Propensity Matters: Measuring and Enhancing Balancing for Recommendation

Haoxuan Li [1] **Yanghao Xiao** [2] **Chunyuan Zheng** [3] **Peng Wu** [4] **Peng Cui** [5]

Abstract

Propensity-based weighting methods have been widely studied and demonstrated competitive performance in debiased recommendations. Nevertheless, there are still many questions to be addressed. How to estimate the propensity more conducive to debiasing performance? Which metric is more reasonable to measure the quality of the learned propensities? Is it better to make the cross-entropy loss as small as possible when learning propensities? In this paper, we first discuss the potential problems of the previously widely adopted metrics for learned propensities, and propose balanced-mean-squared-error (BMSE) metric for debiased recommendations. Based on BMSE, we propose IPS-V2 and DR-V2 as the estimators of unbiased loss, and theoretically show that IPS-V2 and DR-V2 have greater propensity balancing and smaller variance without sacrificing additional bias. We further propose a co-training method for learning balanced representation and unbiased prediction. Extensive experiments are conducted on three real-world datasets including a large industrial dataset, and the results show that our approach boosts the balancing property and results in enhanced debiasing performance.

1. Introduction

For recommender systems (RSs), it is crucial to understand and answer the counterfactual question "what would the feedback be if an intervention had been made to a user", which covers many common tasks in RS (Wu et al., 2022; Chen et al., 2022). For example, in the task of rating prediction (Schnabel et al., 2016), we want to know the rating if a user had rated the item; in the task of post-view click-through rate prediction (Saito, 2019; Saito et al., 2020), we want to know the click-through rate if an item had been exposed to a user; in the task of post-click conversion rate prediction (Zhang et al., 2020; Dai et al., 2022), we want to know the conversion rate if an item had been clicked by a user. However, since users always choose preferred items to rate or click on and the exposure mechanism of the RS is not at random, causing the observed data is no longer a valid representative sample of the target population. In general, there is a large discrepancy between observed and missing events, and ignoring such discrepancy will incur bias and lead to sub-optimal performance (Schnabel et al., 2016).

In order to eliminate the differences between the observed samples and the target population, weighting-based methods are proposed to achieve the unbiasedness (Swaminathan & Joachims, 2015a; Schnabel et al., 2016). The basic idea is to reweight the observed data to the target population, according to the probability of observing an specific event (called propensity), which motivates many studies on the variants of propensity-based weighting methods (Wang et al., 2019a; Guo et al., 2021; Wang et al., 2021; Ding et al., 2022; Li et al., 2023b;e) and demonstrates competitive performance in a wide range of recommendation scenarios (Joachims et al., 2017; Saito et al., 2020; Wang et al., 2022).

Despite their popularity and theoretical appeal, the main practical difficulty of the propensity-based weighting methods is that the true propensity is rarely known and needs to be estimated from the observed data. Given the key role propensity played in debiasing, several approaches are proposed to estimate the propensities, such as Naive Bayes (Schnabel et al., 2016), Logistic Regression (Schnabel et al., 2016), Poisson Factorization (Wang et al., 2020a), Multi-task Learning (Zhang et al., 2020), Variance Regularization Constraint (Wang et al., 2021), Stabilized Constraint (Li et al., 2023b;e), and Minimax (Ding et al., 2022).

Nevertheless, a unified and clear criterion for estimating propensities has not been established yet. Many issues need to be resolved: How to estimate the propensity more conducive to debiasing performance? Which metric is more reasonable to measure the quality of the learned propensities? In practice, the propensities are usually trained by minimizing a cross-entropy loss. But, is it better to make the loss as small as possible when learning propensities?

In this paper, we find that the existing propensity estimation

[1]Peking University [2]University of Chinese Academy of Sciences [3]University of California, San Diego [4]Beijing Technology and Business University [5]Tsinghua University. Correspondence to: Peng Wu <pengwu@btbu.edu.cn>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

methods in debiased recommendations ignore the essence of propensity, i.e., the balancing property (Rosenbaum & Rubin, 1983; Imbens & Rubin, 2015; Hernán & Robins, 2020). That is, for all measurable functions of features (e.g., user and item embeddings, feature representations, and rating predictions), the expectation in the observed events weighting by the inverse of propensities is always equal to that in the target population (Imai & Ratkovic, 2014; Sant'Anna et al., 2022). Based on balancing property, prediction models trained in the weighted population can generalize to the missing events and achieve unbiasedness, which lays the foundation for propensity-based weighting methods.

Toward this end, we propose *balanced-mean-squared-error (BMSE)* as a measure of the quality of the learned propensities for debiased recommendations. We also empirically show that the smaller the BMSE, the better the debiasing performance of the propensity-based weighting methods.

Based on the BMSE, we propose principled enhanced versions of the widely used Inverse Propensity Scoring (IPS) and Doubly Robust (DR) estimators, named IPS-V2 and DR-V2, respectively. We theoretically demonstrate that IPS-V2 and DR-V2 have greater propensity balancing and smaller variance compared to existing IPS and DR, without sacrificing additional bias. For previous estimators with similar forms, we show that the direct use of variance regularizers comes at the cost of introducing additional bias and does not guarantee the balancing property of the propensities.

We further propose a learning method compatible with the balancing property for debiased recommendations. Specifically, the propensity model and the prediction model are jointly trained to permit both the balancing property of the feature representation (minimizing BMSE) and the unbiased prediction (minimizing IPS-V2 or DR-V2 loss). The proposed method has easy operability as well as can alleviate the data sparsity problem through parameter sharing.

The contributions of this paper are summarized as follows.

• We propose a metric (BMSE) to measure the balancing property of learned propensities in debiased recommendations, then theoretically and empirically reveal that the smaller the BMSE, the better debiasing performance of the propensity-based weighting methods.

• Based on BMSE, we propose IPS-V2 and DR-V2 as the estimators of unbiased loss, and theoretically show that IPS-V2 and DR-V2 have greater propensity balancing and smaller variance without sacrificing additional bias.

• We further propose a co-training method for learning balanced representation and unbiased prediction.

• We conduct extensive experiments on three datasets including a large-scale industrial dataset, and the results show that our approach boosts the balancing property of the learned propensities and results in enhanced debiasing performance.

2. Preliminaries

Let $\mathcal{U} = \{u\}$ be the set of users, $\mathcal{I} = \{i\}$ be the set of items. To define a causal problem, the widely adopted potential outcome framework (Rubin, 1974; Neyman, 1990) consists of the following key elements: (1) *Target population*: the set of all user-item pairs $\mathcal{D} = \{(u, i) \mid u \in \mathcal{U}, i \in \mathcal{I}\}$; (2) *Feature*: $x_{u,i}$, the feature of user u and item i; (3) *Treatment*: $o_{u,i} \in \{0, 1\}$, it has different implications for different prediction tasks in RS, e.g., whether the true rating $r_{u,i}$ is observed ($o_{u,i} = 1$) or missing ($o_{u,i} = 0$), whether item i is exposed to user u, and whether user u clicks item i; (4) *Outcome*: $r_{u,i}$, the feedback of user-item pair (u, i), e.g., rating, click indicator, conversion indicator; (5) *Potential outcome*: $r_{u,i}(o)$ for $o \in \{0, 1\}$, it is the outcome that would be observed if $o_{u,i}$ had been set to o. We provide more formulation details and examples in Appendix A.

Let $\mathbb{P}$ and $\mathbb{E}$ be the distribution and expectation on the target population, and $\mathcal{O} = \{(u, i) \mid (u, i) \in \mathcal{D}, o_{u,i} = 1\}$ be the set of treated units. In RS, the widely adopted counterfactual question is "what would the feedback be if an intervention had been made to a user", which is equivalent to learning the causal estimand $\mathbb{E}(r_{u,i}(1) \mid x_{u,i})$[1], i.e., predicting the potential outcomes $r_{u,i}(1)$ using feature $x_{u,i}$ for all $(u, i) \in \mathcal{D}$ (Imbens & Rubin, 2015; Hernán & Robins, 2020).

Note that $r_{u,i}(1)$ is observed only when $o_{u,i} = 1$, missing otherwise, the task of predicting $r_{u,i}(1)$ can be viewed as a missing data problem. However, there is always a discrepancy between observed events $\mathcal{O}$ and all events $\mathcal{D}$, due to the existence of confounders that affect both treatment and outcome. Ignoring this discrepancy will suffer from bias and result in sub-optimal performance (Wang et al., 2019a).

