Leveraging Instrumental Variables in Online Advertising Auctions : Robust Click-Through-Rate Prediction

Ryohei Emori
Keio University / Riken AIP Center
Tokyo, Japan
ryohey3569@keio.jp

Shinya Suzumura
LY Corporation
Tokyo, Japan
ssuzumur@lycorp.co.jp

Nobuyuki Shimizu
LY Corporation
Tokyo, Japan
nobushim@lycorp.co.jp

Takahiro Hoshino
Keio University / Riken AIP Center
Tokyo, Japan
hoshino@econ.keio.ac.jp

ABSTRACT

Predicting the click-through rate (CTR) in online ad auctions is essential for calculating bid amounts and forming rankings. However, predicting CTR from historical data faces some difficulties, one of which is the cold-start problem. Our research uses the instrumental variables (IVs) framework to address the cold-start problem and selection bias, validating robust CTR prediction in online advertising auctions. Although generally identifying IVs in wide applications is notably challenging, their potential use is not limited to CTR prediction; they can potentially be used to address practical issues and research questions in advertising auctions in general. We put forth bid amounts as IVs, discussing their validity as IVs and testing the robustness of predictions using IVs in both simulations and real data scenarios. Moreover, we enhanced our methodology by integrating explicit interactions between bid amounts and other features, demonstrating that accounting for heterogeneity in IVs significantly improves prediction accuracy in actual data. Our proposal on IVs and its refined CTR prediction approach enriches the research fields on causal inference robustness and invariant prediction.

CCS CONCEPTS

• **Information systems → Computational advertising**.

KEYWORDS

Instrumental Variables, Omitted Variable Bias, Robustness, Cold-start Problem, Click-Through-Rate, Online Advertising Auction

ACM Reference Format:

Ryohei Emori, Shinya Suzumura, Nobuyuki Shimizu, and Takahiro Hoshino. 2024. Leveraging Instrumental Variables in Online Advertising Auctions : Robust Click-Through-Rate Prediction. In *30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, AdKDD Workshop 2024 (AdKDD'24).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

AdKDD'24, August 2024, Barcelona, Spain

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Online advertising, an essential backbone of the digital economy, relies heavily on accurate prediction models to allocate ads effectively and enhance the user experience. Crucially, the accuracy of click-through rate (CTR) prediction plays a pivotal role in determining the success in terms of welfare of of online advertising auctions, and at the same time, hover the potential biases that may skew results [4, 11].

In addition to the problem of bias that lurks in some online ad auctions and is often the subject of research, the cold-start problem arises when we must make predictions for new advertisements or infrequent users, leading to decreased predictive accuracy. Against the backdrop of problems arising from those various factors, causal methods of predicting user behavior that capture invariant user behavior have risen as a subject of high research interest [3, 5, 9]. Among them, prior research [3] has highlighted that one of those causal methods, the instrumental variables (IVs) method, has the potential to contribute to solving the cold-start problem. [8] provided a methodology for IVs using neural networks, but specific IVs always need to be identified in a specific research domain. [13] uses the user's search query as an instrumental variable; their use of IVs is limited to search advertising and may not satisfy one of the conditions for IVs, the exclusion restriction.

In this paper, we identify bid amounts as IVs in online ad auction settings and demonstrate that click prediction using the IVs method exhibits robust predictions in the overall prediction and cold start problems.

Although IVs are generally considered difficult to identify, they have the potential to: 1) maximize the use of data, including impressions of ads with low historical win rates; 2) not require random impressions of ads; 3) avoid assumptions that often lead to erroneous predictions due to the unrealistic absence of unobserved confounding factors between treatment and outcome relationships [10]; and 4) potentially infer the causal effect of impressions on conversion as well as clicks.

Furthermore, we demonstrate that the explicit use of first-stage heterogeneity in the IVs method can be strongly recommended in online ad auctions [1, 2]. First-stage heterogeneity in the IVs method has been relatively overlooked compared to heterogeneity in the second stage, namely, user response. However, we find that increasing the association between IVs and impression probability shows robust predictions for the overall prediction and the cold-start problem.

