Predicting Winning Price in Real Time Bidding with Censored Data

Wush Chi-Hsuan Wu
Dept. of Electrical Engineering
National Taiwan Univ.
Taipei, Taiwan
wush978@gmail.com

Mi-Yen Yeh
Inst. of Information Science
Academia Sinica
Taipei, Taiwan
miyen@iis.sinica.edu.tw

Ming-Syan Chen
Dept. of Electrical Engineering
National Taiwan Univ.
Taipei, Taiwan
mschen@cc.ee.ntu.edu.tw

ABSTRACT

In the aspect of a Demand-Side Platform (DSP), which is the agent of advertisers, we study how to predict the winning price such that the DSP can win the bid by placing a proper bidding value in the real-time bidding (RTB) auction. We propose to leverage the machine learning and statistical methods to train the winning price model from the bidding history. A major challenge is that a DSP usually suffers from the censoring of the winning price, especially for those lost bids in the past. To solve it, we utilize the censored regression model, which is widely used in the survival analysis and econometrics, to fit the censored bidding data. Note , however, the assumption of censored regression does not hold on the real RTB data. As a result, we further propose a mixture model, which combines linear regression on bids with observable winning prices and censored regression on bids with the censored winning prices, weighted by the winning rate of the DSP. Experiment results show that the proposed mixture model in general prominently outperforms linear regression in terms of the prediction accuracy.

Categories and Subject Descriptors: H.3.5 [Information Storage And Retrieval]: Online Information Services; I.2.6 [Artificial Intelligence]: Learning

Keywords: Demand-Side Platform, Real-Time Bidding, Display Advertising, Learning with Partial Labels

1. INTRODUCTION

In recent years, programmatic advertising has been taking over the online ad industry. To enable automatic selling and purchasing ad impressions between advertisers and publishers through real-time auctions, Real-Time Bidding (RTB) is quickly becoming the leading method according to the white paper by Google Inc. [8].

In contrast to the traditional online ad market, where a certain amount of impressions is sold at a fixed rate, RTB allows advertisers to bid each impression individually in real time at a cost based on impression-level features. In the RTB paradigm, publishers manage and sell their inventories of ad impressions via the Supply-Side Platform (SSP) while advertisers can bid and manage their ads across multiple inventory sources via the Demand-Side Platform (DSP). While an impression is loaded in a web browser, information about the user reading this webpage will be passed to the advertisers. DSP, on behalf of the advertisers, will help decide whether to bid the impression and set a bid to the ad exchange. Usually RTB takes the second price auction, where the advertiser with the highest bidder wins and pays the second highest bid. Interested readers are referred to the introduction and tutorials in [17] and [19] for more details about RTB.

Since a DSP is the agent of advertisers, it has to know how to compute the bid properly. The common paying models for the advertiser include cost-per-mille (CPM) for each impression, cost-per-click (CPC) for each click, and cost-per-action (CPA) for each conversion. Therefore, DSPs usually estimate the click-through-rate (CTR) and/or the conversion rate of the impression so that they can set a proper bid for advertisers to maximize their profits or maximize the revenue with a limited budget. From the standpoint of DSP in the RTB exchange, we propose in this paper to predict the *winning price* of the bid directly. The main reason is that the winning price is usually the same as the cost of winning the bid. There is usually a budget constraint for DSPs to bid the impressions, and the cost is hence an important factor of the bidding decision making as stated by Zhang et al. [21]. Furthermore, the winning price is an indicator of the importance of an ad impression or the importance of a user in the market. The information of the market value might help DSPs predict the CTR and conversion rate. Our experiment shows that adding the winning price to the feature sets in the prediction of CTR can increase the AUC value from 0.708 to 0.733[1].

Apparently, DSPs meet a grand challenge to predict the winning price. Ghosh et al. [7] first described that the partial observable exchange makes them hard to explore the pattern of winning prices. According to the mechanism of the modern RTB display advertising, the winning price is only observable for the DSP who wins the bid. For the lost bids, DSPs can only observe the lower bound, which is their own bidding price. Moreover, the modern RTB process introducing the soft floor price makes the winning price of some winning bids also unobservable, of which more details

KDD'15, August 10-13, 2015, Sydney, NSW, Australia.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3664-2/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2783258.2783276 .

[1] The data of the experiment is the day 2013-06-07 of the dataset iPinYou session 2. The details of the dataset and the features are shown in Sec. 4.1.

are given in Sec. 2. In such cases, only the upper bound or the lower bound of the winning price, which is the paying price of the bid, is observed. This kind of partially observed winning price is called *censoring*, and is handled in many fields such as survival analysis [12] and econometrics [15].

