}_{\pm 0.002}$	$\mathbf{0.686}^{*}_{\pm 0.002}$	$\mathbf{0.339}^{*}_{\pm 0.001}$	$\mathbf{0.783}_{\pm 0.011}$	$\mathbf{0.792}^{*}_{\pm 0.008}$	$\mathbf{0.600}^{*}_{\pm 0.006}$
DR	$0.710_{\pm 0.006}$	$0.632_{\pm 0.003}$	$0.471_{\pm 0.008}$	$0.656_{\pm 0.009}$	$0.669_{\pm 0.007}$	$0.330_{\pm 0.005}$	$0.745_{\pm 0.004}$	$0.718_{\pm 0.003}$	$0.574_{\pm 0.003}$
Multi-DR	$0.719_{\pm 0.006}$	$0.634_{\pm 0.009}$	$0.480_{\pm 0.007}$	$0.686_{\pm 0.001}$	$0.660_{\pm 0.003}$	$0.323_{\pm 0.002}$	$0.752_{\pm 0.001}$	$0.767_{\pm 0.012}$	$0.581_{\pm 0.003}$
DR-JL	$0.714_{\pm 0.007}$	$\underline{0.646}_{\pm 0.009}$	$0.486_{\pm 0.006}$	$0.682_{\pm 0.001}$	$0.660_{\pm 0.002}$	$0.326_{\pm 0.001}$	$0.759_{\pm 0.002}$	$0.757_{\pm 0.004}$	$0.582_{\pm 0.005}$
MRDR-JL	$0.715_{\pm 0.004}$	$0.653_{\pm 0.006}$	$\underline{0.492}_{\pm 0.005}$	$0.684_{\pm 0.001}$	$0.660_{\pm 0.003}$	$0.326_{\pm 0.002}$	$0.762_{\pm 0.003}$	$0.751_{\pm 0.002}$	$0.579_{\pm 0.003}$
DR-MSE	$0.715_{\pm 0.001}$	$0.630_{\pm 0.009}$	$0.480_{\pm 0.006}$	$0.685_{\pm 0.001}$	$0.648_{\pm 0.004}$	$0.316_{\pm 0.002}$	$0.779_{\pm 0.003}$	$0.773_{\pm 0.004}$	$0.589_{\pm 0.001}$
ESCM^2-DR	$\underline{0.730}_{\pm 0.009}$	$0.642_{\pm 0.010}$	$0.489_{\pm 0.010}$	$\underline{0.688}_{\pm 0.002}$	$\underline{0.669}_{\pm 0.002}$	$\underline{0.326}_{\pm 0.002}$	$\underline{0.788}_{\pm 0.001}$	$\underline{0.796}_{\pm 0.004}$	$\underline{0.606}_{\pm 0.002}$
SDR	$0.719_{\pm 0.006}$	$0.631_{\pm 0.008}$	$0.475_{\pm 0.006}$	$0.687_{\pm 0.001}$	$0.650_{\pm 0.001}$	$0.316_{\pm 0.001}$	$0.764_{\pm 0.003}$	$0.791_{\pm 0.003}$	$0.595_{\pm 0.002}$
DR-WBM	$\mathbf{0.744}^{*}_{\pm 0.001}$	$\mathbf{0.697}^{*}_{\pm 0.002}$	$\mathbf{0.514}^{*}_{\pm 0.002}$	$\mathbf{0.713}^{*}_{\pm 0.000}$	$\mathbf{0.691}^{*}_{\pm 0.001}$	$\mathbf{0.343}^{*}_{\pm 0.001}$	$\mathbf{0.802}^{*}_{\pm 0.002}$	$\mathbf{0.808}^{*}_{\pm 0.001}$	$\mathbf{0.611}^{*}_{\pm 0.000}$

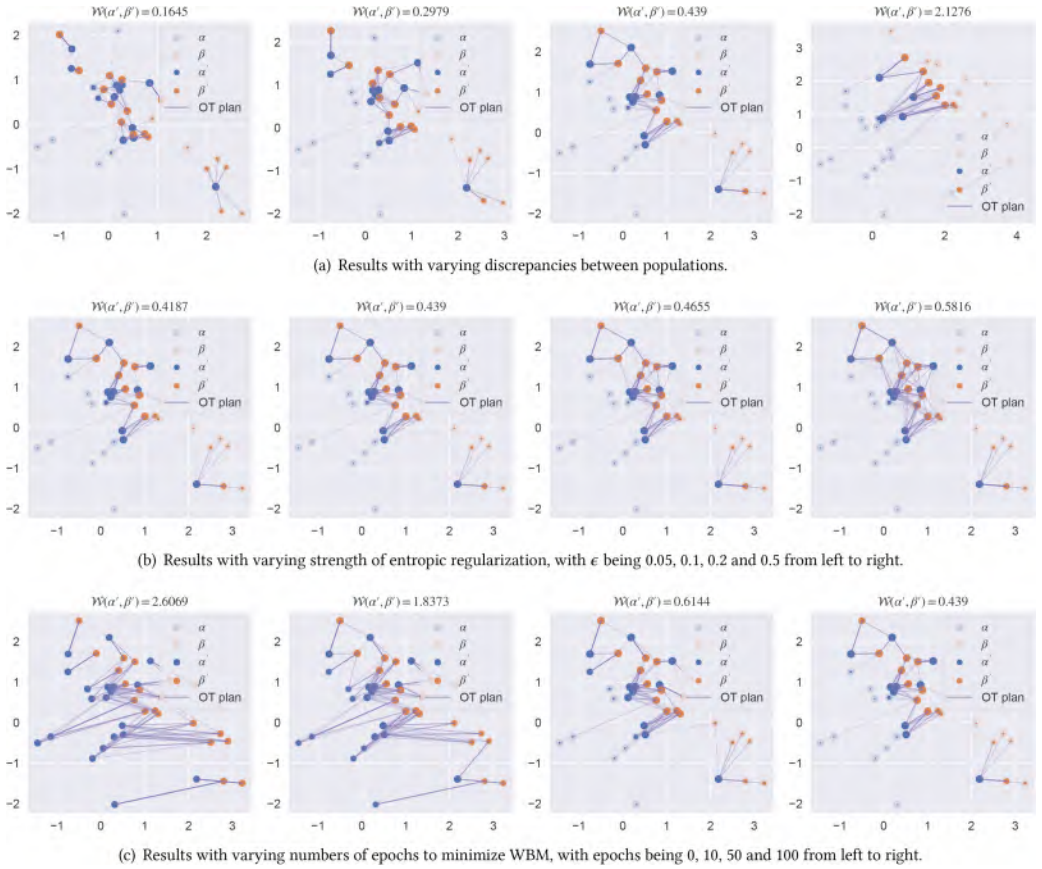
The bold and underlined fonts indicate the best and second-best performance for IPS and DR methods, respectively. "*" indicates the metrics where IPS-WBM (DR-WBM) outperforms the best baselines based on IPS (DR), with p-value < 0.05 under two-sample t-test.