The contributions of the paper have three main points:

(1) We identify and propose valid IVs tailored to online advertising auctions. The IVs suit broad advertising auction contexts, including display and search advertising. Furthermore, the IVs method is expected to have further applications such as causal inference of medium- and long-term effects of ad impressions on conversions, etc., not limited to causal effects on user click behavior in online ad auctions.
(2) There have been few empirical examples the IVs method has been demonstrated to be capable of making invariant behavioral predictions. We identify valid IVs for further application in the setting of online ad auctions, a setting in which the research field has been broaden, and demonstrated the robustness of the IVs method's prediction accuracy for the overall forecast and the cold-start scenario in our experiments.
(3) Notably, our research advances the concept of utilizing the first stage heterogeneity in the IVs method in the context of prediction. By considering heterogeneity in the strength of IVs concerning impression probability, our method shows more significantly robust prediction performance in whole prediction and the cold-start scenario.

2 IDENTIFICATION OF INSTRUMENTAL VARIABLES IN AD AUCTIONS

2.1 Ad Auctions and Biases

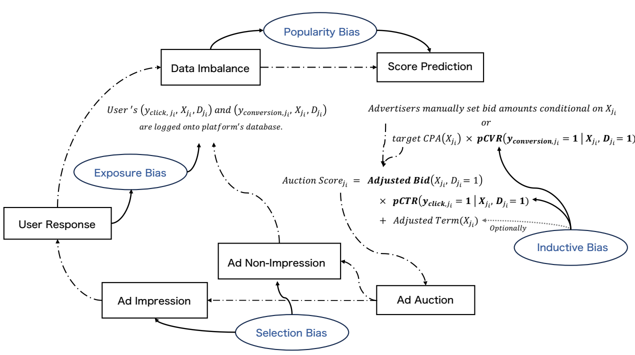


Figure 1: Inductive, Selection, Exposure, and Popularity Bias in Users' Click Behavior and Ad Auction System

Before we explain that the bid amounts is IVs, we describe the setting in ad auctions. This is because it is essential to examine the actual flow of data generation to ascertain the IVs.

The notations used to describe the auction mechanism are as follows: the total number of auctions is N, the number of auctioneers participating in auction $i \in \{1, \cdots, N\}$ is m_i, and the auctioneer's advertisement is $j_i \in \{1, \cdots, m_i\}$. Let Bid_{j_i} be the bid amount that the auctioneer spends on the ad j_i, $pCTR_{j_i}$ be the predictive click-through-rate, and j_i^* be the ad that wins an impression to the user in the auction i. Also, y_{j_i} is the outcome that is 1 if ad j_i is clicked and 0 if not, X_{j_i} is a variables vector used to target ads and users in ad j_i. To simplify complex effects such as position bias, we assume a setting where there is only one ad that wins an impression. Therefore, let D_{j_i} be a binary dummy that is 1 when $j_i = j_i^*$ and 0 otherwise. Also, let y_{j_i} be the outcome that is 1 if the ad j_i^* is clicked and 0 otherwise.

Here, $pCTR_{j_i}$ is as followed:

$$pCTR_{j_i} = p(y_{j_i} = 1 | D_{j_i} = 1, X_{j_i}),$$

where $pCTR_{j_i}$ is the probability of whether ad j_i will be clicked given winning impression, target and other variables.

In ad auctions, there can be various methods for determining auction scores. Here, for instance, the auction score is calculated as follows:

$$Auction\ Score_{j_i} = Bid_{j_i} \times pCTR_{j_i},$$

This determination scheme, which takes into account bid amount and predictive CTR in the auction score, has been studied under the name "weighted GSP" [14, 15]. When the bid amount is a manual bid by the auctioneer, it is generated from the distribution of bid amounts conditional on the target variable of the ad set by the auctioneer. Alternatively, when the bid amount is an automated bid by the platform, the bid amount is generated by, for example, predictive conversion rate (pCVR) and target CPA. In this case, $pCVR_{j_i}$ is a function of X_{j_i}. That is, bid amounts is generated from some distribution conditioned on the target variables of the ad set by the auctioneer or other variables used by the platform. Thus,

$$Bid_{j_i} \sim F(X_{j_i}),$$

where $F(\cdot)$ is the generated distribution of bid amounts.