Let $\hat{r}_{u,i} = f(x_{u,i}; \theta)$ be the prediction model that aims to predict all $r_{u,i}(1)$ accurately. Ideally, if all potential outcome $r_{u,i}(1)$ are known, $\hat{r}_{u,i}$ can be trained by minimizing the average loss of all user-item pairs

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

where $e_{u,i} = L(\hat{r}_{u,i}, r_{u,i}(1))$ is the prediction error and $L(\cdot, \cdot)$ is an appropriately chosen loss function, e.g., squared loss $e_{u,i} = (\hat{r}_{u,i} - r_{u,i}(1))^2$. Although optimizing $\mathcal{L}_{ideal}(\theta)$ directly is infeasible due to the missingness of $r_{u,i}(1)$, $\mathcal{L}_{ideal}(\theta)$ provides a benchmark of unbiased learning and prediction theoretically. As such, various debiasing methods try to construct unbiased estimators of $\mathcal{L}_{ideal}(\theta)$ and train the prediction model by minimizing the estimated ideal loss.

The basic idea of the propensity-based weighting methods is to reweight the observed samples to the target popula-

[1]It is equivalent to $\mathbb{P}(r_{u,i}|x_{u,i}, do(o_{u,i} = 1))$ using *do-calculus* in SCM framework.

tion (Schnabel et al., 2016), which exhibit competitive debiasing performance (Wang et al., 2019a; Dai et al., 2022; Ding et al., 2022). Specifically, the Inverse Propensity Scoring (IPS) estimator (Schnabel et al., 2016) assigns higher training weights to events with smaller probability of being observed in the training set, and is formulated as

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

where $\hat{p}_{u,i}$ is the propensity model used to estimate $p_{u,i} \triangleq \mathbb{P}(o_{u,i} = 1|x_{u,i})$. By further introducing a error imputation model on IPS, the Doubly Robust (DR) estimator (Wang et al., 2019a; Saito, 2020; Li et al., 2023b;e) is defined as

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

where $\hat{e}_{u,i}$ is the error imputation model that estimates $e_{u,i}$. The DR estimator has double robustness, i.e., it is an unbiased estimator of $\mathcal{L}_{ideal}(\theta)$ when either imputed errors or learned propensities are accurate.

3. The Central Role of Propensity in Debiasing

3.1. Motivation

Propensity is ubiquitous and plays a critical role in debiased recommendations. For example, the unbiasedness of the IPS and SNIPS estimators (Schnabel et al., 2016; Swaminathan & Joachims, 2015a; Saito et al., 2020) depends on the accuracy of the learned propensities. Wang et al. (2020a) treats propensities as new features and includes them in the matrix factorization. Zhang et al. (2020) proposes a multi-task learning approach to train the propensity model and prediction model simultaneously. Li et al. (2023b) proposes a collaborative targeted learning approach incorporating the learned propensities to the training process of the imputed errors. Li et al. (2023e) proposes a stabilized constraint to train the propensities. Ding et al. (2022) propose to fluctuate the inverse propensities to mitigate the influence of unmeasured confounders. Wang et al. (2021), Chen et al. (2021) and Li et al. (2023c) suggest using a small uniform dataset for seeking better estimates of propensities.

Some might argue that the DR estimator is unbiased when the imputed errors are accurate, regardless of the accuracy of the learned propensities. However, due to the missing events, the training of the imputed errors relies heavily on the learned propensities (Wang et al., 2019a; Guo et al., 2021; Chen et al., 2021; Dai et al., 2022; Ding et al., 2022), but not vice versa. Specifically, the error imputation model $\hat{e}_{u,i}$ is typically trained by minimizing

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

which depends on a pre-specified propensity model $\hat{p}_{u,i}$. Therefore, if the learned propensities 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 propensities in debiased recommendations, we aim to establish a unified propensity training standard. Importantly, there are several questions that need to be answered. How to learn propensity that is more helpful for debiasing performance? Is it better to predict $o_{u,i}$ as accurately as possible? Which metric reasonably measures the quality of the learned propensities?

3.2. Are NLL and PPL Proper Metrics for Propensity Model Training?

In practice, we usually train the propensity model by optimizing the cross-entropy loss (also known as the negative log-likelihood, NLL)

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

or the perplexity (PPL)

$$\mathcal{L}'_p = 2^{-\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} o_{u,i}\cdot\log_2(\hat{p}_{u,i}) + (1-o_{u,i})\cdot\log_2(1-\hat{p}_{u,i})},$$

which corresponds to finding a propensity model that predicts $o_{u,i}$ as accurately as possible. However, are the learned propensities with smaller NLL and PLL sufficiently lead to a better debiasing performance?

It is obviously not. Consider an extreme case where $\hat{p}_{u,i} = 0$ for $o_{u,i} = 0$ and $\hat{p}_{u,i} = 1$ for $o_{u,i} = 1$. Although such propensities reach the smallest NLL and PLL, it would reduce $\mathcal{L}_{IPS}(\theta)$ to a Naive estimator

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

that is, the simple averaging of losses over the observed events, which leads to biased estimates on the target population. Besides, it also reduces $\mathcal{L}_{DR}(\theta)$ to an Error Imputation-Based (EIB) estimator (Steck, 2010) that

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

where $\hat{e}_{u,i}$ is the error imputation model used to estimate $e_{u,i}$, and EIB methods empirically showed suboptimal performance due to the sparsity of the collected data and the difficulty of obtaining accurate imputed errors (Schnabel et al., 2016; Wang et al., 2019a; Guo et al., 2021).

3.3. Proposed Balancing-Mean-Square-Error Metric

From a causal inference perspective, the role of propensity weighting is to recover the distribution of $x_{u,i}$ in the target

population $\mathcal{D}$ from the observed events $\mathcal{O}$ (Imai & Ratkovic, 2014; Imbens & Rubin, 2015; Wong & Chan, 2018; Rosenbaum, 2020; Sant'Anna et al., 2022). Formally, for any measurable and integrable function $\boldsymbol{\phi} : \mathcal{X} \to \mathbb{R}^m$, recap that $p_{u,i} = \mathbb{P}(o_{u,i} = 1|x_{u,i}) = \mathbb{E}[o_{u,i}|x_{u,i}]$, we have

$$\mathbb{E}\left[\frac{o_{u,i}\boldsymbol{\phi}(x_{u,i})}{p_{u,i}}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{o_{u,i}\boldsymbol{\phi}(x_{u,i})}{p_{u,i}}|x_{u,i}\right]\right] = \mathbb{E}\left[\frac{\boldsymbol{\phi}(x_{u,i})}{p_{u,i}}\mathbb{E}(o_{u,i}|x_{u,i})\right] = \mathbb{E}[\boldsymbol{\phi}(x_{u,i})],$$

where the first equation follows from the law of iterated expectations, the second equation follows from the fact that both $\phi(x_{u,i})$ and $p_{u,i}$ are functions of $x_{u,i}$, thus are considered as constants after given $x_{u,i}$. Similarly, we have

$$\mathbb{E}\left[\frac{(1-o_{u,i})\boldsymbol{\phi}(x_{u,i})}{1-p_{u,i}}\right] = \mathbb{E}[\boldsymbol{\phi}(x_{u,i})].$$

The arbitrariness of $\boldsymbol{\phi}$ indicates that propensity weighting creates a pseudo-population consisting of the observed events weighted by $1/p_{u,i}$, whose distribution of $x_{u,i}$ is identical to the target population $\mathcal{D}$ (Hernán & Robins, 2020). In fact, the balancing property of propensity mimics randomization, as if the collected data comes from a randomized controlled trial, then the entire distribution of features between observed events and the target population will be the same. As a result, training the prediction model in the pseudo-population acts as training in the target population, and thus can be naturally generalized to all events.

Unfortunately, the existing propensity estimation methods do not take into account such essential balancing property. Therefore, the generalizability of the prediction model cannot be guaranteed once using the estimated propensity $\hat{p}_{u,i}$ instead of the true propensity $p_{u,i}$. To fill this gap, we first introduce a metric, named balancing-mean-square-error (BMSE), for measuring the balancing property of the learned propensities

$$\text{BMSE}(\boldsymbol{\phi}, \hat{p}) = \left|\left|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right]\boldsymbol{\phi}(x_{u,i})\right|\right|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm, and $\boldsymbol{\phi}$ is a pre-specified vector-valued function, such as learned embeddings $x_{u,i}$, learned representations $\boldsymbol{\phi}(x_{u,i})$, and predicted ratings $\hat{r}_{u,i}$.

Proposition 3.1 (Balancing Property)**.** *If $\hat{p}_{u,i}$ estimates $p_{u,i}$ accurately, i.e, $\hat{p}_{u,i} = p_{u,i}$, then for any integrable vector-valued functions $\boldsymbol{\phi}(x)$, BMSE$(\boldsymbol{\phi}, \hat{p}) \to 0$ almost surely.*

From Proposition 3.1, BMSE$(\boldsymbol{\phi}, \hat{p})$ will converge to zero if the learned propensities are accurate. We also empirically show that choosing $\boldsymbol{\phi}$ as a constant, or the concatenation / element-wise product of user and item embeddings, or the predicted ratings are all beneficial for improving the debiasing performance (see Section 5 for details).