classes, to assess the ranking performance of the models. To further compare performance in top-k recommendation, we introduce **normalized discounted cumulative gain at k (NDCG@k)**, which evaluates the quality of the ranking by considering the positions of relevant items, and F1@k, which balances precision and recall within the top-k recommendations, as supplementary metrics. Here, k is set to 5 for COAT and MUSIC and 50 for KuaiRec follow Li et al. [27].

5.2 Overall Performance

The debiased recommendation performance of IPS-WBM and DR-WBM is evaluated against competing models in Table 3. Notable findings are summarized as follows:

- —The biased approaches exhibit practical performance across datasets, with ESMM exhibiting the best overall performance among these approaches. On the Coat dataset, for instance, ESMM achieves a 0.88% higher AUC and 3.19% higher F1@5 over the Naïve method. In comparison to Multi-IMP, which also incorporates untreated samples as negative samples, ESMM performs better in top-k metrics. The efficacy of ESMM can be attributed to its multitask learning nature, as well as its implicit modeling of potential outcome which bypasses the selection bias issue.
- —Most debiased baselines significantly outperform biased methods. For instance, on the COAT dataset, the unbiased method ESCM^2-IPS records an AUC of 0.721, a 5.10% improvement over ESMM's AUC of 0.686. This trend is even more pronounced for top-k metrics, where ESCM^2-IPS achieves NDCG@5 = 0.645 and an F1@5 = 0.490, significantly surpassing those of biased methods. The superior performance of unbiased methods across AUC and top-k metrics underscores the importance of debiased learning in real-world recommendation tasks.
- —DR estimators outperform their IPS counterparts, demonstrating the benefits of error imputation techniques. For instance, ESCM^2-DR on the COAT dataset achieves an AUC of 0.730, a 1.23% improvement over ESCM^2-IPS's AUC of 0.721. These improvements are also evident in top-k metrics; for example, on the KuaiRec dataset, Multi-DR achieves NDCG@50 of 0.767, surpassing Multi-IPS's NDCG@50 of 0.738 by 3.92%. The consistent performance



(a) Results with varying discrepancies between populations.

(b) Results with varying strength of entropic regularization, with ϵ being 0.05, 0.1, 0.2 and 0.5 from left to right.

(c) Results with varying numbers of epochs to minimize WBM, with epochs being 0, 10, 50 and 100 from left to right.

Fig. 3. The balancing capability of the balancing scores calculated by minimizing the proposed WBM. The light area indicates the (uniform) weight of samples in α and β; the dark area indicates the weight of samples in the reweighted sets α' and β'. The dashed lines indicate the transport matrix. $\mathcal{W}(\alpha',\beta')$ indicates the discrepancy between the reweighted sets.

boost of DR methods can be attributed to their ability to reduce variance while leveraging prediction error from both treated and untreated populations, leading to more stable and accurate recommendations in real-world applications.

— Our proposed methods, IPS-WBM and DR-WBM, demonstrate superior performance compared to other IPS and DR methods. For example, on the KuaiRec dataset, IPS-WBM and DR-WBM achieve the AUC of 0.783 and 0.802, outperforming ESCM^2-IPS and ESCM^2-DR by 0.51% and 1.77%, respectively. In terms of top-k metrics, DR-WBM achieves NDCG@50 = 0.808 and F1@50 = 0.611 on KuaiRec, outperforming other debiasing methods significantly. The superior performance is attributed to the improved balancing capacity of learned propensity scores by WBM, which effectively handles selection bias and improves debiased recommendation performance.

5.3 A Case Study on the Balancing Capability of WBM

To showcase the efficacy of WBM's learned propensity scores to balance populations, a case study is conducted in Figure 3, using simulated datasets consistent with those in Figure 2. The primary observations are summarized below.