As summarized by [4], bias in the recommendation system is a looping process. **Figure 1** depicts the looping of several biases, focused in ad auctions setting, which are interdependent. In particular, the auction score will be biased if the platform's prediction of the pCTR is a biased estimator. The same is true for pCVR and adjust term. The assignment of impressions by the auction score with bias is as follows:

$$j_i^* = \underset{j_i \in \{1, \cdots, m_i\}}{\arg\max}\ Auction\ Score_{j_i}^{\text{biased}}.$$

2.2 Causal View of Online Ad Auctions

Treatment D_{j_i}, impressions in ad auctions, can be easily correlated with the error term for the unobserved heterogeneity of users' click behavior. This can be explicitly expressed in the pCTR formulation as follows:

$$p(y_{j_i} = 1) := \theta^*(X_{j_i}, \eta_{j_i}, \epsilon_{j_i} | D_{j_i} = 1),$$

where ϵ_{j_i} represents the error term in the user's click response, and η_{j_i} is unobserved heterogeneity of click behavior that correlates with some or all of X_{j_i} consisting of user and ad features but cannot be observed, known as the omitted variable. $\theta^*(\cdot)$ is a function returns a predictive probability when $y_{j_i} = 1$.

Treatments are determined in the auction system together with predicted values such as pCTR and pCVR, which are conditioned on the user and ad features involved in ad auctions, and the advertiser's bid amount. At this point, pCTR and pCVR are not conditioned on omitted variables η_{j_i}, which generates a bias in the estimates of predictive outcome. Since the bid amount is determined from the predictions with this bias and an auction is formed, there is

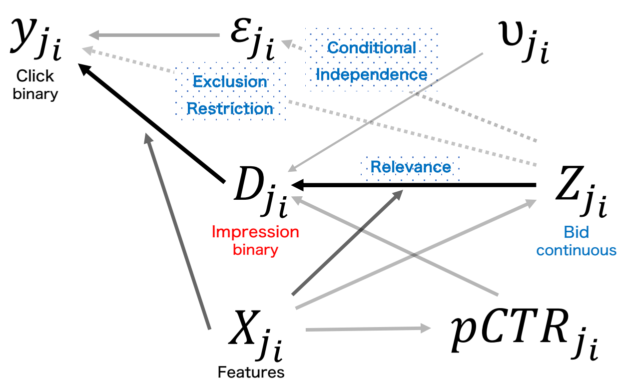


Figure 2: Users' Click Behavior and Bid Amounts as Instrumental Variables in Ad Auctions

a strong suspicion that the impressions D_{j_i} are endogenous variables, which are variables correlated with the error term amplified through the auction with the omitted variable bias. We consider the assumption that no omitted variables exist as a type of inductive bias, a convenient assumption for pCTR model.

Unconfoundedness, i.e., a situation where no omitted variables exist, is a somewhat severe assumption for real-world data. Therefore, IVs methods that do not require the assumption of unconfoundedness can be compelling and valuable.

2.3 Validating Bid Amounts as IVs

There are three conditions that valid IVs satisfy. The first is the relevance of the IVs to a treatment variable. The second is an exclusion restriction, where the IVs does not directly affect the outcome but rather affects the outcome through the treatment variable. The third is the independence of the IVs with respect to the treatment and the outcome. Notating IVs vector in ad j_i as Z_{j_i} and combining these conditions, we can write them as follows:

$$\begin{aligned} &Relevance: && D_{j_i} \not\perp Z_{j_i}, \\ &Exclusion\ Restriction: && \{\epsilon_{j_i}, D_{j_i}\} \perp Z_{j_i}, \\ &Conditional\ Independence: && \epsilon_{j_i} | X_{j_i} \perp Z_{j_i}, \end{aligned}$$

We argue that bid amounts is valid as IVs in ad auctions. The reason bid amounts function as IVs is summarized in **Figure 2** under our proposed IVs formulation.

With regard to the relevance between bid amounts and impressions, the relevance is explicitly acknowledged by the fact that the main item in the auction score is the bid amount. Concerning the exclusion restriction, the bid amount only influences impressions through the auction score. Therefore, the bid amounts does not influence the user's click behavior. Conditional on the variables used by advertisers and platforms to set bid amounts, bid amounts are valid instruments.