4. Balancing-Enhanced Learning Framework

4.1. Proposed Balancing-Enhanced Estimators

The BMSE metric is designed to evaluate the balancing property of the learned propensities rather than directly training an unbiased prediction model. To fill this gap, we further propose balancing-enhanced IPS and DR estimators, named IPS-V2 and DR-V2, using BMSE as a regularization on the vanilla IPS and DR estimators, which have greater balancing and smaller variance without sacrificing additional bias.

The proposed balancing-enhanced IPS estimator is

$$\mathcal{L}_{IPS-V2}(\theta) = \mathcal{L}_{IPS}(\theta) + \lambda \cdot \text{BMSE}(\boldsymbol{\phi}, \hat{p}),$$

where $\lambda > 0$ is a scalar weight which trade-offs the balancing property and the prediction performance. Similarly, the balancing-enhanced DR estimator is

$$\mathcal{L}_{DR-V2}(\theta) = \mathcal{L}_{DR}(\theta) + \lambda \cdot \text{BMSE}(\boldsymbol{\phi}, \hat{p}).$$

Theorem 4.1 (Unbiasedness of IPS-V2 and DR-V2)**.** *When learned propensities are accurate,*

(a) $\mathcal{L}_{IPS-V2}(\theta)$ is an unbiased estimator of $\mathcal{L}_{ideal}(\theta)$.

(b) $\mathcal{L}_{DR-V2}(\theta)$ is an unbiased estimator of $\mathcal{L}_{ideal}(\theta)$, whether the imputed errors are accurate or not.

As discussed in Section 3.1, the unbiasedness of vanilla IPS and DR relies on the accuracy of the learned propensities. From Theorem 4.1, the proposed IPS-V2 and DR-V2 inherit the unbiasedness compared to vanilla IPS and DR under the same conditions, and the introduced constraint can enhance the balancing of the feature representations without sacrificing additional bias. Moreover, we show in Theorem 4.2 that the balancing constraints on IPS-V2 and DR-V2 can further reduce the variance compared with vanilla IPS and DR.

Theorem 4.2 (Variance Reduction of IPS-V2 and DR-V2)**.** *(a) Given imputed errors and learned propensities, the variance of $\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})$ reaches its minimum at*

$$\lambda_{opt} = \frac{2}{|\mathcal{D}|^2 \cdot \mathbb{V}(\textit{BMSE}(\boldsymbol{\phi}, \hat{p}))} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}^2} \text{Cov}\Big(e_{u,i}, \frac{1}{|\mathcal{D}|}\sum_{(s,t)\in\mathcal{D}}\left[\frac{1-o_{s,t}}{1-\hat{p}_{s,t}} - \frac{o_{s,t}}{\hat{p}_{s,t}}\right]\boldsymbol{\phi}(x_{u,i})^\top\boldsymbol{\phi}(x_{s,t})\Big),$$

where $\boldsymbol{o} = \{o_{u,i}|(u,i) \in \mathcal{D}\}$ is all the treatment indicators.

(b) $\mathcal{L}_{DR-V2}(\theta)$ has a smaller variance than $\mathcal{L}_{DR}(\theta)$,

$$\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})\Big|_{\lambda=\lambda_{opt}} = \left(1 - \rho_{L,B}^2\right)\mathbb{V}(\mathcal{L}_{DR}(\theta) \mid \boldsymbol{o}) \leq \mathbb{V}(\mathcal{L}_{DR}(\theta) \mid \boldsymbol{o}),$$

where $\rho_{L,B} = \text{Corr}\left(\mathcal{L}_{DR}(\theta), \textit{BMSE}(\boldsymbol{\phi}, \hat{p})\right)$, and similar results hold for $\mathcal{L}_{IPS-V2}(\theta)$.

4.2. Are Previous Regularizers Unbiased?

Several works have proposed estimators similar in form to the IPS-V2 and DR-V2, but with different regularization constraints (Swaminathan & Joachims, 2015b; Wang et al., 2021; Guo et al., 2021; Dai et al., 2022). For example, by using the bi-level optimization, Wang et al. (2021) adopts the sample variance (SV) regularization constraints[2]

$$\mathcal{L}_{IPS-SV}(\theta) = \mathcal{L}_{IPS}(\theta) + \lambda \cdot \mathcal{L}_{SV},$$
$$\mathcal{L}_{DR-SV}(\theta) = \mathcal{L}_{DR}(\theta) + \lambda \cdot \mathcal{L}_{SV},$$

where $\mathcal{L}_{SV} = \frac{1}{|\mathcal{D}|-1} \sum_{(u,i)\in\mathcal{D}} \left(\hat{p}_{u,i} - \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \hat{p}_{s,t}\right)^2$. There are other alternative regularizers, such as mean inverse square (MIS) (Wang et al., 2021)

$$\mathcal{L}_{MIS} = \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{1}{\hat{p}_{u,i}^2},$$

and the estimated variance of IPS and DR estimators (see Appendix B for more details) (Guo et al., 2021)

$$\hat{\mathbb{V}}(\mathcal{L}_{IPS}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{o_{u,i}\,(1-\hat{p}_{u,i})\,e_{u,i}^2}{\hat{p}_{u,i}^2},$$
$$\hat{\mathbb{V}}(\mathcal{L}_{DR}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{o_{u,i}(1-\hat{p}_{u,i})}{\hat{p}_{u,i}^2} \left(e_{u,i} - \hat{e}_{u,i}\right)^2.$$

However, in Proposition 4.3, we show all these regularizers are at the cost of introducing *additional bias* to reduce the variance of either the learned propensities or the estimator itself, and the regularized estimators are no longer unbiased.

Proposition 4.3 (Bias of Previous Regularizers)**.** *Regardless of whether the imputed errors or the learned propensities are accurate, the sample variance regularization is biased*

$$\mathbb{E}\left[\mathcal{L}_{DR-SV}(\theta)\right] = \mathbb{E}[\mathcal{L}_{DR}(\theta)] + \lambda \cdot \mathbb{E}[\mathcal{L}_{SV}] \neq \mathcal{L}_{ideal}(\theta),$$

and same for $\mathcal{L}_{IPS-SV}(\theta)$, as well as other regularizers.

4.3. Co-training Method

In practice, the interaction data available for training is usually sparse in RS, e.g., in rating prediction, items that a user interacted with only constitute a small fraction among the total item set (Schnabel et al., 2016; Wang et al., 2019a), and in post-click conversion rate prediction, the amount of training data in click-through rate (CTR) task is generally larger than that in conversion rate (CVR) task by $1 \sim 2$ order of magnitudes (Ma et al., 2018). In addition, the sparsity of the interacted data can cause the propensity-based reweighting methods to be less robust (Li et al., 2023e).

To address the above issues, previous studies suggested the use of multi-task learning by sharing the learned feature representations in the propensity (or CTR) model to the prediction (or CVR) model (Ma et al., 2018; Zhang et al., 2020; Wang et al., 2022). This would allow the latter to satisfy unbiased learning while further benefiting from the additional information obtained from parameter sharing, which would alleviate the data sparsity problem.

Specifically, let the prediction model be $\hat{r}_{u,i} = h\left(\boldsymbol{\phi}(x_{u,i})\right)$, where $\boldsymbol{\phi}(x_{u,i})$ is a representation layer to first obtain balanced representations for each user and item, followed by connecting a prediction head $h(\cdot)$. From Theorem 4.2(b), the proposed DR-V2 has a smaller variance compared with DR as long as $\rho_{L,B} = \mathrm{Corr}\left(\mathcal{L}_{DR}(\theta), \mathrm{BMSE}(\boldsymbol{\phi}, \hat{p})\right) \neq 0$, where θ denotes the parameters in $\boldsymbol{\phi}(\cdot)$ and $h(\cdot)$. We propose to co-train the propensity model and the prediction model by minimizing IPS-V2 with $\mathcal{L}_p$ in Section 3.2

$$\mathcal{L}_{Co-IPS}(\hat{r}, \hat{p}) = \mathcal{L}_{IPS-V2}(\hat{r}, \hat{p}) + \mathcal{L}_p(\hat{p}),$$

or minimizing DR-V2 with $\mathcal{L}_e, \mathcal{L}_p$ in Sections 3.1 and 3.2

$$\mathcal{L}_{Co-DR}(\hat{r}, \hat{p}, \hat{e}) = \mathcal{L}_{DR-V2}(\hat{r}, \hat{p}, \hat{e}) + \mathcal{L}_e(\hat{p}, \hat{e}) + \mathcal{L}_p(\hat{p}),$$

which has easy operability and permits both the balancing property of the representation (minimizing BMSE) and the unbiased prediction of the head (minimizing IPS-V2 or DR-V2 loss). We empirically show the advantages of the co-training in Section 5.