Table 4. Varying Weight of WBM (λ_b) Results

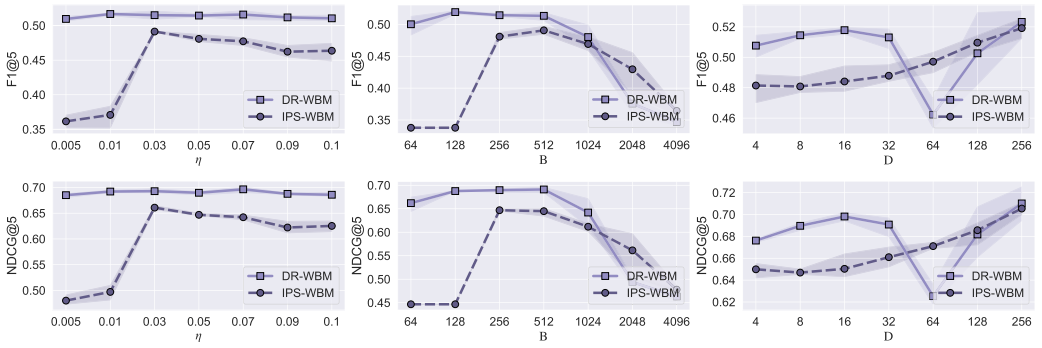
		Coat			Music		
	λ_b	AUC	NDCG@5	F1@5	AUC	NDCG@5	F1@5
IPS-WBM	$\lambda_b = 0$	0.711	0.604	0.463	0.651	0.667	0.331
	$\lambda_b = 0.05$	0.739	0.650	0.487	0.707	0.682	0.336
	$\lambda_b = 0.1$	0.739	0.650	0.487	0.708	**0.686**	**0.340**
	$\lambda_b = 0.5$	0.739	0.650	0.487	**0.708**	0.685	0.338
	$\lambda_b = 1$	0.739	**0.650**	**0.487**	0.706	0.685	0.338
	$\lambda_b = 5$	0.739	0.649	0.486	0.702	0.679	0.335
	$\lambda_b = 10$	**0.740**	0.649	0.483	0.696	0.679	0.335
DR-WBM	$\lambda_b = 0$	0.719	0.634	0.480	0.686	0.660	0.323
	$\lambda_b = 0.05$	0.744	0.692	0.514	0.719	0.679	0.364
	$\lambda_b = 0.1$	0.744	0.692	0.514	0.719	0.678	**0.364**
	$\lambda_b = 0.5$	**0.744**	**0.693**	**0.515**	0.719	0.678	0.364
	$\lambda_b = 1$	0.744	0.693	0.514	**0.719**	0.679	0.363
	$\lambda_b = 5$	0.744	0.693	0.514	0.719	0.680	0.364
	$\lambda_b = 10$	0.744	0.692	0.513	0.719	**0.681**	0.364

—First, the learned propensity scores prove effective in balancing biased populations. The inverse propensity tends to be higher for samples within the overlapping fields and smaller for others, which effectively reduces selection bias and diminishes the discrepancy between the reweighted populations.
—Second, WBM exhibits better capacity to balance biased populations compared to NLL. Specifically, the discrepancies between populations reweighted by WBM's propensity scores are significantly smaller than those observed in Figure 2, with relative reduction of Wasserstein discrepancy at least 60%. It is attributed to the efficacy of OT to directly minimize discrepancy between biased populations.
—Third, the incorporation of the entropic regularizer, which enables the differentiation of WBM with respect to propensity scores, influences the calculation of discrepancy. As illustrated in Figure 3(b), as ϵ decreases, the matching strategy is refined, and the discrepancy converges to the Wasserstein discrepancy [2]. Nevertheless, as ϵ decreases, the difference in discrepancy becomes smaller, particularly when compared to the discrepancies observed in NLL-based cases. Therefore, it is plausible to incorporate entropic regularizer to make WBM differentiable to propensity scores.
—Finally, WBM can be effectively optimized through stochastic gradient methods. As shown in Figure 3(c), WBM reduces swiftly as we increase the number of epochs to minimize WBM, with WBM being 2.606 in the initial stage and 0.439 after optimizing 100 epochs.

5.4 Ablation Study

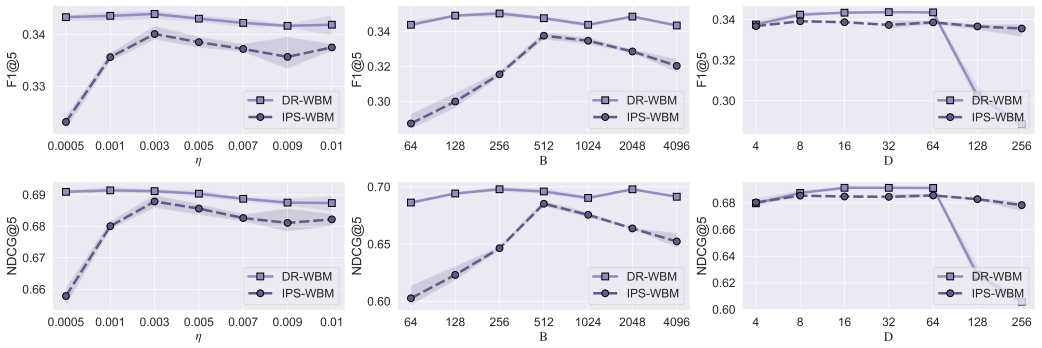
In this section, we examine the effectiveness of incorporating WBM to improve debiased recommendation performance. Table 4 presents the results for different WBM weights, denoted as λ_b, on the Coat and Music datasets. Key observations are summarized as follows.