2.4 Reasons Other Variables are Not Valid IVs

Here, we introduce why other variables, such as bid times used for targeting, do not meet the conditions of an instrumental variable in ad auctions.

Relevance : Take targeting variables as an example. From the perspective of relevance, advertisers determine bid amounts based on targeting users, which should relate to the probability of assignment. Bid amounts influence the auction score directly, ensuring more vital relevance than targeting variables, while targeting variables have an "indirect" relevance to the auction score.

Conditional Independence : The more crucial condition, however, is that targeting variables do not satisfy the independence from the unobserved factors affecting the user's probability of clicking. For instance, consider bid times as one of the targeting variables. The time when a user requests an advertisement, that is, the user's visitation process, and the probability of clicking the ad can be related. Users visiting at 10 AM may have a higher or lower probability of clicking an ad, and even if conditioned on other targeting variables, the presence of unobserved factors makes it impossible to guarantee the independence of bid times from the click probability. On the other hand, the probability that a user will click is considered independent of the bid amount, conditioned on the targeting variables, since the user cannot know how much was paid for the specific advertising at the time of the click.

Exclusion Restriction : From the perspective of the exclusion restriction, targeting variables affect the probability of a user's click, and do not ensure that their influence on the click probability is exerted solely through the assignment of impressions.

3 CLICK PREDICTION WITH FIRST-STAGE IVS HETEROGENEITY

In the methodology section, we propose several variants of the IVs method to examine the following questions:

- **Q.1** Do prediction methods using simple neural networks with IVs perform in the online ad auction setting? and
- **Q.2** Is IVs heterogeneity strongly present in online ad auction settings and is explicitly addressing it effective in prediction?,
- **Q.3** Heterogeneity in treatment effects is widely known, but by how much improvement relative to accounting for heterogeneity in IVs?

To introduce models that respond to those questions, the methodology section is organized as follows. For **Q.1**, We first introduce the basic structure of the nonparametric IVs method and highlight its heterogeneous relevance to the probability of winning impressions in ad auctions. Next, **Q.2**, we present a method based on an attention network that explicitly considers interactions between IVs and their other features. Finally, **Q.3**, we explicitly incorporate heterogeneity in click probabilities by employing an interaction structure similar to the heterogeneity of instrumental variables. **Figure 3** summarizes our proposed final IVs method.

For simplicity in subscripting the training data, l corresponds to the record number in this section.

3.1 First-stage IVs Heterogeneity in Ad Auctions

In principle, we can estimate a user's click response y_l using IVs in a two-stage approach. Following nonparametric IVs notation by [7], the incorporation of heterogeneity in the first stage can be written

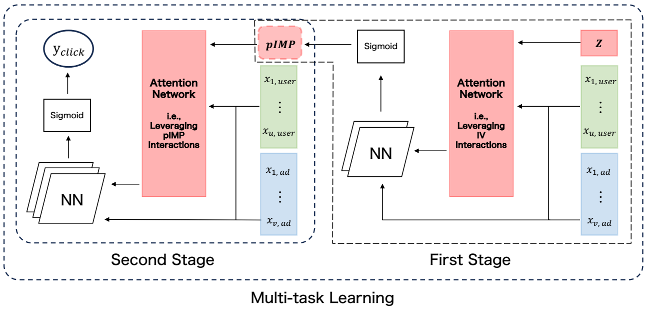


Figure 3: IV-IMP Approach Leveraging First- and Second-stage Heterogeneity with Multi-task Learning Structure

as follows:

$$p(y_l = 1) = \phi^*(X_l, p(Z_l, X_l), \epsilon_l),$$
$$p(Z_l, X_l) = p(D_l = 1 | X_l, Z_l),$$

where $p(Z_l, X_l)$ is an instrument summarized by the interaction of multiple IVs, and we assume that D_l depends only on X_l through $p(Z_l, X_l)$ and call it first stage. ϕ^* is a function that returns a predictive probability of the event $y_l = 1$, which is called second stage. In the ad auctions, $p(Z_l, X_l)$ is the predicted impression probability, henceforth $pIMP$, which is a multi-task learning frame and can be trained in one step together with $pCTR$. Using neural networks, a layer structure can be used that follows the simplified manner of IVs, which we henceforth refer to as the **IV-BS** approach.