To solve the problem of predicting the winning price for a DSP, say D_i, we propose to learn a mixture model of winning price from historical bids, where the winning prices are either observable or unseen for D_i. Our rationale is as follows. Intuitively, we can train a linear regression model based only on the bids, of which the winning price is observable to D_i (or for simplicity, we call these bids *winning bids* of D_i hereafter). Apparently, the model could be overfitting to these limited winning bids. As a result, we further propose to use the censored regression model [9] to consider also the censored data, which are the bids with unobservable winning prices (or we call them *losing bids* of D_i hereafter for simplicity). However, the real bidding data in fact violate an assumption of the censored regression model: the pattern of winning prices on observed data and that on censored data should be consistent. We will show this violation in Sec. 4.2. To remedy the violation, we propose to mix the result of linear regression model and that of the censored regression model, weighted by the *winning rate* of D_i on a bid. This is because the linear regression model works better on the winning bids in general and the censored regression model takes the censored information into account. Therefore, if the winning rate, which is the chance that D_i wins the bid, is higher, then the linear regression model accounts for more percentage in the mixture model, and vice versa.

Our contributions can be summarized as follows. We are the first, to our knowledge, to predict the winning price on the censored data by applying the censored regression model. Furthermore, we propose an approach to predicting the winning rate for a DSP on a bid and leverage it to mix the results of the linear regression model on winning bids and the censored regression model. We conduct a series of experiments on real RTB data sets from two existing DSPs. Experiment results show that our proposed mixture model outperforms the baseline methods in all cases in terms of having the smaller predicting errors. This shows the effectiveness of our proposed model in practice.

The rest of the paper is organized as follows. Section 2 reviews the auction process of RTB and describes the challenge of modeling winning prices. Section 3 presents our methodology to predict the winning price. The experiment results are provided in Section 4. We give a brief literature survey related to our topic in Section 5, followed by the conclusion in Section 6.

2. PRELIMINARIES

In this section, we describe the mechanism of modern RTB and define the winning price we discussed in this paper. Also, we show the availability of the true winning price in different bidding cases and thus the challenges of modeling it.

2.1 The Censoring of Winning Price

We only focus on the interaction between a DSP and the ad exchange in the RTB process. Suppose we are the DSP and our bidding price is b. The winning price we discuss here refers to the price value that we, as a DSP, have to offer at least to win the bid. In the current RTB process, the publishers can set a soft floor price and the hard floor price for each ad impression. If all the bids from all DSPs are below the hard floor price, the impression will fail to be sold. In this case, the winning price is the hard floor price set by the publisher of the ad impression. On the other hand, if there is at least one bidding price greater than the hard floor price, the winning price is then the highest bidding price from all other DSPs.

Usually RTB runs a second-price auction. Suppose the highest bidding value of all other DSPs is $b2$ and we win the bid, then we only need to pay $b2$. Thus the winning price is $b2$. In practice, however, the winning price is affected by the soft floor price set by SSPs at the same time. If b is smaller than the soft floor price of the ad impression, RTB will ask the winner to pay b. If b is larger than the soft floor price of the ad impression and $b2$ is smaller than the soft floor price, RTB will ask the bidder to pay the soft floor price instead of $b2$. In this case, the winning price is *left-censored*, which means we only know an upper bound of the winning price. If we lose a bid, the winning price is not available and generally *right-censored*, i.e., our bidding price is a lower bound of the winning price. The winning price of each bid can be known by us only if the RTB process makes the bidding price public or every DSP makes its own bidding public. In summary, the auction process and the resulting data censoring are shown in Fig. 1.

2.2 Problem Statement

The goal of this paper is to learn the winning price, in the aspect of a DSP, with historical RTB bids and to deal with the problem of various winning price censoring at the same time. We aim to propose a machine-learning based model that can help a DSP to predict the winning price and thus decide the bidding price.

3. METHODOLOGY

In this section, we describe the proposed method to predict winning price. We state the true, but unavailable in practice, model of the winning price and how we approximate it with a statistical model in Sec. 3.1. The modeling of the censored data is introduced in Sec. 3.2. We show how and why to use the estimated winning rate to mitigate the impact of the unrealistic assumption in Sec. 3.3.

For ease of presentation, we assume that all winning bids are observed and all losing bids are right-censored in the following content. We will also show how to extend our model to left-censored data.

3.1 Modeling Winning Price

Suppose there are J DSPs, $D_1, ..., D_J$, in the RTB market and without loss of generality the following process is described from the standpoint of DSP D_1. For the i^{th} bid, we denote the true winning price by v_i, the observed one by w_i, and the bid placed by D_1 by b_i. If D_1 wins the bid, then $w_i = v_i$. Otherwise, w_i=0.