—Incorporating WBM effectively enhances debiased recommendation performance. For example, increasing λ_b from 0 to 1 on the Coat and Music datasets results in a 3.93% and 8.75% AUC increase, respectively. These enhancements are also reflected in top-k metrics, with NDCG@5 improving by 7.61% on Coat and 2.69% on Music. A similar trend is observed for the DR-based



(a) Results with varying learning rate (η) (b) Results with varying batch size (B) (c) Results with varying embedding size (D)

Fig. 4. Hyperparameter sensitivity study with F1@5 and NDCG@5 on the Coat dataset, with colored lines for means and shaded areas for 99.9% CIs.



(a) Results with varying learning rate (η) (b) Results with varying batch size (B) (c) Results with varying embedding size (D)

Fig. 5. Hyperparameter sensitivity study with F1@5 and NDCG@5 on the Music dataset, with colored lines for means and shaded areas for 99.9% CIs.

estimator, with NDCG@5 increasing by 9.30% on Coat and 2.87% on Music. These results underscore the critical role of balancing capability in propensity score learning.

— The WBM-enhanced estimators exhibit notable improvements across a wide range of λ_b values. For instance, the IPS-WBM estimator on the Music dataset shows gains of over 8% in AUC and 2% in NDCG@5 for λ_b values between 0.05 and 1. A notable phenomenon is that the improvement is consistent under different values of λ, which is also observed in prevailing studies [26]. We attribute this phenomenon to the nature of the proposed WBM, which emphasizes the balance of reweighted distributions rather than the accuracy of propensity score estimation. In this case, large values of λ strengthen the balancing property without causing overfitting, which makes consistent performance improvement. The slight performance drop in some cases could be due to the optimizer's instability when handling excessively large weights.

5.5 Hyper-Parameter Sensitivity Study

In this section, we examine the influence of critical hyperparameters on the proposed estimators. The results on the Coat and Music datasets are presented in Figures 4 and 5, respectively. Key observations are summarized as follows.

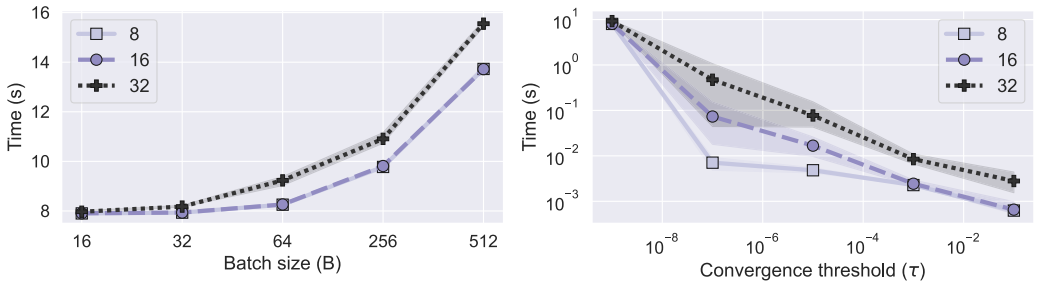


Fig. 6. Running time of calculating WBM given different settings of batch size (B), embedding size (D), and convergence threshold (τ). Different colors indicate different embedding sizes. The dashed lines indicate the average from 5 runs, and the shaded areas indicate 99.9% CIs.

— The update rate (η) controls the volume of model parameter updates each epoch. Overall, the DR-WBM estimator is robust to variations in the update rate across a wide range of values on both datasets, whereas the IPS-WBM estimator exhibits an increasing-then-decreasing trend as η increases. An update rate of approximately 0.03 and 0.003 was found to be optimal for IPS-WBM on the Coat and Music datasets, respectively, effectively balancing update stability and speed.
— The batch size (B) determines the number of samples used in each parameter update, with sizes ranging from 64 to 4,096 examined. For both models, there is a consistent improvement in F1@5 and NDCG@5 as the batch size increases up to B = 256. However, increasing the batch size beyond 512 does not yield further performance gains and introduces unnecessary computational overhead. Therefore, a batch size of around 256 to 512 offers an effective balance between performance and computational efficiency on both datasets.
— The embedding size (D) controls the capacity of the parametric estimator. Overall, the IPS-WBM approach benefits from larger embedding sizes. On the Coat dataset, increasing D from 4 to 256 consistently improves F1 and NDCG metrics. On the Music dataset, performance remains relatively stable across a wide range of D values. In contrast, the DR-WBM estimator is sensitive to the embedding size. On the Coat dataset, performance improves as D increases up to 16 but significantly declines at D = 64. Similarly, on the Music dataset, a substantial performance drop is observed when D exceeds 64. This sensitivity may stem from the incorporation of an error imputation model, which is challenging to train and prone to overfitting, thereby limiting the scalability of DR-based estimators to larger embedding sizes.

5.6 Complexity Study

In this section, we investigate the running time cost of calculating the proposed WBM metric, which primarily involves solving the OT problem in (7) using the Sinkhorn algorithm [10]. The time complexity is mainly influenced by the batch size (B), embedding size (D), and convergence threshold (τ). An empirical analysis of the OT problem solution is presented in Figure 6, with key observations summarized below.

— The batch size directly impacts the scale of the optimization problem and, consequently, the computational complexity of calculating WBM. As shown in the left panel of Figure 6, the running time increases super-linearly with batch size, reaching nearly 16 seconds for B = 512 with a convergence threshold of $\tau = 1e^{-9}$. This demonstrates a limitation of the WBM metric, where large batch sizes result in high computational cost.