Although there can be several approaches incorporating interactions between features and IVs, we use an attention network. This is because it is suitable merely for validating the idea of bid amount heterogeneity.

3.2 Leveraging First-Stage IVs by Interactions

Given a dataset, let the input feature matrix be represented as K after passing through an input layer where all units are fully connected, including units from $pIMP$ and features. Let B denote the batch size and L represent the number of units in the input layer, leading to K having dimensions of $B \times L$. The instrumental variable, represented as matrix Z, has dimensions $B \times 1$. To align with the shape of K, matrix Q^{iv} is formed by performing a tiling operation on Z. Specifically, each row of Z is replicated on the basis of the number of columns in K. Furthermore, the weight matrix for IVs interaction is denoted as W^{iv} and has dimensions $L \times L$. Using these matrices, the attention score α_{iv} is calculated as:

$$\alpha^{\text{iv}} = Softmax(W^{\text{iv}}(Q^{\text{iv}} \odot K) + b^{\text{iv}}).$$

Here, we use the swish function as an activation function in the weight matrix W^{iv} so as to represent the non-linear strength in the heterogeneity of bid amounts. We feed element-wise products as interactions into the fully connected layer with the softmax function as the activation function to generate the attention score α^{iv}. Then, we obtain the representation g by the element-wise product of the input layer K and the generated attention scores α^{iv}.

$$g^{\text{iv}} = \alpha^{\text{iv}} \odot K$$

We combine the representation g obtained by the attention layer and the features input in a fully connected neural network to form the hidden layer.

3.3 Second-stage Heterogeneity

In the second stage, namely in $pCTR$ side, it is evident that heterogeneity exists when conditioning on user and advertisement features regarding the effect of impressions. Similarly to how we took the dot product of bid amounts and feature units in the input layer in the first stage, we symmetrically use the same in the second stage. The input layer consists of fully connected units from $pIMP$ and features. The structure of the entire network including $pIMP$ and $pCTR$ is drawn in Figure 3. The attention score and representation g can be written as follows:

$$\alpha^{\text{imp}} = Softmax(W^{\text{imp}}(Q^{\text{imp}} \odot K) + b^{\text{imp}}),$$
$$g^{\text{imp}} = \alpha^{\text{imp}} \odot K,$$

where Q^{imp} is formed by performing a tiling operation on $pIMP$ to align with the shape of K. Specifically, each row of $pIMP$ is replicated on the basis of the number of columns in K. W^{imp} is a weight matrix of $L \times L$ for $pIMP$ interaction.

3.4 Loss Function for Multi-task Learning

In the multi-task learning framework for pIMP and pCTR, we adjust the loss function for pCTR by applying sample weights through an indicator function, $1_{\{D_l=1\}}$:

$$Loss_{pCTR} = Loss_{pCTR} \times 1_{\{D_l=1\}}$$

This function ensures that the $Loss_{pCTR}$ is only computed for data points with impressions, when $D_l = 1$, filtering out instances without impressions from affecting the pCTR loss calculation. This approach allows us to concentrate on the performance of the model to predict CTR.

4 EXPERIMENTS

The experimental section is divided into two parts: simulation and evaluation in scenarios approximating the cold-start problem with real data sets. The code for replication is available at the following link: https://github.com/ryohei-emori/NPIV-pCTR. Please note that the repository excludes sections related to private data.

The notation is consistent with that used in Section 3.

4.1 Simulated Datasets

The procedures for simulating the auction data are summarized in **Algorithm 1**, aligning with procedure and notation in section 3.1. The experiment is replicated 20 times. The subscripts k and l correspond to the number of records in step 1 and 4, respectively. $\theta(X_k)$ is learned by logistic regression. We use the Beta distribution for generating bid amounts, which satisfies non-negative constraints. Specifically, we use the reparametrized Beta distribution by [6] to model the mean of bid amounts. For simplicity, the number of auctioneers m_i participating in auction i is fixed, but in reality, it may vary depending on the attractiveness of users, represented by X_{j_i}. The link function Logistic$(\cdot)$ is defined as $(1 + \exp(-\cdot))^{-1}$. The feature vectors X_k, X_{j_i}, and X_l are 25×1 vectors respectively. Each $X_{s,k}$ is drawn from a specific