In practical RTB, SSPs will deliver several features of the ad impression for sale to all DSPs with the bid request, which are denoted by x_i^1 here. Of course, DSPs themselves can collect features other than x_i^1 of the bid on their own and do not make them available to others. We denote the features collected and observed only by D_1 as x_i^2 and features collected by other DSPs and unknown to D_1 as x_i^3. For example, according to the report of iPinYou [21], the feature

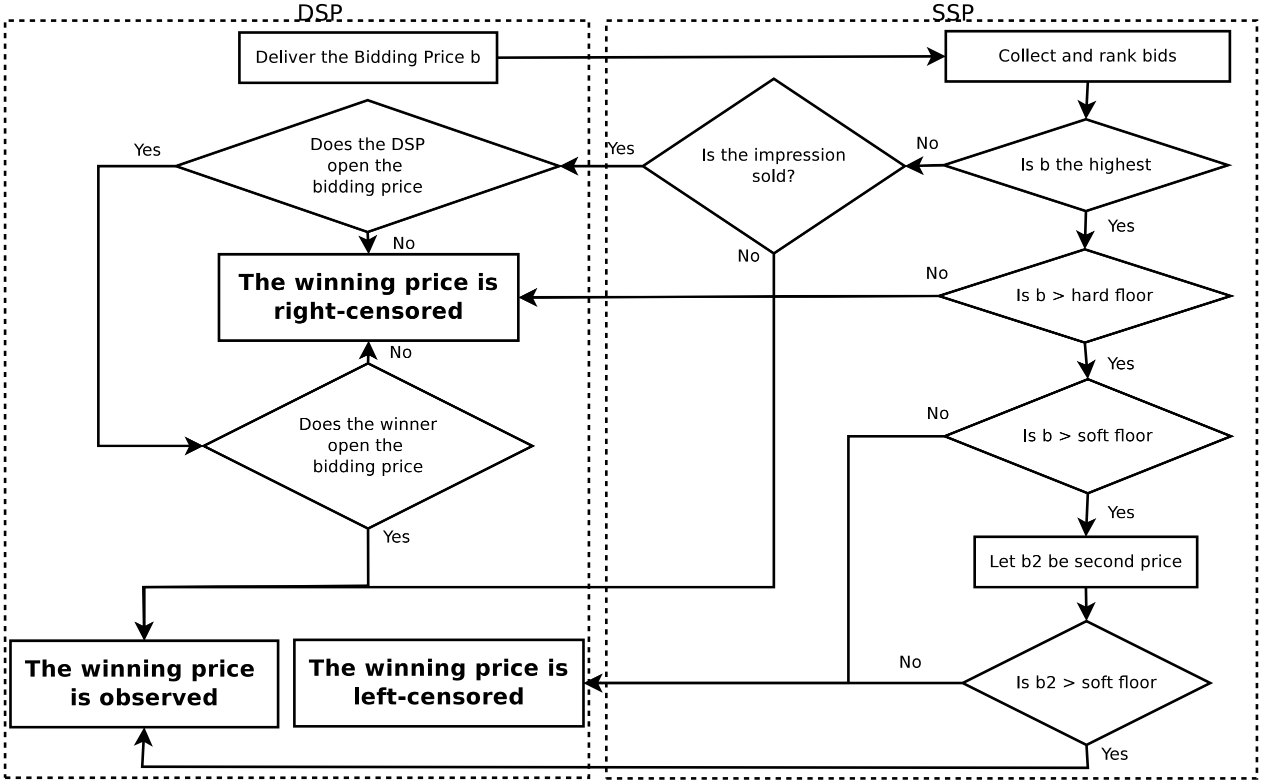
DSP
SSP
Deliver the Bidding Price b
Collect and rank bids
Is b the highest
No
Yes
Is the impression sold?
Yes
No
Does the DSP open the bidding price
Yes
No
The winning price is right-censored
No
Is b > hard floor
No
Yes
Does the winner open the bidding price
Yes
Is b > soft floor
No
Yes
Let b2 be second price
The winning price is observed
The winning price is left-censored
No
Is b2 > soft floor
Yes

Figure 1: Flow chart of the RTB auction and the resulting winning price censoring.

User Tags is only observed by iPinYou. Hence this feature belongs to x_i^2 for iPinYou and is deemed as x_i^3 by other DSPs in the market.

In the bidding process, each DSP offers a price according to its bidding strategy function $f_j(\cdot)$, which could be either deterministic or stochastic, for D_j and the available features. Therefore, the ad exchange receives bids $f_1(x_i^1, x_i^2)$, $f_2(x_i^1, x_i^3)$, ..., and $f_J(x_i^1, x_i^3)$. Suppose the hard floor price of the publisher, which is the webpage owner of the auctioned impression, is denoted as $f_p(x_i^1)$ because the publisher is one of the public feature. Then we can model the true winning price of the i^{th} bid for D_1 as

$$v_i = max\{f_p(x_i^1), f_2(x_i^1, x_i^3), ..., f_J(x_i^1, x_i^3)\}. \quad (1)$$

D_1 will win the impression if it offers the price higher than v_i.

In reality, x_i^3, and thus v_i in Eq. (1), is unknown to D_1. We propose to use a linear regression model to approximate the true winning price as follows.

$$v_i \approx \beta^T x_i + \varepsilon_i, \quad (2)$$

where $x_i = \{x_i^1, x_i^2\}$ is the set of all features observed by D_1, $\beta^T x_i$ captures the expected value of the winning price, and ε_i is assumed to be an independent and identically distributed (*i.i.d.*) and normally distributed random variable with zero mean and σ^2 variance. With Eq. (2), we can predict the winning price v_i based on the features x_i and the vector β. Therefore, we call β an *estimator* in the following content.