— The embedding dimension influences the size of the pairwise distance matrix $\mathbf{D}$ in (7), which grows with increasing embedding size. This requires more iterations to reach the convergence threshold, as demonstrated in both panels of Figure 6. For instance, in the left panel, the running time increases from approximately 14–16 seconds as the embedding size grows from 8 to 32.
— The convergence threshold determines the number of iterations required for the algorithm to terminate. The right panel of Figure 6 shows the running time for different values of τ, with B = 64. The running time increases as τ decreases, reaching nearly 10 seconds at $\tau = 1e^{-9}$. However, relaxing the threshold to $1e^{-7}$, which is sufficient to ensure the accuracy of the balancing result, reduces the running time to under 1 second. Thus, adjusting the stopping threshold offers a practical tradeoff between accuracy and running time, providing a strategy to reduce computational cost in large-scale applications.

6 Related Work

Recommendation systems serve as cornerstones in industries such as e-commerce [53], advertising [39, 66], and social media [50, 70]. However, the data collected for these systems are observational rather than experimental, leading to selection bias. This bias creates a discrepancy between the training and test datasets, causing trained models to perform suboptimally during online deployment. To mitigate selection bias, debiased recommendation methods seek to estimate and optimize an unbiased learning objective using the biased training data [59, 61]. This area has garnered substantial attention [26, 30, 67], focusing on two primary challenges [8]: (1) how to estimate the propensity score; (2) how to derive an unbiased learning objective based on the propensity score. In this section, we provide a comprehensive review of the key works to address these challenges.

6.1 Learning Objective for Debiased Recommendation

The development of unbiased estimators stems from the IPS approach [43]. IPS addresses selection bias by inversely weighting the prediction error using the propensity score associated with each sample in the treated population. Since samples that are less likely to be treated have lower propensity scores, IPS assigns them higher weights, thereby reducing selection bias [26]. Theoretically, IPS provides an unbiased estimator of the ideal learning objective when the propensity scores are accurately estimated [43, 53]. However, IPS is prone to high variance when propensity scores are small, and it can produce biased estimates if the propensity scores are inaccurately estimated. Both issues limit its effectiveness in real-world recommendations [45].

To counteract these defects, the DR estimator was introduced, enhancing IPS with error imputation techniques. It reduces variance compared to IPS under mild assumptions and is less dependent on accurate propensity score estimation. Subsequent advancements have focused on further reducing variance and increasing resilience to small propensity scores. For example, Li and Sui [31] treat the propensity score as a user self-selection mechanism and propose a machine unlearning approach to reduce the effect of users with more non-random self-selection behavior; Guo et al. [17] propose a more robust DR estimator that employs a double learning procedure to minimize estimation variance; Li et al. [23] introduce an ensemble learning framework to achieve multiple robustness, while Li et al. [29] develop a cyclic optimization technique to ensure model stability and calibration, effectively reducing variance and enhancing stability for small propensity scores. Additionally, Song et al. [48] suggest filtering out false error imputations to decrease variance and improve tail bounds, and Li et al. [24] demonstrate methods to simultaneously reduce bias and variance in DR estimators when the error imputation model is misspecified.

6.2 Propensity Estimation for Debiased Recommendation

Accurate estimation of propensity scores is critical for the efficacy of the IPS and DR estimators [26, 30]. Various strategies have been employed to learn these propensity scores. Initially, some studies estimated propensity scores using a power-law function of item popularity, based on the number of interactions an item receives [44]. While these methods are straightforward to implement and interpret, their oversimplified assumptions often lead to biased recommendations. A more prevalent approach involves estimating propensity scores via logistic regression, as introduced by Rosenbaum and Rubin [43] and widely adopted in recommendation practices [3, 11, 22, 67].

On this basis, one line of research has focused on enhancing the accuracy of propensity estimation by integrating advanced learning techniques, such as joint optimization [53, 67] and alternative training [72]. Recently, they have innovatively incorporated a small subset of unbiased data during training [25, 27], effectively addressing missing confounders and substantially improving estimation quality with minimal data collection costs. Another line of research aims to learn propensity scores that ensure the unbiasedness of the recommendation model under non-ideal conditions, where there exists user and item model misspecifications [27] and noisy feedbacks [28].

7 Conclusion

In this work, we focus on the challenge of selection bias in recommendation systems stemming from the use of observational data. We highlight the limitations of existing propensity estimation methods that rely on NLL minimization, which may not sufficiently correct for bias. To overcome this, we introduce the WBM, a novel approach to measure and enhance the balancing capacity of propensity scores by minimizing the Wasserstein distance between reweighted and ideal distributions. We develop the *IPS-WBM* and *DR-WBM* estimators by integrating WBM as a regularizer into the standard IPS and DR frameworks, enhancing their effectiveness without introducing additional bias. Our extensive experiments on real-world datasets demonstrate that incorporating WBM significantly improves the balancing capability of propensity scores, leading to better debiasing performance in recommendation tasks. These findings underscore the potential of directly addressing distributional discrepancies in propensity estimation and open avenues for future research in developing more advanced propensity learning techniques for debiased recommendations.