Algorithm 1 Simulating auction data and validating baselines

1. Initializing paramaters:
Set parameters (α, β, γ)
$k := 0$
while $k < 5,000$ **do**
  Generate X_k and η_k
  $D_k \sim \text{Bernoulli}(p_{D_k})$, where $p_{D_k} = \text{Logistic}(X'_k \alpha + \eta_k)$
  if $D_k = 1$ **then**
    $y_k \sim \text{Bernoulli}(p_{y_k})$, where $p_{y_k} = \text{Logistic}(X'_k \beta + \eta_k)$
    $k := k + 1$
  end if
end while
Train pCTR: $p(y_k = 1 | D_k = 1) := \theta(X_k)$
2. Generating historical auction data:
for each auction i in $5,000$ **do**
  $m_i = 20$
  Generate X_{j_i} and η_{ji}
  $Bid_{j_i} \sim \text{Beta}(\mu, 2)$ by [6], where $\mu := \text{Logistic}(X'_{j_i} \gamma)$
  $pCTR_{j_i} = \theta(X_{j_i})$
  $j_i^* := \arg\max_{j_i \in \{1,\cdots,m_i\}} \text{Auction Score}_{j_i}$,
    where $\text{Auction Score}_{j_i} := Bid_{j_i} \times pCTR_{j_i}$
  $y_{j_i} \sim \text{Bernoulli}(p_{j_i})$ & $D_{j_i} = 1$ if $j_i = j_i^*$
    where $p_{j_i} = \text{Logistic}(X'_{j_i} \beta + \eta_{j_i})$
  $y_{j_i} = 0$ & $D_{j_i} = 0$, otherwise
end for
3. Learning $pCTR$ with historical data:
  $\{(y_{j_i}, X_{j_i}, Bid_{j_i}, D_{j_i}), j_i = 1, \cdots, m_i, i = 1, \cdots, 5,000\}$
4. Validating $pCTR$ with independently displayed data:
  $\{(y_l, X_l, D_l = 1), l \in \{1, \cdots, 50,000\}\}$,
    where $y_l \sim \text{Bernoulli}(p_l)$, $p_l = \text{Logistic}(X'_l \beta + \eta_l)$,
    generated X_l and η_l.

distribution: Uniform$[-5, 5]$ for $s \in \{1, \cdots, 10\}$, Bernoulli(0.5) for $s \in \{11, \cdots, 20\}$, and Uniform$[-2, 2]$ for $s \in \{21, \cdots, 25\}$. These vectors are generated similarly. The vectors η_k, η_j, and η_l are generated from a Uniform$[-5, 5]$ distribution. The parameters α, β, and γ are coefficient vectors with 25×1 elements each, independently generated from a normal distribution with a mean of 0.1 and variance of 1.

We assume that rare ads and users have more prominent unobserved confounding factors, and thus evaluate predictive CTR by dividing the degree of magnitude of the omitted variable values. Thus, the test data is separated by the distance of η_l from the mean. Out of a total number of $50,000$ records, we move the outside quantiles of the distribution of η_l by 10% on each side.

4.2 Real Datasets

The actual dataset consists of user responses to advertisements displayed on websites such as Yahoo! JAPAN operated by LY corporation and auction history records including bidding. The datasets are divided into a training dataset, in which ad impressions and clicks are observed through ad auctions, and a test dataset, in which ad impressions are randomly made to visiting users.

4.2.1 ***Training data.*** The training data covers a sample of $50,000$ records randomly drawn from the population for a past seven-day period. The training data were generated from ad auctions system, which produced data not satisfying the condition of conditional independence between the treatment D_{j_i} and unobserved confounders ϵ_{j_i}.

4.2.2 ***Test data.*** In the test data, the prediction baselines using the day after the 7 days of training data is evaluated. The test dataset consists of all independently displayed records conditional on ads' targeting variables.

To evaluate the model's performance in cold-start scenarios, the test data was divided based on previous ad impressions. Specifically, the data was split into 20 subsets at every 5% quantile, with each subset containing data points below the respective quantile. To ensure sufficient sample size, the test data included 2,000,000 records. Predicting clicks with more past impressions is generally easier, even with a simple baseline.