Although the linear regression model with i.i.d. normal distributed assumption of ε_i may not be the best fit for predictive analysis, as we can see, this model is generally elegant and effective to capture the censored information, which we will discuss in the next subsection, compared to other sophisticated model.

We use the negative log-likelihood function to measure how good Eq. (2) can approximate the observed winning price:

$$\sum_{i \in W} -log(\phi(\frac{w_i - \beta_{lm}^T x_i}{\sigma})), \quad (3)$$

where W represents all the historical bids that D_1 wins and thus it observed the true winning prices. ϕ is the probability density function (pdf) of the standard normal distribution. The vector β_{lm} can be learned from historical bidding data by finding the value that minimizes Eq. (3).

3.2 Modeling Censored Data

In Eq. (3), the estimator β_{lm} learns the winning bids of D_1. However, we may lose too much winning price information from those bids that D_1 lost. Second, the winning price from the observed data usually is lower than the winning price of the censored data as shown in Sec. 4.1, and the estimator β_{lm} hence usually underestimates the winning price on the censored data. Therefore, in this section we describe how to apply the censored regression model to explore the information from those losing bids.

The linear regression described in Eq. (2) predicts not only the winning price but also the winning rate. That is, the probability that D_1 will win the i^{th} bid, can be derived from Eq. (2) as follow.

$$P(v_i < b_i) = P(\varepsilon_i < b_i - \beta^T x_i) = \Phi(\frac{b_i - \beta^T x_i}{\sigma}), \quad (4)$$

where Φ is the cumulative density function of the standard normal distribution. Note that one can modify Eq. (4) to the left censoring case by changing "<" to ">" properly. Therefore, the censored regression model can be modified to include the unobserved winning price due to the soft floor price discussed in Sec. 2.

The idea of censored regression is to combine Eq. (3) and Eq. (4). Eq. (3) measures how well the model in Eq. (2) approximates the observed data and Eq. (4) measures how well the model in Eq. (2) approximates the censored data. Following the principle of maximum likelihood and the Tobit model [16], the estimator β_{clm} is learned via minimizing the negative log-likelihood function as follows.

$$\sum_{i \in W} -log(\phi(\frac{w_i - \beta_{clm}^T x_i}{\sigma})) + \sum_{i \in L} -log(\Phi(\frac{b_i - \beta_{clm}^T x_i}{\sigma})), \quad (5)$$

where W represents the set of all the winning bids and L represents the set of all the losing bids of D_1, respectively.

Apparently, Eq. (5) is a mixture of winning price prediction in Eq. (2) and the winning rate prediction in Eq. (4). Because the censored regression learns from more data compared to the linear regression, which learns from the winning bids only, the estimator of the censored regression β_{clm} is supposed to have a smaller variance compared to the estimator β_{lm}.

However, there is an important assumption for deriving the likelihood function in Eq. (5): the pattern of the winning prices on the observed data should be the same as that of the winning prices on censored data. In Eq. (2), the estimator β is a realization of the pattern of winning price because β describes the changes of winning price after changing the features x_i. More specifically, the assumption for deriving the likelihood function in Eq. (5) is that β estimator on the observed data and that on the censored data are the same. If these patterns are different, the estimator β_{clm} becomes a mixture of two different βs. Therefore, the estimator β_{clm} is biased when it is used to predict the winning price on the observed data. Similarly, the estimator β_{lm} is biased when it is used to predict the winning price on the censored data.

In practice, the estimator β_{lm} usually underestimates the winning price, and β_{clm} usually provides higher estimation than β_{lm} does because it learns the censored data.

We will further show in Sec. 4.2 that the patterns of the winning price on the observed data are different from those on the censored data in practice. Therefore, it is hard to conclude that β_{clm} will be better than β_{lm} or not because the assumption is invalid. To solve this problem, we propose a mixture model in the next subsection.

3.3 Mixture Model

We introduce the censored regression to learn the pattern of winning price from both observed and censored data. However, the patterns on the observed and censored data are inconsistent, which may violate the assumption of the censored regression model. Therefore, in this section, we propose a mixture model to address the effect of mixing different patterns.

The winning rate prediction provides the likelihood of the bid whether it is going to be observed or not. Therefore, we use the predicted winning rate as the weight to mix the estimator β_{lm} and the estimator β_{clm}. That is to say, we predict the winning price according to the following equation.

$$v_i = [P(v_i < b_i)\beta_{lm} + (1 - P(v_i < b_i))\beta_{clm}]^T x_i = \beta_{mix}^T x_i. \quad (6)$$

Unlike that of the estimator β_{clm}, the performance of the estimator β_{mix} is better than that of the estimator β_{lm} if the winning rate prediction is accurate enough. Our reasons are as follows. If the winning rate prediction is absolutely correct, i.e. the winning rate is 1 on the observed data and 0 on the censored data, the estimator β_{lm} outperforms on the observed data and the estimator β_{clm} outperforms on the censored data, the performance of the estimator β_{mix} will be the same as that of β_{lm} on the observed data, and will be the same as that of β_{clm} on the censored data. That is to say, the winning rate helps β_{mix} to pick the better estimator.