5. Experiments

5.1. Experimental Setup

Dataset and Preprocessing. Following the previous studies (Saito, 2020; Wang et al., 2019a; 2021; Chen et al., 2021), we conduct extensive experiments on two real-world datasets, **COAT**[3], **YAHOO! R3**[4], and a public large-scale industrial dataset, **PRODUCT**[5] (Gao et al., 2022). Specifically, **COAT** has 6,960 biased ratings and 4,640 unbiased ratings from 290 users to 300 items. **YAHOO! R3** has 311,704 biased ratings and 54,000 unbiased ratings from 15,400 users to 1,000 items. Both datasets are five-scale, and we binarize the ratings greater than three as 1, otherwise as 0. **PRODUCT** is collected from a short video sharing platform, and it is an almost fully exposed industrial dataset. There are 4,676,570 outcomes from 1,411 users on 3,327 items with a density of 99.6%. The video watching ratios that greater than two are denoted as 1, otherwise as 0.

Baselines. We take the matrix factorization (**MF**) (Koren et al., 2009) as the base model, and compare the proposed

[2] We use red color here and in Proposition 4.3 to emphasis the biasedness of previous regularizers.

[3] https://www.cs.cornell.edu/~schnabts/mnar/

[4] http://webscope.sandbox.Yahoo! R3.com/

[5] https://github.com/chongminggao/KuaiRec

Table 1. Recommendation performances in terms of AUC, Recall@5 (R@5), NDCG@5 (N@5) on COAT and YAHOO! R3. The best results are bolded, and second-best results are underlined.

	COAT			YAHOO! R3		
Method	AUC	R@5	N@5	AUC	R@5	N@5
Base model	0.749	0.546	0.499	0.723	0.719	0.553
+ IPS	0.760	0.567	0.511	0.722	0.724	0.551
+ DR	0.764	0.572	0.521	0.723	0.727	0.555
+ RD-IPS	0.763	0.570	0.516	0.731	0.731	0.571
+ RD-DR	0.766	0.574	0.534	0.732	0.735	0.574
+ ESMM	0.746	0.537	0.506	0.705	0.702	0.549
+ Multi-IPS	0.749	0.523	0.502	0.711	0.694	0.530
+ Multi-DR	0.755	0.551	0.534	0.712	0.684	0.516
+ ESCM^2-IPS	0.765	0.569	0.545	0.750	0.782	0.641
+ ESCM^2-DR	0.766	0.570	0.544	0.742	0.789	0.648
+ IPS-V2	**0.774**	**0.592**	0.579	0.755	0.801	0.655
+ DR-V2	**0.774**	0.590	**0.582**	**0.758**	**0.802**	**0.663**

Table 2. Recommendation performances in terms of AUC, Recall@50 (R@50), NDCG@50 (N@50) on PRODUCT. The best results of IPS and DR methods are bolded, respectively.

	PRODUCT					
Method	AUC	RI	R@50	RI	N@50	RI
Base model	0.830	-	0.842	-	0.578	-
+ ESMM	0.823	-0.84%	0.852	1.19%	0.563	-2.60%
+ IPS	0.826	-	0.849	-	0.574	-
+ RD-IPS	0.830	0.48%	0.873	2.83%	0.588	2.44%
+ Multi-IPS	0.810	-1.94%	0.875	3.06%	0.554	-3.48%
+ ESCM^2-IPS	0.833	0.85%	0.875	3.06%	0.598	4.18%
+ IPS-V2	**0.855**	**3.51%**	**0.895**	**5.42%**	**0.607**	**5.75%**
+ DR	0.832	-	0.866	-	0.582	-
+ RD-DR	0.834	0.24%	0.878	1.39%	0.587	0.86%
+ Multi-DR	0.829	-0.36%	0.859	-0.81%	0.562	-3.44%
+ ESCM^2-DR	0.834	0.24%	0.877	1.27%	0.596	2.41%
+ DR-V2	**0.853**	**2.52%**	**0.900**	**3.93%**	**0.608**	**4.47%**

Note: RI refers to the relative improvement over the corresponding baseline.

methods with the following approaches: **MF** (Koren et al., 2009), **IPS** (Schnabel et al., 2016), **DR** (Wang et al., 2019a; Saito, 2020), **RD-IPS** (Ding et al., 2022), **RD-DR** (Ding et al., 2022). We also compare with the following multi-task learning approaches: **ESMM** (Ma et al., 2018; Wen et al., 2020), **Multi-IPS** (Zhang et al., 2020), **Multi-DR** (Zhang et al., 2020), **ESCM^2-IPS** (Wang et al., 2022), and **ESCM^2-DR** (Wang et al., 2022).

Experimental Protocols and Details. We adopt three widely used evaluation metrics AUC, Recall@K (R@K), and NDCG@K (N@K) to measure the debiasing performance. We set K = 5 on COAT and YAHOO! R3, and K = 50 on PRODUCT. All the methods are implemented on PyTorch with Adam as the optimizer. We tune learning rate in $\{0.0005, 0.001, 0.005, 0.01\}$, weight decay in $\{0, 1e-6, 1e-5, ..., 1e-1\}$, and the regularization hyperparameter λ in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$.

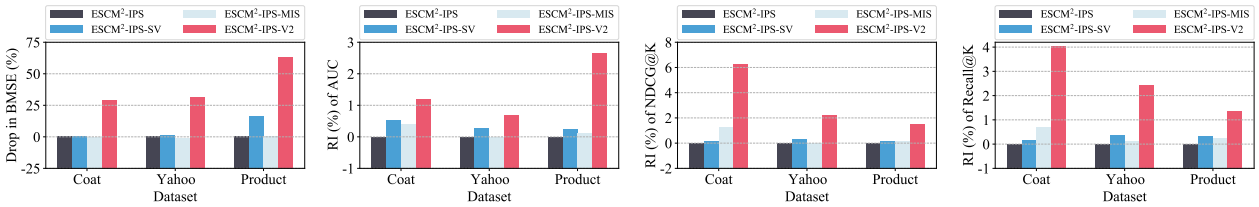
5.2. Performance Comparison

We train the prediction models with biased ratings and evaluate them with unbiased ratings on two widely used real-world datasets, COAT and YAHOO! R3, and the results are shown in Table 1. We have the following findings. First, most of the debiasing methods have better prediction performance compared to MF. ESCM^2-IPS and ESCM^2-DR are the most competitive baseline methods. The Multi-IPS and Multi-DR methods are even worse than MF on YAHOO! R3, which is explained by the lack of separate propensity and imputation model losses, leading to inaccurate learned propensities and imputed errors. Second, the proposed IPS-V2 and DR-V2 demonstrate the optimal performance in all three evaluation metrics. This is interpreted as the learned representations can be balanced by the learned propensities, which significantly enhances the debiasing ability.

Table 2 shows the performance of various debiasing methods on a large industrial dataset, PRODUCT. We compare the proposed IPS-V2 and DR-V2 methods with the previous IPS-based and DR-based methods, respectively. First, the ESMM and IPS-based methods do not have much improvement compared with MF, whereas the DR-based methods demonstrate stronger competitiveness, which may stem from the inaccurate learned propensities. Second, RD-IPS and RD-DR use robust methods to combat unobserved confounding, and show certain performance improvements compared with the vanilla IPS and DR methods. Finally, the proposed IPS-V2 and DR-V2 methods achieve the most significant improvement compare with all IPS-based and DR-based baseline methods in all metrics, while the ablated versions, ESCM^2-IPS and ESCM^2-DR methods, do not consider balancing representations, and lead to worse performance compare with the proposed methods. It further demonstrates the necessity of propensity balancing.

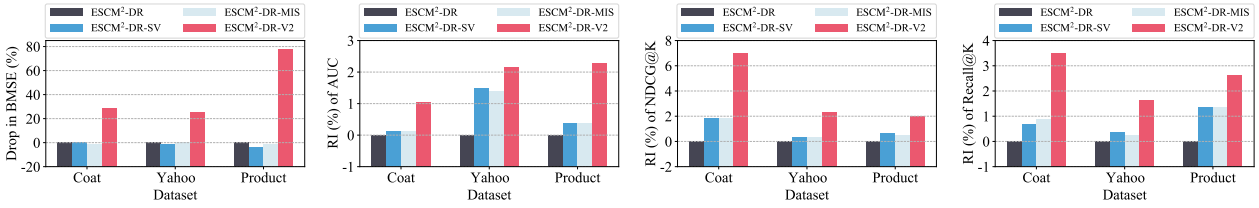
5.3. Ablation Study

Effects of Regularizers. We have discussed the biasedness of the previous regularizers in Section 4.2, and now we further investigate how the various regularizers affect BMSE and consequently the prediction performance, using AUC, NDCG@K, and Recall@K as evaluation metrics. Figure 1 shows the performance of using ESCM^2-IPS as backbone with the MIS, SV (see Section 4.2 for details), and BMSE as the regularizers on COAT, YAHOO! R3, and PRODUCT, respectively. Similarly, Figure 2 shows the corresponding results using ESCM^2-DR as backbone. ESCM^2-IPS and ESCM^2-DR serve as the most competitive baseline methods, introducing MIS, SV, and BMSE as regularizers can still further improve the performance in terms of AUC, NDCG@K, and Recall@K, as shown in Figures 1(b-d) and 2(b-d), respectively. However, there is no significant change in BMSE for the previous MIS and SV regularizers,



(a) IPS with BMSE (b) IPS with AUC (c) IPS with NDCG@K (d) IPS with Recall@K

Figure 1. Effects of **drop in BMSE (%)** on **RI (%) of AUC, NDCG@K, and Recall@K**, with varying regularizers on ESCM^2-IPS.