4.3 Evaluation Score

We used log loss, known as a standard evaluation metric for pCTR, and the area under the curve (AUC) scores. AUC is a proper metric for evaluating rankings in assessing the ability to predict the correct position in auction rankings. For the simulation data, we employes the actual scores and relative scores to compare improvements. For our real dataset, we present relative evaluation scores due to confidentiality. The relative scores are defined as follows:

$$\text{Relative LogLoss} = \frac{\text{Naive LogLoss} - \text{Compared LogLoss}}{\text{Naive LogLoss}} \times 100,$$

$$\text{Relative AUC} = (\frac{\text{Compared AUC} - 0.5}{\text{Naive AUC} - 0.5} - 1) \times 100.$$

4.4 Ablation studies

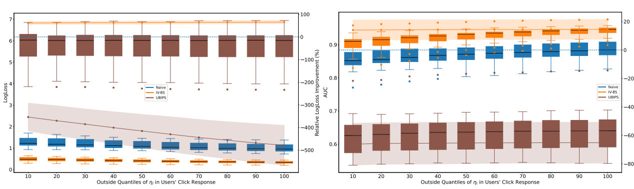
To evaluate our proposed methods with instrumental variables, we took a naive benchmark and comparative baselines.

(1) **Naive**: The **Naive** has three hidden layers between the input layer of features and their passage to the sigmoid function, building a pCTR model. Each of these hidden layers consists of 256 units. The first layer uses the swish activation function, while the second and third layers use the ReLU activation function.
(2) **IV-BS**: The baseline is described in section 3.1. Its pCTR model has the same network structure as **Naive**, including $pIMP$ in the input layer.
(3) **IV-FS**: The baseline is described in section 3.2. In $pCTR$ side, it has the same network structure as **IV-BS**.
(4) **IV-SSFS**: The baseline in $pCTR$ side is described in section 3.3, while its network has the same structure as **IV-FS** in $pIMP$ side.
(5) **UBIPS** : It consists of $pIMP$ times $pCTR$ for unbiased inverse propensity weighting estimator [12]. Its network structure is consistent with **IV-BS** for $pIMP$ and $pCTR$ excluding $pIMP$ in the input of $pCTR$. It also uses a multitasking framework.

The **IV-FS** and **IV-SSFS** are not tested in our simulated dataset for two reasons: one is the **IV-BS** is sufficient to test whether bid amounts are efficient and valid IVs in ad auctions. Another is those approaches are not suitable to the simplicity, such as the linear interactions, in the heterogeneity of IVs and the user's click probability in our simulated dataset.

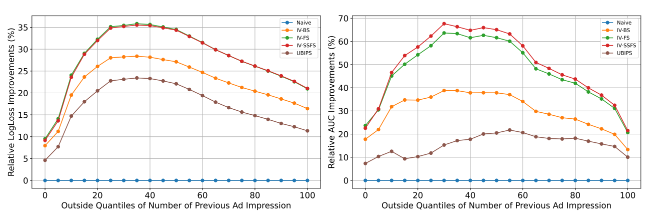
In this experiments, the loss function is unified across comparative beselines. *pCTR* and *pIMP* models both use binary cross entropy as their loss function. We trained the comparison models until convergence, where no further improvement in the loss function in *pCTR* was observed. For all comparative approaches, the optimization method was Adamax, and the learning rate was fixed at 0.001.

4.5 Comparing Each Baselines



(a) LogLoss & Relative LogLoss **(b) AUC & Relative AUC**

Figure 4: Simulation: Performance scores at each outside quantile of η_l. Box plots show actual scores. Line plots show relative scores, with the bold line as the mean and shaded area showing replication variation.



(a) Relative LogLoss **(b) Relative AUC**

Figure 5: Real data: Performance scores at each quantile of previous ad impressions.

4.5.1 ***In Simulated datasets***.

Figure 4 shows that **IV-BS** improves AUC and LogLoss performance even with omitted variables. **IV-BS** remains stable and robust, especially on the left side where the test data's η_l value is high. Notably, omitted variable bias cannot be ignored even in the Weighted GSP impression assignment algorithm, and in this regard, **IV-BS** demonstrates superior performance.