Note that, in practice, the winning rate prediction will not be absolutely correct and the estimator β_{clm} might not outperform β_{lm} on the censored data. Also, the winning price on the censored data is usually higher than the winning price on the observed data because these bids are lost due to the bidding price is not high enough. Therefore, β_{lm} underestimates the winning price at censored data while β_{clm} provides a higher estimation. If β_{clm} underperforms, it suggests that β_{clm} overestimates too much. Interestingly, β_{mix} aggregates the merits of these two estimators and provides a better estimation. β_{mix} gives higher estimation than β_{lm} does because β_{clm} is higher and β_{mix} gives more conservatively lifting of the estimation compared to β_{clm}.

We will compare these estimators β_{lm}, β_{clm}, and β_{mix} on all data, observed data and censored data in Sec. 4.3 and Sec. 4.4.

To realize the mixture model, however, we understand that Eq. (4) is not directly applicable to predict the winning rate $P(v_i < b_i)$. First, the accuracy of Eq. (4) depends on the distribution of the random variable v_i. Therefore, the accuracy is affected by the accuracy of the winning price prediction. Second, predicting winning rate with Eq. (4) with the estimator β_{lm} or β_{clm} might introduce hidden dependency because we already mix them to derive the estimator β_{mix}. In view of these, we do not use the model in Eq. (4) directly.

Instead, we use the logistic regression to predict the winning rate used in Eq. (6)

$$P(v_i < b_i) = \frac{1}{1 + e^{-\beta_{lr}^T x_i}}. \quad (7)$$

The logistic regression describes the winning rate based on the features directly, so it is not derived from Eq. (2). Therefore, it is not affected by the accuracy of the winning price and it does not introduce any hidden dependency from any estimator derived from Eq. (2).

4. EXPERIMENTS

In this section, we shall address the following questions. (1) Are the patterns of the winning price on the observed data and those on the censored data different? (2) Does the censored regression model have better performance on predicting winning price compared to the linear regression

day	# of bids	# of win. bids	WR	EWR	WR AUC	Avg. WP	Avg. WP on W	Avg. WP on L
2013-06-06	1,821,479	1,514,416	0.83	0.83	0.89	74.86498	52.46772	185.3269
2013-06-07	1,806,062	1,524,314	0.84	0.85	0.90	72.31279	51.12051	186.9674
2013-06-08	1,634,967	1,352,038	0.83	0.83	0.87	81.14319	58.48506	189.4200
2013-06-09	1,651,630	1,366,097	0.83	0.83	0.88	81.31667	58.95707	188.2934
2013-06-10	1,920,576	1,603,798	0.84	0.83	0.91	79.83572	58.91341	185.7621
2013-06-11	1,745,905	1,461,085	0.84	0.86	0.85	79.62260	58.91626	185.8431
2013-06-12	1,657,578	1,378,728	0.83	0.85	0.84	79.99693	58.80196	184.7920

Table 1: The statistics for the dataset iPinYou Season 2.

day	# of bids	# of win. bids	WR	EWR	WR AUC	Avg. WP	Avg. WP on W	Avg. WP on L
2013-10-19	228,133	170,739	0.75	0.75	0.85	86.73686	50.89386	193.3647
2013-10-20	214,295	159,646	0.74	0.74	0.87	87.89860	51.95333	192.9054
2013-10-21	848,760	621,626	0.73	0.73	0.89	90.08863	51.40931	195.9472
2013-10-22	681,700	503,108	0.74	0.74	0.87	91.51073	52.88855	200.3125
2013-10-23	226,791	170,850	0.75	0.75	0.83	89.27304	53.44846	198.6853
2013-10-24	245,897	197,279	0.80	0.80	0.83	74.18814	45.58807	190.2397
2013-10-25	318,240	245,671	0.77	0.77	0.85	82.27562	49.19676	194.2589
2013-10-26	268,380	198,709	0.74	0.74	0.89	89.61273	49.18726	204.9104
2013-10-27	110,467	84,730	0.77	0.77	0.90	86.00950	48.43570	209.7080

Table 2: The statistics for the dataset iPinYou Season 3.

model trained from only the historical winning bids? (3) Does the proposed mixture model successfully mitigate the violation of the censored regression model and leverage the results of both linear and censored models?

4.1 Settings

For the following experiments, we use two real datasets, the iPinYou RTB data set[2] and the Bridgewell dataset[3]. Since both the two DSPs can only provide the winning prices of their own won bids, we simulate the bidding results of both winning and losing bids as follows. For each historical bid, we divide the bidding price originally offered by each DSP by a factor as the new bidding price. If not specified, the factor is set to 2. The new bidding price is compared to the original winning price to produce the simulated winning (new bidding price > original winning price) and losing bids. As such, we can always have the ground truth of the winning price of each bid and the pattern of the winning prices will not vary. Note that *winning bids* and *observed data* are used interchangeably, and so does *losing bids* and *censored data*. Although the winning price is obtained from our simulation results instead of randomly drawn from the winning prices in the real market, we believe the experimental study of the winning price pattern conditioned on the simulated winning and losing bids reveals reasonable trends accordingly. We release the codes for experiments on the iPinYou dataset on `https://github.com/wush978/KDD2015wpp`.