(a) DR with BMSE (b) DR with AUC (c) DR with NDCG@K (d) DR with Recall@K

Figure 2. Effects of **drop in BMSE (%)** on **RI (%) of AUC, NDCG@K, and Recall@K**, with varying regularizers on ESCM^2-DR.

Table 3. Varying balanced representation results on YAHOO! R3.

	YAHOO! R3					
Method	AUC	RI	R@5	RI	N@5	RI
ESCM^2-IPS	0.750	-	0.782	-	0.641	-
IPS-V2-1	**0.758**	**1.07%**	**0.801**	**2.43%**	0.652	1.72%
IPS-V2-K	0.754	0.53%	0.793	1.41%	0.647	0.94%
IPS-V2-2K	0.755	0.67%	0.797	1.92%	0.650	1.40%
IPS-V2-R	0.755	0.67%	**0.801**	**2.43%**	**0.655**	**2.18%**
ESCM^2-DR	0.742	-	0.789	-	0.648	-
DR-V2-1	0.758	2.16%	0.799	1.27%	0.662	2.16%
DR-V2-K	**0.760**	**2.43%**	**0.803**	**1.77%**	**0.664**	**2.47%**
DR-V2-2K	0.758	2.16%	0.802	1.65%	0.663	2.31%
DR-V2-R	0.758	2.16%	0.802	1.65%	0.663	2.31%

Note: RI refers to the relative improvement over the corresponding baseline.

Table 4. Varying balanced representation results on PRODUCT.

	PRODUCT					
Method	AUC	RI	R@50	RI	N@50	RI
ESCM^2-IPS	0.833	-	0.875	-	0.598	-
IPS-V2-1	0.850	2.04%	0.887	1.37%	0.599	0.17%
IPS-V2-K	0.847	1.68%	0.889	1.60%	0.603	0.84%
IPS-V2-2K	0.854	2.52%	0.894	2.17%	0.605	1.17%
IPS-V2-R	**0.855**	**2.64%**	**0.895**	**2.29%**	**0.607**	**1.51%**
ESCM^2-DR	0.834	-	0.877	-	0.596	-
DR-V2-1	0.852	2.16%	0.890	1.48%	0.602	1.01%
DR-V2-K	0.850	1.92%	0.894	1.94%	0.604	1.34%
DR-V2-2K	**0.854**	**2.40%**	**0.900**	**2.62%**	0.606	1.68%
DR-V2-R	0.853	2.28%	**0.900**	**2.62%**	**0.608**	**2.01%**

Note: RI refers to the relative improvement over the corresponding baseline.

as shown in Figure 1(a) and Figure 2(a). This is due to the fact that MIS and SV simply restrict the variance of the learned propensities without considering the balancing properties they should achieve. In contrast, the proposed IPS-V2 and DR-V2 methods use BMSE as regularizer have the most significant BMSE decreases and steadily have the most significant AUC, NDCG@K and Recall@K boosts. This empirically demonstrates that the reduction of BMSE contributes to the debiasing performance.

Effects of Balanced Representations. The proposed IPS-V2 and DR-V2 can improve the debiasing performance by reducing the BMSE, and we further explore the impact of varying choices of representation functions $\phi(x)$ as follows: (i) IPS/DR-V2-1: $\phi(x_{u,i}) = 1$; (ii) IPS/DR-V2-K: $\phi(x_{u,i}) = \mathbf{p}_u \odot \mathbf{q}_i \in \mathbb{R}^K$; (iii) IPS/DR-V2-2K: $\phi(x_{u,i}) = [\mathbf{p}_u : \mathbf{q}_i] \in \mathbb{R}^{2K}$; (iv) IPS/DR-V2-R: $\phi(x_{u,i}) = \hat{r}_{u,i} = f(u, i \mid \mathbf{p}_u, \mathbf{q}_i) \in \mathbb{R}$, where $\mathbf{p}_u$ and $\mathbf{q}_i$ denote the latent vector for user u and item i, respectively. For (i), the prediction model is set to $\hat{r}_{u,i} = f(x_{u,i})$ for comparison purpose, i.e., there is no representation layer, and for (ii)-(iv), we consider the prediction model $\hat{r}_{u,i} = h(\boldsymbol{\phi}(x_{u,i}))$ in Section 4.3. We take ESCM^2-IPS and ESCM^2-DR as the backbone of the prediction models on YAHOO! R3 and PRODUCT, and the results are shown in Table 3 and Table 4, respectively.

From Table 3 and Table 4, the proposed IPS-V2 and DR-V2 outperform in both ESCM^2-IPS and ESCM^2-DR settings for all $\boldsymbol{\phi}(x)$, achieving 2.64% AUC growth on the PRODUCT dataset. This is because the proposed BMSE regularization adds an additional constraint on the learned propensities, i.e., balancing $\boldsymbol{\phi}(x)$. Recall that we showed in Theorem 4.2 that when vanilla IPS and DR losses and $\boldsymbol{\phi}(x)$ corresponding to BMSE$(\boldsymbol{\phi}, \hat{p})$ are correlated, IPS-V2 and DR-V2 will lead to smaller variances, thus improving the generalization ability. Notably, this is empirically easy to be satisfied, since the representation $\boldsymbol{\phi}(x)$ in the prediction model $\hat{r}_{u,i} =$

Table 5. Effects of balancing hyper-parameter λ on IPS-V2.

IPS-V2	YAHOO! R3				PRODUCT			
λ	R@5	RI	N@5	RI	R@50	RI	N@50	RI
0	0.782	-	0.641	-	0.875	-	0.598	-
0.001	0.796	1.79%	0.651	1.56%	0.895	2.29%	0.605	1.17%
0.005	0.800	2.30%	0.654	2.03%	0.895	2.29%	0.606	1.34%
0.01	0.801	2.43%	**0.655**	**2.18%**	**0.896**	**2.40%**	0.606	1.34%
0.05	0.800	2.30%	0.652	1.72%	0.893	2.06%	0.603	0.84%
0.1	0.801	2.43%	0.650	1.40%	0.895	2.29%	0.607	1.51%
0.5	0.804	2.81%	0.653	1.87%	0.895	2.29%	0.605	1.17%
1	**0.805**	**2.94%**	0.654	2.03%	0.893	2.06%	**0.608**	**1.67%**

Note: RI refers to the relative improvement over the baseline with $\lambda = 0$.

Table 6. Effects of balancing hyper-parameter λ on DR-V2.

DR-V2	YAHOO! R3				PRODUCT			
λ	R@5	RI	N@5	RI	R@50	RI	N@50	RI
0	0.789	-	0.648	-	0.877	-	0.596	-
0.001	0.801	1.52%	0.662	2.16%	0.895	2.05%	0.604	1.34%
0.005	**0.802**	**1.65%**	**0.663**	**2.31%**	**0.900**	**2.62%**	0.606	1.68%
0.01	0.797	1.01%	0.661	2.01%	0.899	2.51%	0.606	1.68%
0.05	0.797	1.01%	0.659	1.70%	0.898	2.39%	0.607	1.85%
0.1	0.795	0.76%	0.658	1.54%	**0.900**	**2.62%**	0.608	2.01%
0.5	0.794	0.63%	0.655	1.08%	**0.900**	**2.62%**	0.607	1.85%
1	0.792	0.38%	0.650	0.31%	0.899	2.51%	**0.609**	**2.18%**

Note: RI refers to the relative improvement over the baseline with $\lambda = 0$.

$h\left(\boldsymbol{\phi}(x_{u,i})\right)$ is also included in BMSE$(\boldsymbol{\phi}, \hat{p})$, resulting in the correlation between the two.

5.4. In-depth Analysis

It is now clear that BMSE regularization plays an important role in the proposed IPS-V2 and DR-V2, so it is meaningful to analyze the impact of the regularization hyperparameter λ on the debiasing performance. Table 5 and Table 6 show the effects of varying BMSE constraint strength λ on IPS-V2 and DR-V2 on YAHOO! R3 and PRODUCT, respectively. It can be seen that the methods with BMSE as the regularization stably outperform the vanilla IPS and DR methods, and the optimal performance is reached at proper constraint strengths (about 0.005-0.01) for IPS-V2 and DR-V2.