4.5.2 ***In Real dataset***.

An evaluation of our proposed methods on the real dataset is shown in **Figure 5**. It is expected that **Naive** performs relatively well since the training data includes many ads with numerous impressions. However, our proposed methods, **IV-BS**, **IV-FS**, and **IV-SSFS**, show significant improvement in relative AUC, particularly for ads with few previous impressions. The improvement of **UBIPS** over **Naive**, unlike in the simulation experiment, is likely attributable to the confounder being associated with the variable observed in the actual data.

Improvement for ads with few impressions matches that for ads with many, likely due to the infrequent inclusion of rare ads in training data, causing popularity bias. Notably, the increasing improvement of IVs methods for the $0 - 20$ quantile of previous impressions demonstrates their robustness in predicting rare ads.

5 CONCLUSION

This paper argues that bid amount is a valid instrumental variable under the assumption of conditional independence, and tested its validity by applying it to predictive CTR. Our experiment on a real dataset showed that explicitly accounting for heterogeneity in the strength of IVs allows for efficient and robust predictions. For greater extensibility, incorporating complex interactions between IVs and other features with more developed approachs such asgraph neural networks is recommended. Additionally, addressing other looping bias and validating prediction methods in repeated auctions would be valuable.

REFERENCES

[1] Alberto Abadie, Jiaying Gu, and Shu Shen. 2023. Instrumental variable estimation with first-stage heterogeneity. *Journal of econometrics* (2023), 105425–.
[2] Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 6 (2012), 2369–2429.
[3] Peter Bühlmann. 2020. Invariance, Causality and Robustness. *Statistical science* 35, 3 (2020), 404–426.
[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 Transactions on Information Systems* 41, 3 (2023), 1–39.
[5] Amir Feder, Guy Horowitz, Yoav Wald, Roi Reichart, and Nir Rosenfeld. 2022. In the Eye of the Beholder: Robust Prediction with Causal User Modeling. *Advances in Neural Information Processing Systems* 35 (2022), 14419–14433.
[6] Silvia Ferrari and Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of applied statistics* 31, 7 (2004), 799–815.
[7] Markus Frolich. 2007. Nonparametric IV estimation of local average treatment effects with covariates. *Journal of econometrics* 139, 1 (2007), 35–75.
[8] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. 2017. Deep IV: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*. PMLR, 1414–1423.
[9] Yue He, Zimu Wang, Peng Cui, Hao Zou, Yafeng Zhang, Qiang Cui, and Yong Jiang. 2022. Causpref: Causal preference learning for out-of-distribution recommendation. In *Proceedings of the ACM Web Conference 2022*. 410–421.
[10] Guido W. Imbens. 2014. Instrumental Variables: An Econometrician's Perspective. *Statist. Sci.* 29, 3 (2014), 323–358. http://www.jstor.org/stable/43288511
[11] Veronica Marotta, Yue Wu, Kaifu Zhang, and Alessandro Acquisti. 2022. The Welfare Impact of Targeted Advertising Technologies. *Information Systems Research* 33, 1 (March 2022), 131–151. https://doi.org/10.1287/isre.2021.1024
[12] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (Houston, TX, USA) *(WSDM '20)*. Association for Computing Machinery, New York, NY, USA, 501–509. https://doi.org/10.1145/3336191.3371783
[13] Zihua Si, Xueran Han, Xiao Zhang, Jun Xu, Yue Yin, Yang Song, and Ji-Rong Wen. 2022. A Model-Agnostic Causal Learning Framework for Recommendation Using Search Data. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 224–233. https://doi.org/10.1145/3485447.3511951
[14] Yang Sun, Yunhong Zhou, and Xiaotie Deng. 2014. Optimal reserve prices in weighted GSP auctions. *Electronic Commerce Research and Applications* 13, 3 (2014), 178–187. https://doi.org/10.1016/j.elerap.2014.02.003
[15] David RM Thompson and Kevin Leyton-Brown. 2013. Revenue optimization in the generalized second-price auction. In *Proceedings of the fourteenth ACM conference on Electronic commerce*. 837–852.

Received 15 June 2024