Limitations and Future Work. This study focuses on selection bias in recommendation systems. However, in industrial settings, additional biases such as position bias and popularity bias are prevalent. Extending the WBM framework to account for these biases could further enhance recommendation performance. Additionally, while the selected Wasserstein discrepancy effectively quantifies distributional differences, the calculation process is computationally intensive, particularly with large batch sizes common in practice. Future work could explore alternative discrepancy measures that maintain similar statistical properties while offering reduced computational complexity, thereby improving the scalability and applicability in large-scale debiased recommendation scenarios.

References

[1] Ravindra K. Ahuja and James B. Orlin. 1992. The scaling network simplex algorithm. *Oper. Res.* 40, Supplement-1 (1992), S5–S13.

[2] Jason M. Altschuler, Jonathan Weed, and Philippe Rigollet. 2017. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Proc. Adv. Neural Inf. Process. Syst.*, 1964–1974.

[3] Jiawei Chen, Hande Dong, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, and Keping Yang. 2021. Autodebias: Learning to debias for recommendation. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 21–30.

[4] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and debias in recommender system: A survey and future directions. *ACM T. Inform. Syst.* 41, 3 (2023), 1–39.

[5] Zhichao Chen, Haoxuan Li, Fangyikang Wang, Haotian Zhang, Hu Xu, Xiaoyu Jiang, Zhihuan Song, and Hao Wang. 2024. Rethinking the diffusion models for missing data imputation: A gradient flow perspective. In *Proc. Adv. Neural Inf. Process. Syst*.

[6] Zhichao Chen, Licheng Pan, Yiran Ma, Zeyu Yang, Le Yao, Jinchuan Qian, and Zhihuan Song. 2025. E^2AG: Entropy-regularized ensemble adaptive graph for industrial soft sensor modeling. *IEEE/CAA J. Autom. Sinica.* (2025), 1–16.

[7] Zhichao Chen, Hao Wang, Zhihuan Song, and Zhiqiang Ge. 2024. Improving data-driven inferential sensor modeling by industrial knowledge: A Bayesian perspective. *IEEE Trans. Syst., Man, Cybern., Syst*. (2024), 1–13.

[8] Zhixuan Chu, Jianmin Huang, Ruopeng Li, Wei Chu, and Sheng Li. 2023. Causal effect estimation: Recent advances, challenges, and opportunities. arXiv:2302.00848. Retrieved from https://arxiv.org/abs/2302.00848

[9] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. 2017. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell*. 39, 9 (2017), 1853–1865.

[10] Marco Cuturi. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *Proc. Adv. Neural Inf. Process. Syst*., 2292–2300.

[11] Quanyu Dai, Haoxuan Li, Peng Wu, Zhenhua Dong, Xiao-Hua Zhou, Rui Zhang, Rui Zhang, and Jie Sun. 2022. A generalized doubly robust learning framework for debiasing post-click conversion rate prediction. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 252–262.

[12] Yann Disser and Martin Skutella. 2019. The simplex algorithm is NP-Mighty. *ACM Trans. Algorithms* 15, 1 (2019), 5:1–5:19.

[13] Pavel E. Dvurechensky, Alexander V. Gasnikov, and Alexey Kroshnin. 2018. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn's Algorithm. In *Proc. Int. Conf. Mach. Learn.*, Vol. 80, 1366–1375.

[14] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier. 2021. POT: Python optimal transport. *J. Mach. Learn. Res.* 22 (2021), 78:1–78:8.

[15] Chongming Gao, Shijun Li, Wenqiang Lei, Jiawei Chen, Biao Li, Peng Jiang, Xiangnan He, Jiaxin Mao, and Tat-Seng Chua. 2022. KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In *Proc. ACM Int. Conf. Inf. Knowl. Manag*. ACM, 540–550.

[16] Chongming Gao, Shijun Li, Yuan Zhang, Jiawei Chen, Biao Li, Wenqiang Lei, Peng Jiang, and Xiangnan He. 2022. KuaiRand: An unbiased sequential recommendation dataset with randomly exposed videos. In *Proc. ACM Int. Conf. Inf. Knowl. Manag*. ACM, 3953–3957.

[17] 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 *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 275–284.

[18] Leonid V. Kantorovich. 2006. On the translocation of masses. *J. Math. Sci.* 133, 4 (2006), 1381–1382.

[19] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Represent*., 1–9.

[20] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[21] Wonbin Kweon and Hwanjo Yu. 2024. Doubly calibrated estimator for recommendation on data missing not at random. In *Proc. Int. Conf. World Wide Web*, 3810–3820.

[22] Jae-woong Lee, Seongmin Park, and Jongwuk Lee. 2021. Dual unbiased recommender learning for implicit feedback. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 1647–1651.

[23] Haoxuan Li, Quanyu Dai, Yuru Li, Yan Lyu, Zhenhua Dong, Xiao-Hua Zhou, and Peng Wu. 2023. Multiple robust learning for recommendation. In *Proc. AAAI Conf. Artif. Intell*., Vol. 37, 4417–4425.

[24] Haoxuan Li, Yan Lyu, Chunyuan Zheng, and Peng Wu. 2023. TDR-CL: Targeted doubly robust collaborative learning for debiased recommendations. In *Proc. Int. Conf. Learn. Represent*., 1–9.

[25] Haoxuan Li, Kunhan Wu, Chunyuan Zheng, Yanghao Xiao, Hao Wang, Zhi Geng, Fuli Feng, Xiangnan He, and Peng Wu. 2024. Removing hidden confounding in recommendation: A unified multi-task learning approach. *Proc. Adv. Neural Inf. Process. Syst*. 36 (2024), 54614–54626.