6. Related Work

Biases are prevalent in the collected interactions in RS and have received much attention in recent years (Chen et al., 2022; Wu et al., 2022). Many methods were developed to eliminate biases and improve the prediction performance for different tasks in RS, such as rating prediction (Marlin & Zemel, 2009; Schnabel et al., 2016; Wang et al., 2019a; 2020b; Huang et al., 2022), post-view click-through rate (CTR) prediction (Yuan et al., 2019), post-click conversion rate (CVR) prediction (Ma et al., 2018; Zhang et al., 2020; Guo et al., 2021; Dai et al., 2022; Wang et al., 2022), and uplift modeling (Saito et al., 2019; Sato et al., 2019; 2020). A common question faced in these tasks is "what would the feedback be if an intervention is made to a user", which can be formulated as a missing data problem (Wang et al., 2020a; 2023; Li et al., 2023a;b;c;e), where the distribution of observed events is different from that of missing events.

To address this problem, the error imputation-based (EIB) methods (Steck, 2010; Hernández-Lobato et al., 2014) tried to construct a sample of all events by imputing the missing events and then training the recommendation model on them. However, due to the difficulty of accurate imputations, the EIB methods usually have sub-optimal performance empirically (Guo et al., 2021). Schnabel et al. (2016) proposed inverse probability weighting (IPS) methods for debiasing, which aims to recover the distribution of all events by reweighing the observed events with inverse propensities. Wang et al. (2019a) further proposed a doubly robust (DR) approach by combing EIB and IPS methods. Due to the competitive performance and theoretical guarantees, IPS and DR methods have gained widespread attention and have inspired a series of new variants (Bonner & Vasile, 2018; Liu et al., 2020; Chen et al., 2021; Saito, 2020; Wang et al., 2021; Guo et al., 2021; Dai et al., 2022; Ding et al., 2022; Li et al., 2023a;b;c;e). The above methods are also known as propensity-based weighting methods. However, the true propensity is unknown and needs to be estimated.

Empirically, previous works have found that the estimation of propensity is critical for debiasing performance (Schnabel et al., 2016; Wang et al., 2021) and tried different strategies for learning propensities, such as Naive Bayes (Schnabel et al., 2016), Logistic Regression (Schnabel et al., 2016; Li et al., 2023d), Poisson Factorization (Wang et al., 2020a), Multi-task Learning (Zhang et al., 2020; Wang et al., 2022; Li et al., 2023f), Variance Regularization (Wang et al., 2021; Guo et al., 2021; Dai et al., 2022), and Minimax (Ding et al., 2022). However, there is a lack of theoretically guaranteed criterion for estimating propensity. In this paper, we propose a new metric to measure the quality of learned propensities and propose two estimators IPS-V2 and DR-V2.

7. Conclusion

In this paper, we propose a principled approach to measure and enhance the balancing property of propensity-based methods for debiased recommendations. First, we propose a metric, called BMSE, to measure the balancing property of learned propensities, and theoretically discuss its importance for unbiased learning. Then, we propose IPS-V2 and DR-V2 as unbiased loss estimators with BMSE as the regularization term, and theoretically show that IPS-V2 and DR-V2 have greater propensity balancing and smaller variance without sacrificing additional bias. Then, we further propose to use IPS-V2 or DR-V2 to co-train the propensity and the prediction model, which permits both the propensity balancing property and the unbiased prediction. Extensive experiments are conducted on two real-world and a large-scale industrial dataset, and the results show that our method has practical benefits while being easy to operate.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (No. 62141607, U1936219), National Key R&D Program of China (No. 2018AAA0102004). Peng Wu was supported by the Disciplinary Funding of Beijing Technology and Business University.

References

Bonner, S. and Vasile, F. Causal embeddings for recommendation. In *RecSys*, 2018.

Chen, J., Dong, H., Qiu, Y., He, X., Xin, X., Chen, L., Lin, G., and Yang, K. Autodebias: Learning to debias for recommendation. In *SIGIR*, 2021.

Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., and He, X. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 2022.

Dai, Q., Li, H., Wu, P., Dong, Z., Zhou, X.-H., Zhang, R., He, X., Zhang, R., and Sun, J. A generalized doubly robust learning framework for debiasing post-click conversion rate prediction. In *KDD*, 2022.

Ding, S., Wu, P., Feng, F., He, X., Wang, Y., Liao, Y., and Zhang, Y. Addressing unmeasured confounder for recommendation with sensitivity analysis. In *KDD*, 2022.

Gao, C., Li, S., Lei, W., Chen, J., Li, B., Jiang, P., He, X., Mao, J., and Chua, T.-S. Kuairec: A fully-observed dataset and insights for evaluating recommender systems. In *CIKM*, 2022.

Guo, S., Zou, L., Liu, Y., Ye, W., Cheng, S., Wang, S., Chen, H., Yin, D., and Chang, Y. Enhanced doubly robust learning for debiasing post-click conversion rate estimation. In *SIGIR*, 2021.

Hernán, M. A. and Robins, J. M. *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC, 2020.

Hernández-Lobato, J. M., Houlsby, N., and Ghahramani, Z. Probabilistic matrix factorization with non-random missing data. In *ICML*, 2014.

Huang, J., Oosterhuis, H., and de Rijke, M. It is different when items are older: Debiasing recommendations when selection bias and user preferences are dynamic. In *WSDM*, 2022.

Imai, K. and Ratkovic, M. Covariate balancing propensity score. *Journal of the Royal Statistical Society (Series B)*, 76(1):243–263, 2014.

Imbens, G. W. and Rubin, D. B. *Causal Inference For Statistics Social and Biomedical Science*. Cambridge University Press, 2015.

Joachims, T., Swaminathan, A., and Schnabel, T. Unbiased learning-to-rank with biased feedback. In *WSDM*, 2017.

Koren, Y., Bell, R., and Volinsky, C. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37, 2009.

Li, H., Dai, Q., Li, Y., Lyu, Y., Dong, Z., Zhou, X.-H., and Wu, P. Multiple robust learning for recommendation. In *AAAI*, 2023a.

Li, H., Lyu, Y., Zheng, C., and Wu, P. TDR-CL: Targeted doubly robust collaborative learning for debiased recommendations. In *ICLR*, 2023b.

Li, H., Xiao, Y., Zheng, C., and Wu, P. Balancing unobserved confounding with a few unbiased ratings in debiased recommendations. In *WWW*, 2023c.

Li, H., Zheng, C., Cao, Y., Geng, Z., Liu, Y., and Wu, P. Trustworthy policy learning under the counterfactual no-harm criterion. In *ICML*, 2023d.

Li, H., Zheng, C., and Wu, P. StableDR: Stabilized doubly robust learning for recommendation on data missing not at random. In *ICLR*, 2023e.

Li, H., Zheng, C., Wu, P., Kuang, K., Liu, Y., and Cui, P. Who should be given incentives? counterfactual optimal treatment regimes learning for recommendation. In *KDD*, 2023f.

Liu, D., Cheng, P., Dong, Z., He, X., Pan, W., and Ming, Z. A general knowledge distillation framework for counterfactual recommendation via uniform data. In *SIGIR*, 2020.

Ma, X., Zhao, L., Huang, G., Wang, Z., Hu, Z., Zhu, X., and Gai, K. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *SIGIR*, 2018.

Marlin, B. M. and Zemel, R. S. Collaborative prediction and ranking with non-random missing data. In *RecSys*, 2009.

Neyman, J. S. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5:465–472, 1990.

Rosenbaum, P. R. *Design of Observational Studies*. Springer Nature Switzerland AG, second edition, 2020.

Rosenbaum, P. R. and Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983.

Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational psychology*, 66:688–701, 1974.

Saito, Y. Unbiased pairwise learning from implicit feedback. In *NeurIPS Workshop*, 2019.

Saito, Y. Doubly robust estimator for ranking metrics with post-click conversions. In *RecSys*, pp. 92–100, 2020.

Saito, Y. Asymmetric tri-training for debiasing missing-not-at-random explicit feedback. In *SIGIR*, 2020.

Saito, Y. Doubly robust estimator for ranking metrics with post-click conversions. In *RecSys*, 2020.

Saito, Y., Sakata, H., and Nakata, K. Doubly robust prediction and evaluation methods improve uplift modeling for observational data. In *SIAM*, 2019.

Saito, Y., Yaginuma, S., Nishino, Y., Sakata, H., and Nakata, K. Unbiased recommender learning from missing-not-at-random implicit feedback. In *WSDM*, 2020.

Sant'Anna, P. H. C., Song, X., and Xu, Q. Covariate distribution balance via propensity scores. *Journal of Applied Econometrics*, 37(6):1093–1120, 2022.

Sato, M., Singh, J., Takemori, S., Sonoda, T., Zhang, Q., and Ohkuma, T. Uplift-based evaluation and optimization of recommenders. In *RecSys*, 2019.

Sato, M., Takemori, S., Singh, J., and Ohkuma, T. Unbiased learning for the causal effect of recommendation. In *RecSys*, 2020.