There are three seasons of RTB records on the iPinYou dataset. We use the data of season 2 and season 3 because the private feature *User Tags*, which corresponds to x_i^2 of iPinYou, is not provided in season 1. All the features we used from the logs of iPinYou are shown in Table 4. These features are categorical and converted to binary features via hashing trick proposed by Weinberger et al. [18]. This kind of feature extraction is widely used in the literature such as He et al. [10], Zhang et al. [21] and Chapelle [3]. The hash size in the experiments is set to 2^{20}.

The proposed mixture model of the winning price requires the winning rate as input which is predicted by the simulated bidding result. According to Zhang et al. [20], the expected KPI, which is defined according to different business models and usually is the expected CTR or the expected conversion rate, is an important feature for predicting winning rate. As a result, we put the expected CTR into the feature set x. Because the iPinYou dataset only provides the log of click events instead of the expected CTR directly, we estimate the CTR on our own by the logistic regression as stated by Zhang et al. [21]. The accuracy of our expected CTR is shown in Table 5. Compared to the benchmark reported by Zhang et al. [21], the total AUC of our model is only 0.0009 lower than theirs. Hence, the expected CTR is good enough for being a part of x.

To simulate the online environment, we predict CTR and the winning rate in an online fashion. After receiving a bid request, the model predicts the corresponding CTR and the winning rate immediately. Then we check whether the impression that occurs ten minutes ago is clicked or not and update the CTR model. Similarly, we check the result of bidding that occurs one minute ago and update the winning rate model. The updated model takes effect immediately. Both the CTR and winning rate models are optimized by FTPRL proposed by McMahan et al. [13].

There are 9-day logs in the Bridgewell dataset. These logs record the bidding history for a specific advertiser. Unlike the iPinYou dataset, there are real expected KPI in the dataset, so we estimate the winning rate according to the features and expected KPI directly. The features we used are similar to those in the iPinYou dataset. Because there are some extremely high winning prices in the Bridgewell dataset, we convert the winning prices to a logarithmic scale before using them.

The data are split by the day according to the occurrence time of the bid request. Table 1, Table 2 and Table 3 show

[2] `http://data.computational-advertising.org/`

[3] The dataset is sampled from the real RTB logs from *Bridgewell Inc.*(`http://www.bridgewell.com.tw/index.html`) , which is one of DSPs in Taiwan.

day	# of bids	# of win. bids	WR	EWR	WR AUC	Avg. WP	Avg. WP on W	Avg. WP on L
2014-12-01	2,799,327	1,911,988	0.68	0.68	0.77	256.8171	254.3374	262.1601
2014-12-02	3,042,964	1,952,962	0.64	0.64	0.77	253.0018	253.0107	252.9859
2014-12-03	2,667,145	1,712,442	0.64	0.64	0.76	254.9517	251.9175	260.3940
2014-12-04	2,418,780	1,487,997	0.62	0.61	0.74	259.4486	253.1226	269.5615
2014-12-05	1,734,819	972,121	0.56	0.56	0.72	265.4780	258.3856	274.5178
2014-12-06	1,209,382	646,522	0.53	0.53	0.71	236.3764	227.0063	247.1393
2014-12-07	1,314,122	707,196	0.54	0.54	0.71	230.5689	220.1978	242.6534
2014-12-08	1,721,061	949,686	0.55	0.55	0.73	251.8090	244.4737	260.8399
2014-12-09	1,675,202	933,429	0.56	0.56	0.72	278.5520	268.3400	291.4026

Table 3: The statistics for the dataset Bridgewell.

the basic statistics of the dataset, where each row shows the statistics of a specific day. In these tables, *# of bids* and *# of win. bids* are the count of total bids and simulated winning bids respectively. *WR* is the resulting simulated winning rate which equals # of win. bids / # of bids. *EWR* is the expected winning rate according to our winning rate model in Eq. (7). *WR AUC* is the Area Under Curve (AUC) of our winning rate model based on our simulated bidding results. *Avg. WP*, *Avg. WP on W*, and *Avg. WP on L* are the average winning price on all bids, the winning bids, and the censored bids, respectively.

It can be seen that the characteristics of the three datasets are very different. First, the winning rate varies significantly among three tables. Also, the proportions of the winning bids to the total bids are also different. The number of winning bids is almost equal to the number of losing bids on the Bridgewell dataset while the number of winning bids is around 4.88 times as many as the losing ones on the iPinYou Season 2 dataset.

Example	Feature Name
183.18.197.*	IP
216	Region
236	City
2	AdExchange
3d68edb4b8f5bba8bea6782e33c8e228	Domain
e63cfda49ec2a36a0cafcd646906227b	URL
3844656199	AdSlotId
250	AdSlotWidth
250	AdSlotHeight
OtherView	AdSlotVisibility
Na	AdSlotFormat
7321	CreativeID
6	weekday
02	hour
2259	adid
10684,10102,10006	usertag

Table 4: Example features of the iPinYou dataset [21].