[26] Haoxuan Li, Yanghao Xiao, Chunyuan Zheng, Peng Wu, and Peng Cui. 2023. Propensity Matters: Measuring and enhancing balancing for recommendation. In *Proc. Int. Conf. Mach. Learn.*, Vol. 202, PMLR, 20182–20194.

[27] Haoxuan Li, Chunyuan Zheng, Shuyi Wang, Kunhan Wu, Eric Wang, Peng Wu, Zhi Geng, Xu Chen, and Xiao-Hua Zhou. 2024. Relaxing the accurate imputation assumption in doubly robust learning for debiased collaborative filtering. In *Proc. Int. Conf. Mach. Learn.*, Vol. 235, 29448–29460.

[28] Haoxuan Li, Chunyuan Zheng, Wenjie Wang, Hao Wang, Fuli Feng, and Xiao-Hua Zhou. 2024. Debiased recommendation with noisy feedback. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 1576–1586.

[29] Haoxuan Li, Chunyuan Zheng, and Peng Wu. 2023. StableDR: Stabilized doubly robust learning for recommendation on data missing not at random. In *Proc. Int. Conf. Learn. Represent.*, 1–9.
[30] Haoxuan Li, Chunyuan Zheng, Yanghao Xiao, Peng Wu, Zhi Geng, Xu Chen, and Peng Cui. 2024. Debiased collaborative filtering with kernel-based causal balancing. In *Proc. Int. Conf. Learn. Represent.*, 1–9.
[31] Meng Li and Haochen Sui. 2025. Causal recommendation via machine unlearning with a few unbiased data. In *AAAI 2025 Workshop on Artificial Intelligence with Causal Techniques*.
[32] Chang Liu, Jingwei Zhuo, Pengyu Cheng, Ruiyi Zhang, and Jun Zhu. 2019. Understanding and accelerating particle-based variational inference. In *Proc. Int. Conf. Mach. Learn.* PMLR, 4082–4092.
[33] Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Jiajie Su, Yanchao Tan, and Fan Wang. 2024. User distribution mapping modelling with collaborative filtering for cross domain recommendation. In *Proc. Int. Conf. World Wide Web*, 334–343.
[34] Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Yanchao Tan, Fan Wang, Xiaolin Zheng, and Yew Soon Ong. 2024. Learning accurate and bidirectional transformation via dynamic embedding transportation for cross-domain recommendation. In *Proc. AAAI Conf. Artif. Intell.*, Vol. 38, 8815–8823.
[35] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Mengling Hu, Xinting Liao, Fan Wang, Yanchao Tan, Dan Meng, and Jun Wang. 2023. Differentially private sparse mapping for privacy-preserving cross domain recommendation. In *Proc. ACM Int. Conf. Multimedia*, 6243–6252.
[36] 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 *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 1137–1140.
[37] Benjamin M. Marlin, Richard S. Zemel, Sam Roweis, and Malcolm Slaney. 2007. Collaborative filtering and the missing at random assumption. In *Proc. Conf. Uncertainty in Artificial Intelligence*, 267–275.
[38] Gaspard Monge. 1781. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.* (1781), 666–704.
[39] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling fairness and bias in dynamic learning-to-rank. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 429–438.
[40] Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. 2021. OT-Flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proc. AAAI Conf. Artif. Intell.*, 9223–9232.
[41] Gabriel Peyré and Marco Cuturi. 2019. Computational optimal transport. *Found. Trends Mach. Learn.* 11, 5–6 (2019), 355–607.
[42] James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. 1994. Estimation of regression coefficients when some regressors are not always observed. *J. Am. Stat. Assoc.* 89, 427 (1994), 846–866.
[43] Paul Rosenbaum and Donald B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 1 (1983), 41–55.
[44] Yuta Saito. 2020. Unbiased pairwise learning from biased implicit feedback. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 5–12.
[45] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased recommender learning from missing-not-at-random implicit feedback. In *Proc. Int. Conf. Web Search Data Mining*, 501–509.
[46] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *Proc. Int. Conf. Mach. Learn.*, Vol. 48, 1670–1679.
[47] Uri Shalit, Fredrik D. Johansson, and David Sontag. 2017. Estimating individual treatment effect: Generalization bounds and algorithms. In *Proc. Int. Conf. Mach. Learn.*, 3076–3085.
[48] Zijie Song, Jiawei Chen, Sheng Zhou, Qihao Shi, Yan Feng, Chun Chen, and Can Wang. 2023. CDR: Conservative doubly robust learning for debiased recommendation. In *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, 2321–2330.
[49] Hongzu Su, Lichao Meng, Lei Zhu, Ke Lu, and Jingjing Li. 2024. DDPO: Direct dual propensity optimization for post-click conversion rate estimation. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 1179–1188.
[50] Julien Subercaze, Christophe Gravier, and Frederique Laforest. 2018. Real-time, scalable, content-based Twitter users recommendation. In *Proc. Int. Conf. World Wide Web*, 1367.
[51] Karan Vombatkere, Sepehr Mousavi, Savvas Zannettou, Franziska Roesner, and Krishna P. Gummadi. 2024. TikTok and the art of personalization: Investigating exploration and exploitation on social media feeds. In *Proc. Int. Conf. World Wide Web*, 3789–3797.
[52] Fangyikang Wang, Huminhao Zhu, Chao Zhang, Hanbin Zhao, and Hui Qian. 2024. GAD-PVI: A general accelerated dynamic-weight particle-based variational inference framework. In *Proc. AAAI Conf. Artif. Intell.*, 15466–15473.
[53] Hao Wang, Tai-Wei Chang, Tianqiao Liu, Jianmin Huang, Zhichao Chen, Chao Yu, Ruopeng Li, and Wei Chu. 2022. ESCM2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 363–372.