Schnabel, T., Swaminathan, A., Singh, A., Chandak, N., and Joachims, T. Recommendations as treatments: Debiasing learning and evaluation. In *ICML*, 2016.

Steck, H. Training and testing of recommender systems on data missing not at random. In *KDD*, 2010.

Swaminathan, A. and Joachims, T. The self-normalized estimator for counterfactual learning. In *NeurIPS*, 2015a.

Swaminathan, A. and Joachims, T. Counterfactual risk minimization: Learning from logged bandit feedback. In *ICML*, 2015b.

Wang, H., Chang, T.-W., Liu, T., Huang, J., Chen, Z., Yu, C., Li, R., and Chu, W. ESCM^2: Entire space counterfactual multi-task model for post-click conversion rate estimation. In *SIGIR*, 2022.

Wang, W., Zhang, Y., Li, H., Wu, P., Feng, F., and He, X. Causal recommendation: Progresses and future directions. In *Tutorial on SIGIR*, 2023.

Wang, X., Zhang, R., Sun, Y., and Qi, J. Doubly robust joint learning for recommendation on data missing not at random. In *ICML*, 2019a.

Wang, X., Zhang, R., Sun, Y., and Qi, J. Combating selection biases in recommender systems with a few unbiased ratings. In *WSDM*, 2021.

Wang, Y., Liang, D., Charlin, L., and Blei, D. M. The deconfounded recommender: A causal inference approach to recommendation. *arXiv:1808.06581*, 2019b.

Wang, Y., Liang, D., Charlin, L., and Blei, D. M. Causal inference for recommender systems. In *RecSys*, 2020a.

Wang, Z., Chen, X., Wen, R., Huang, S.-L., Kuruoglu, E. E., and Zheng, Y. Information theoretic counterfactual learning from missing-not-at-random feedback. *NeurIPS*, 2020b.

Wen, H., Zhang, J., Wang, Y., Lv, F., Bao, W., Lin, Q., and Yang, K. Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction. In *SIGIR*, 2020.

Wong, R. K. W. and Chan, K. C. G. Kernel-based covariate functional balancing for observational studies. *Biometrika*, 105(1):199–213, 2018.

Wu, P., Li, H., Deng, Y., Hu, W., Dai, Q., Dong, Z., Sun, J., Zhang, R., and Zhou, X.-H. On the opportunity of causal learning in recommendation systems: Foundation, estimation, prediction and challenges. In *IJCAI*, 2022.

Yuan, B., Hsia, J.-Y., Yang, M.-Y., Zhu, H., Chang, C.-Y., Dong, Z., and Lin, C.-J. Improving ad click prediction by considering non-displayed events. In *CIKM*, 2019.

Zhang, W., Bao, W., Liu, X., Yang, K., Lin, Q., Wen, H., and Ramezani, R. Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning. In *WWW*, 2020.

A. Details and Examples of Potential Outcomes Formalization

Let $\mathcal{U} = \{u\}$ be the set of users, $\mathcal{I} = \{i\}$ be the set of items. To define a causal problem, the widely adopted potential outcome framework (Rubin, 1974; Neyman, 1990) consists of the following key elements: (1) *Target population*: the set of all user-item pairs $\mathcal{D} = \{(u, i) \mid u \in \mathcal{U}, i \in \mathcal{I}\}$; (2) *Feature*: $x_{u,i}$, the feature of user u and item i; (3) *Treatment*: $o_{u,i} \in \{0, 1\}$, it has different implications for different prediction tasks in RS, e.g., whether the true rating $r_{u,i}$ is observed ($o_{u,i} = 1$) or missing ($o_{u,i} = 0$), whether item i is exposed to user u, and whether user u clicks item i; (4) *Outcome*: $r_{u,i}$, the feedback of user-item pair (u, i), e.g., rating, click indicator, conversion indicator; (5) *Potential outcome*: $r_{u,i}(o)$ for $o \in \{0, 1\}$, it is the outcome that would be observed if $o_{u,i}$ had been set to o.

Example 1 (Rating prediction). The feedback $r_{u,i}$ is the true rating of user u for item i. However, the rating suffers the problem of missing not at random. Let $o_{u,i}$ be the observing indicator of $r_{u,i}$. If we consider the observing indicator as the treatment, then $r_{u,i}(1)$ denotes the true rating of user u for item i if $o_{u,i} = 1$. Our goal is to predict $r_{u,i}(1)$ for all $(u, i) \in \mathcal{D}$ (Schnabel et al., 2016; Wang et al., 2019b; 2020a; Li et al., 2023a;b;c;e).

Example 2 (Post-view click-through rate (CTR) prediction). The treatment $o_{u,i} = 1$ if item i has been exposed to user u, $o_{u,i} = 0$ otherwise, the feedback $r_{u,i} = 1$ or 0 indicates whether user u clicked item i or not. Then $\mathbb{E}[r_{u,i}(1)|x_{u,i}] = \mathbb{P}(r_{u,i}(1) = 1|x_{u,i})$ denotes the CTR.

Example 3 (Post-click conversion rate (CVR) prediction). The treatment $o_{u,i} = 1$ if user u has clicked item i, $o_{u,i} = 0$ otherwise. The feedback $r_{u,i}$ is the indicator of the observed conversion label of user u on item i. Then $\mathbb{E}[r_{u,i}(1)|x_{u,i}] = \mathbb{P}(r_{u,i}(1) = 1|x_{u,i})$ denotes the CVR (Zhang et al., 2020; Guo et al., 2021; Dai et al., 2022).

B. Proofs of Proposition 3.1, Theorem 4.1, Theorem 4.2, and Proposition 4.3

Proposition 3.1 (Balancing Property). *If $\hat{p}_{u,i}$ estimates $p_{u,i}$ accurately, i.e, $\hat{p}_{u,i} = p_{u,i}$, then for any integrable vector-valued functions $\boldsymbol{\phi}(x)$, BMSE$(\boldsymbol{\phi}, \hat{p}) \to 0$ almost surely.*

Proof of Proposition 3.1. Let $\phi_k(x)$ be the k-th components of $\boldsymbol{\phi}(x)$, $k = 1, \ldots, K$. It suffices to show that

$$\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1 - o_{u,i}}{1 - \hat{p}_{u,i}}\right] \phi_k(x_{u,i}) = 0, \quad \text{almost surely.}$$

This follows immediately by the balancing property of true propensity, i.e.,

$$\mathbb{E}\left[\frac{o_{u,i}\phi_k(x_{u,i})}{p_{u,i}}\right] = \mathbb{E}\left[\frac{(1 - o_{u,i})\phi_k(x_{u,i})}{1 - p_{u,i}}\right] = \mathbb{E}[\phi_k(x_{u,i})],$$

and

$$\text{BMSE}(\boldsymbol{\phi}, p) = \left|\left|\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\frac{o_{u,i}}{p_{u,i}} - \frac{1 - o_{u,i}}{1 - p_{u,i}}\right] \boldsymbol{\phi}(x_{u,i})\right|\right|_F^2 \to 0, \quad \text{almost surely.}$$

□

Theorem 4.1 (Unbiasedness of IPS-V2 and DR-V2). *When learned propensities are accurate for all user-item pairs,*

(a) $\mathcal{L}_{IPS-V2}(\theta)$ is an unbiased estimator of $\mathcal{L}_{ideal}(\theta)$.

(b) $\mathcal{L}_{DR-V2}(\theta)$ is an unbiased estimator of $\mathcal{L}_{ideal}(\theta)$, whether the imputed errors are accurate or not.

Proof of Theorem 4.1. When learned propensities are accurate, the balancing constraint

$$\left|\left|\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1 - o_{u,i}}{1 - \hat{p}_{u,i}}\right] \boldsymbol{\phi}(x_{u,i})\right|\right|_F^2$$

will be zero almost surely according to Proposition 3.1. In this case, the IPS-V2/DR-V2 is equivalent to the vanilla IPS/DR estimator and thus unbiased.

□

Theorem 4.2 (Variance Reduction of IPS-V2 and DR-V2).