4.2 The Difference of Winning Price Patterns

To answer question (1), we observe the patterns of the winning price in two ways. First, we read the average winning price on both winning and losing bids as shown in Table 1, Table 2, and Table 3. We can see that the average winning prices on losing bids, which are the values in column *Avg. WP on L*, are usually higher than the average winning prices in column *Avg. WP on W* in Table 3. This shows one difference of the patterns of the winning price on the winning and the losing bids.

adid	log.loss	auc
1458	0.0068	0.7350
2259	0.0075	0.5665
2261	0.0024	0.5652
2821	0.0051	0.5411
2997	0.0249	0.5908
3358	0.0093	0.7221
3386	0.0081	0.7086
3427	0.0074	0.6987
3476	0.0053	0.6837
total-2	0.0073	0.7153
total-3	0.0074	0.7433
total	0.0073	0.7211

Table 5: The performance of our CTR model of the iPinYou dataset.

Furthermore, we use the linear regression model and observe the estimator learning the winning bids only, denoted by β^w_{lm}, and the estimator learning the losing bids only, denoted by β^l_{lm}, by minimizing Eq. (3). Because there are too many features, we minimize Eq. (3) numerically via the L-BFGS-B algorithm proposed by Byrd et al. [2] with L_2 regularization. The parameter of L_2 regularization is selected via progressive validation on the winning data. Note that the estimator β^l_{lm} is not available in reality because the winning prices on the losing bids are unknown to the DSP. Each estimator learns one day of the data and its performance is tested using the data in the very next day.

The Mean Squared Error (MSE) between the true winning price and the estimated winning price evaluated on the winning bids of each day by either β^w_{lm} learning from the winning bids or β^l_{lm} from the losing bids in the previous day are shown in Fig. 2. We can see that β^w_{lm} outperforms β^l_{lm} in terms of having smaller MSE. In contrast to the result of Fig. 2, β^l_{lm} has smaller MSE on the losing bids compared to β^w_{lm} as shown in Fig. 3. In contrast, β^l_{lm} outperforms.

The results here also show that the winning price patterns on the winning and losing bids are different, otherwise the results of the two estimators should be similar. To ameliorate this effect, we consider the interaction [1] between the estimated winning rate and the features used to train the model. In other words, the model is trained by considering both the original features, i.e., Table. 4, and the features after interaction with the estimated winning rate. In this way, the different patterns of winning and losing bids can

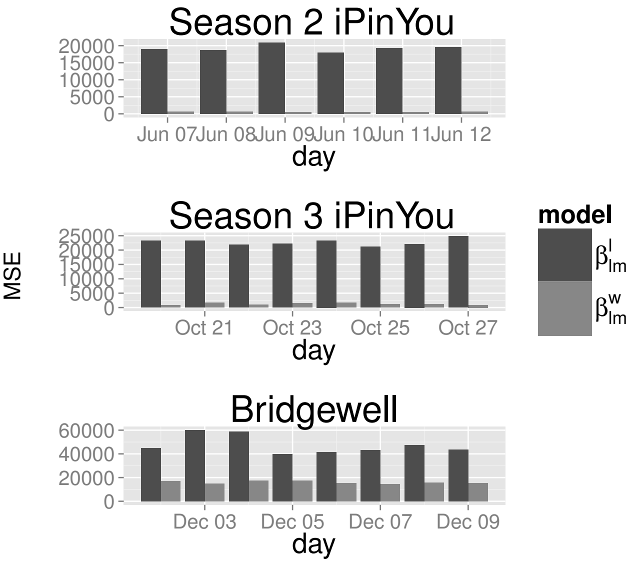

Figure 2: Mean Squared Error (MSE) of linear regression on the winning price based on the winning bids, β^w_{lm}, and based on the losing bids, β^l_{lm}. The result is evaluated on the winning bids.

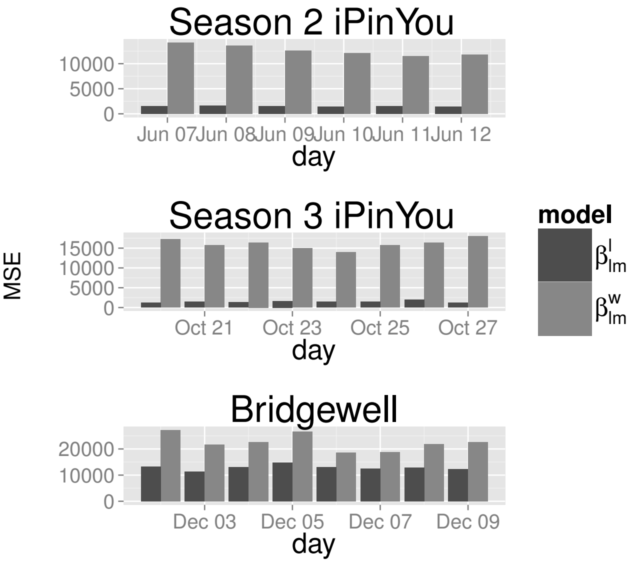

Figure 3: Mean squared error (MSE) of linear regression on the winning price based on the winning bids, β^w_{lm}, and based on the losing bids, β^l_{lm}. The result is evaluated on the losing bids.

thus be differentiated with the aid of the estimated winning rate.