[54] Hao Wang, Zhichao Chen, Jiajun Fan, Haoxuan Li, Tianqiao Liu, Weiming Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, and Ruiming Tang. 2023. Optimal transport for treatment effect estimation. In *Proc. Adv. Neural Inf. Process. Syst*., Vol. 36, 5404–5418.

[55] Hao Wang, Zhichao Chen, Zhaoran Liu, Haozhe Li, Degui Yang, Xinggao Liu, and Haoxuan Li. 2024. Entire space counterfactual learning for reliable content recommendations. *IEEE Trans. Inf. Forensics Security* (2024), 1–12.

[56] Hao Wang, Zhichao Chen, Zhaoran Liu, Licheng Pan, Hu Xu, Yilin Liao, Haozhe Li, and Xinggao Liu. 2024. SPOT-I: Similarity preserved optimal transport for industrial IoT data imputation. *IEEE Trans. Ind. Informat*. 20, 12 (2024), 14421–14429. DOI: https://doi.org/10.1109/TII.2024.3452241

[57] Hao Wang, Zhengnan Li, Haoxuan Li, Xu Chen, Mingming Gong, Bin Chen, and Zhichao Chen. 2025. Optimal transport for time series imputation. In *Proc. Int. Conf. Learn. Represent*., 1–9.

[58] Hao Wang, Xinggao Liu, Zhaoran Liu, Haozhe Li, Yilin Liao, Yuxin Huang, and Zhichao Chen. 2024. LSPT-D: Local similarity preserved transport for direct industrial data imputation. *IEEE Trans. Autom. Sci. Eng*. (2024).

[59] Wenjie Wang, Yang Zhang, Haoxuan Li, Peng Wu, Fuli Feng, and Xiangnan He. 2023. Causal recommendation: Progresses and future directions. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*. ACM, 3432–3435.

[60] Xiaojie Wang, Rui Zhang, Yu Sun, and Jianzhong Qi. 2021. Combating selection biases in recommender systems with a few unbiased ratings. In *Proc. Int. Conf. Web Search Data Mining*, 427–435.

[61] Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, and Xiao-Hua Zhou. 2022. On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges. In *Proc. Int. Joint Conf. Artif. Intell*., 23–29.

[62] Jingjing Xu, Hao Zhou, Chun Gan, Zaixiang Zheng, and Lei Li. 2021. Vocabulary learning via optimal transport for neural machine translation. In *ACL/IJCNLP (1)*. Association for Computational Linguistics, 7361–7373.

[63] Mengyue Yang, Guohao Cai, Furui Liu, Jiarui Jin, Zhenhua Dong, Xiuqiang He, Jianye Hao, Weiqi Shao, Jun Wang, and Xu Chen. 2023. Debiased recommendation with user feature balancing. *ACM T. Inform. Syst*. 41, 4 (2023), 1–25.

[64] Mengyue Yang, Quanyu Dai, Zhenhua Dong, Xu Chen, Xiuqiang He, and JunWang. 2021. Top-n recommendation with counterfactual user preference simulation. In *Proc. ACM Int. Conf. Inf. Knowl. Manag*., 2342–2351.

[65] Mengyue Yang, Jun Wang, and Jean-Francois Ton. 2023. Rectifying unfairness in recommendation feedback loop. In *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 28–37.

[66] Tan Yu, Yi Yang, Yi Li, Xiaodong Chen, Mingming Sun, and Ping Li. 2020. Combo-attention network for baidu video advertising. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2474–2482.

[67] 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 *Proc. Int. Conf. World Wide Web*, 2775–2781.

[68] Xinyue Zhang, Cong Huang, Kun Zheng, Hongzu Su, Tianxu Ji, Wei Wang, Hongkai Qi, and Jingjing Li. 2024. Adversarial-enhanced causal multi-task framework for debiasing post-click conversion rate estimation. In *Proc. Int. Conf. World Wide Web*, 3287–3296.

[69] Jiajing Zheng. 2021. *Sensitivity Analysis for Causal Inference with Unobserved Confounding*. University of California, Santa Barbara.

[70] Jiawei Zheng, Hao Gu, Chonggang Song, Dandan Lin, Lingling Yi, and Chuan Chen. 2023. Dual interests-aligned graph auto-encoders for cross-domain recommendation in WeChat. In *Proc. ACM Int. Conf. Inf. Knowl. Manag*.

[71] Chang Zhou, Jianxin Ma, Jianwei Zhang, Jingren Zhou, and Hongxia Yang. 2021. Contrastive learning for debiased candidate generation in large-scale recommender systems. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 3985–3995.

[72] Ziwei Zhu, Yun He, Yin Zhang, and James Caverlee. 2020. Unbiased implicit recommendation and propensity estimation via combinational joint learning. In *RecSys*., 551–556.

[73] Zhan Zhuang, Yu Zhang, and Ying Wei. 2024. Gradual domain adaptation via gradient flow. In *Proc. Int. Conf. Learn. Represent*., 1–26.

Received 4 October 2024; revised 20 December 2024; accepted 8 February 2025