(a) Given imputed errors and learned propensities, the variance of $\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})$ *reaches its minimum at*

$$\lambda_{opt} = \frac{2}{|\mathcal{D}|^2 \cdot \mathbb{V}(\textit{BMSE}(\boldsymbol{\phi}, \hat{p}))} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}^2} \operatorname{Cov}\Big(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{1-o_{s,t}}{1-\hat{p}_{s,t}} - \frac{o_{s,t}}{\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\Big),$$

where $\boldsymbol{o} = \{o_{u,i} | (u,i) \in \mathcal{D}\}$ *is all the treatments.*

(b) $\mathcal{L}_{DR-V2}(\theta)$ *has a smaller variance than* $\mathcal{L}_{DR}(\theta)$*,*

$$\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})\Big|_{\lambda=\lambda_{opt}} = \big(1-\rho_{L,B}^2\big)\, \mathbb{V}(\mathcal{L}_{DR}(\theta) \mid \boldsymbol{o}) \le \mathbb{V}(\mathcal{L}_{DR}(\theta) \mid \boldsymbol{o}),$$

where $\rho_{L,B} = \operatorname{Corr}\left(\mathcal{L}_{DR}(\theta), \textit{BMSE}(\boldsymbol{\phi}, \hat{p})\right)$*, and similar results hold for* $\mathcal{L}_{IPS-V2}(\theta)$*.*

Proof of Theorem 4.2. The balancing-enhanced IPS estimator is formulated as

$$\begin{aligned}
\mathcal{L}_{IPS-V2}(\theta) &= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}} + \lambda \cdot \Big|\Big| \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \boldsymbol{\phi}(x_{u,i}) \Big|\Big|_F^2, \\
&= \frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}} + \lambda \cdot \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})
\end{aligned}$$

where λ is a tuning parameter. The variance of IPS-V2 estimator is given by

$$\begin{aligned}
&\mathbb{V}(\mathcal{L}_{IPS-V2}(\theta) \mid \boldsymbol{o}) = \mathbb{V}(\mathcal{L}_{IPS}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ 2\lambda \cdot \operatorname{Cov}\left(\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}}, \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t}) \right) \\
&= \mathbb{V}(\mathcal{L}_{IPS}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ \frac{4\lambda}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \operatorname{Cov}\left(\frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}}, \frac{1}{|\mathcal{D}|} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t}) \right) \\
&= \mathbb{V}(\mathcal{L}_{IPS}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ \frac{4\lambda}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \operatorname{Cov}\left(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t}) \right)
\end{aligned}$$

Next, note that $o_{u,i}$ and $o_{s,t}$ are binary variables, then $o_{u,i}^2 = o_{u,i}$ and $o_{u,i}(1-o_{u,i}) = 0$, we have

$$\begin{aligned}
\mathbb{V}(\mathcal{L}_{IPS-V2}(\theta) \mid \boldsymbol{o}) &= \mathbb{V}(\mathcal{L}_{IPS}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&- \frac{4\lambda}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}^2} \operatorname{Cov}\left(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{1-o_{s,t}}{1-\hat{p}_{s,t}} - \frac{o_{s,t}}{\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t}) \right).
\end{aligned}$$

Then $\mathbb{V}(\mathcal{L}_{IPS-V2}(\theta) \mid \boldsymbol{o})$ is a quadratic function in λ, and reaches its optimum when $\lambda = \lambda_{opt}$ that

$$\lambda_{opt} = \frac{2}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}^2} \operatorname{Cov}\left(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{1-o_{s,t}}{1-\hat{p}_{s,t}} - \frac{o_{s,t}}{\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t}) \right) \Big/ \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})),$$

and at λ_{opt}, the minimum variance of $\mathbb{V}(\mathcal{L}_{IPS-V2}(\theta) \mid \boldsymbol{o})$ equals to

$$\mathbb{V}(\mathcal{L}_{IPS-V2}(\theta) \mid \boldsymbol{o}) = \big(1-\rho_{L,B}^2\big)\, \mathbb{V}(\mathcal{L}_{IPS}(\theta)) \le \mathbb{V}(\mathcal{L}_{IPS}(\theta)),$$

where $\rho_{L,B} = \text{Corr}\left(\mathcal{L}_{IPS}(\theta), \text{BMSE}(\boldsymbol{\phi}, \hat{p})\right)$.

Similarly, the variance of DR-V2 estimator is given by

$$
\begin{aligned}
&\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o}) = \mathbb{V}(\mathcal{L}_{DR}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ 2\lambda \cdot \text{Cov}\left(\mathcal{L}_{DR}(\theta), \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\right) \\
&= \mathbb{V}(\mathcal{L}_{DR}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ 2\lambda \cdot \text{Cov}\left(\frac{1}{|\mathcal{D}|} \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}}, \frac{1}{|\mathcal{D}|^2} \sum_{(u,i)\in\mathcal{D}} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\right) \\
&= \mathbb{V}(\mathcal{L}_{DR}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ \frac{4\lambda}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \text{Cov}\left(\frac{o_{u,i} e_{u,i}}{\hat{p}_{u,i}}, \frac{1}{|\mathcal{D}|} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \cdot \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\right) \\
&= \mathbb{V}(\mathcal{L}_{DR}(\theta)) + \lambda^2 \cdot \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})) \\
&+ \frac{4\lambda}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}} \left[\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right] \text{Cov}\left(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{o_{s,t}}{\hat{p}_{s,t}} - \frac{1-o_{s,t}}{1-\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\right)
\end{aligned}
$$

The second equality holds due to $\hat{e}_{u,i}$ and $\hat{p}_{u,i}$ are given without randomness. Then $\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})$ is a quadratic function in λ, and reaches its optimum when $\lambda = \lambda_{opt}$ that

$$
\lambda_{opt} = \frac{2}{|\mathcal{D}|^2} \cdot \sum_{(u,i)\in\mathcal{D}} \frac{o_{u,i}}{\hat{p}_{u,i}^2} \text{Cov}\left(e_{u,i}, \frac{1}{|\mathcal{D}|} \sum_{(s,t)\in\mathcal{D}} \left[\frac{1-o_{s,t}}{1-\hat{p}_{s,t}} - \frac{o_{s,t}}{\hat{p}_{s,t}}\right] \boldsymbol{\phi}(x_{u,i})^\top \boldsymbol{\phi}(x_{s,t})\right) \Big/ \mathbb{V}(\text{BMSE}(\boldsymbol{\phi}, \hat{p})),
$$

and at λ_{opt}, the minimum variance of $\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o})$ equals to

$$
\mathbb{V}(\mathcal{L}_{DR-V2}(\theta) \mid \boldsymbol{o}) = \left(1 - \rho_{L,B}^2\right) \mathbb{V}(\mathcal{L}_{DR}(\theta)) \leq \mathbb{V}(\mathcal{L}_{DR}(\theta)),
$$

where $\rho_{L,B} = \text{Corr}\left(\mathcal{L}_{DR}(\theta), \text{BMSE}(\boldsymbol{\phi}, \hat{p})\right)$. □

Proposition 4.3 (Bias of Previous Regularizers). *Regardless of whether the imputed errors or the learned propensities are accurate, the sample variance regularization is biased*

$$
\mathbb{E}\left[\mathcal{L}_{DR-SV}(\theta)\right] = \mathbb{E}[\mathcal{L}_{DR}(\theta)] + \lambda \cdot \mathbb{E}[\mathcal{L}_{SV}] \neq \mathcal{L}_{ideal}(\theta),
$$

and same for $\mathcal{L}_{IPS-SV}$, as well as other regularizers.

Proof of Proposition 4.3. We first show that the estimated variance of IPS and DR estimators are given as

$$
\hat{\mathbb{V}}(\mathcal{L}_{IPS}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{o_{u,i}\left(1-\hat{p}_{u,i}\right) e_{u,i}^2}{\hat{p}_{u,i}^2}, \quad \hat{\mathbb{V}}(\mathcal{L}_{DR}(\theta)) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{o_{u,i}(1-\hat{p}_{u,i})}{\hat{p}_{u,i}^2} \left(e_{u,i} - \hat{e}_{u,i}\right)^2.
$$

In fact, for IPS estimator, we have

$$
\mathbb{V}_{\mathcal{O}}\left(\mathcal{L}_{IPS}(\theta)\right) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{p_{u,i}\left(1-p_{u,i}\right) e_{u,i}^2}{\hat{p}_{u,i}^2} \approx \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{o_{u,i}\left(1-\hat{p}_{u,i}\right) e_{u,i}^2}{\hat{p}_{u,i}^2},
$$

and similarly for the DR estimator

$$
\mathbb{V}_{\mathcal{O}}\left(\mathcal{L}_{DR}(\theta)\right) = \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} \frac{p_{u,i}\left(1-p_{u,i}\right)\left(e_{u,i} - \hat{e}_{u,i}\right)^2}{\hat{p}_{u,i}^2} \approx \frac{1}{|\mathcal{D}|^2} \sum_{u,i\in\mathcal{D}} o_{u,i} \frac{1-\hat{p}_{u,i}}{\hat{p}_{u,i}^2} \left(e_{u,i} - \hat{e}_{u,i}\right)^2.
$$

Then Proposition 4.3 follows immediately from the fact that $\mathcal{L}_{SV}$, $\mathcal{L}_{MIS}$, $\hat{\mathbb{V}}_{\mathcal{O}}\left(\mathcal{L}_{IPS}(\theta)\right)$, and $\hat{\mathbb{V}}_{\mathcal{O}}\left(\mathcal{L}_{DR}(\theta)\right)$ always does not converge to zero, regardless of whether the learned propensities or the imputed errors are accurate.

□