4.3 Evaluation on β_{lm} vs. β_{clm}

To answer question (2), we conduct experiments to compare the estimator from the linear regression β_{lm} on the winning bids only, and the estimator from the censored regression model β_{clm} on both the winning and losing bids. Note that we cannot see the true winning price of the losing bids thus they are censored.

It is difficult to solve Eq. (5) with the parameter σ numerically. The σ usually diverges to infinity. In the experiment, we treat σ as a tuning parameter and replace it heuristically with the sample standard deviation of the observed winning price. The optimization function is solved numerically via L-BFGS-B algorithm with L_2 regularization. The parameter of L_2 regularization is the same as the one of β_{lm}. Then we calculate the MSE value of them on the losing bids of *the same day* to see the performance of the estimators at the censored data.

In Fig. 4, we can see that β_{clm} outperforms β_{lm} by having smaller MSE values. To explore the reason, we compare the average estimated winning price and the average true winning price at the censored data.

The average winning prices from different estimators are shown in Fig. 5. y_{true} represents the average true winning price, and y_{lm}, y_{clm}, and y_{mix} are the average estimated winning price by β_{lm}, β_{clm}, and β_{mix}, respectively. The censoring in the second price auction is due to high winning price. Hence, the winning price at observed data is usually lower than at censored data as shown in Table 1, Table 2, and Table 3. Consequently, the average winning price from β_{lm}, i.e. y_{lm}, is lower than the true average winning price, i.e. y_{true}, in all experiments as shown in Fig. 5.

On the other hand, β_{clm} learns from the censored data, and it hence has higher average estimated winning prices compared to β_{lm} as shown in Fig. 5. However, it is hard to tell the relationship between the average estimated winning price from β_{clm} and the average true winning price. Fig. 5 suggests that different data sources have different relationships. On the dataset iPinYou Season 2 and season 3, β_{clm} still underestimates the winning price on the censored data on average. Therefore, it is not surprising that the MSE value of β_{clm} is better than β_{lm} because the latter gives even lower estimation on average. β_{clm} overestimates the winning price on the Bridgewell dataset sometimes. Because the difference between y_{clm} and y_{true} is lower than the difference between y_{lm} and y_{true}, β_{clm} has smaller MSE than β_{lm} does in Fig. 4.

In summary, the overall MSE value of β_{clm} is lower than that of β_{lm}. This result depends on the bias, i.e., the difference between the predicted winning price and the true winning price, of β_{clm} on the censored data. It turns out that the predicted winning price of β_{lm} is lower than the true one and the winning price estimated by β_{clm} is higher than the one predicted by β_{lm}.

4.4 Evaluation on Mixture Model

To answer the question (3), we conduct experiments to compare estimators β_{lm}, β_{clm} and β_{mix} at the same time. All of them learn from the data of a specific day as a training day via minimizing the objective function. The detailed optimization is the same as Sec. 4.2 and Sec. 4.3.

The results are shown in Fig. 6, Fig. 7 and Fig. 8. A very interesting point is that the performance of β_{mix} is close to the better one of the other two estimators β_{lm} and β_{clm}. Sometimes, β_{mix} is the best of the three.

Fig. 5 shows that the value of estimated winning price by β_{mix} is between that by β_{lm} and that by β_{clm}. In all exper-

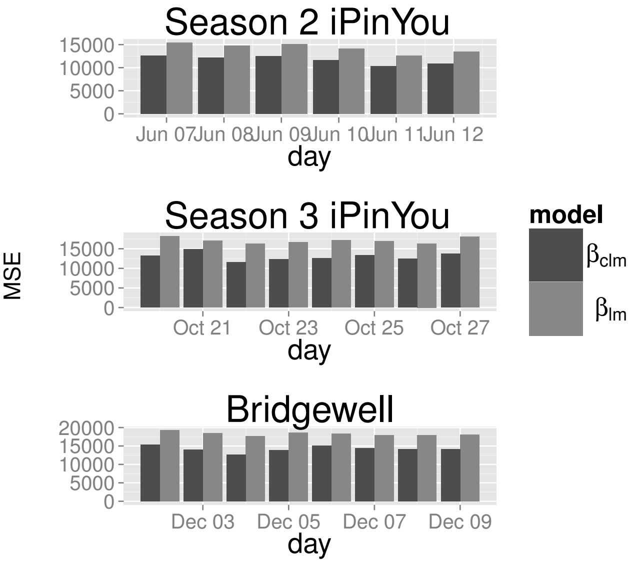

Figure 4: Mean Square Error of the winning prices on the losing bids. The evaluation is conducted on the same day of the training data.

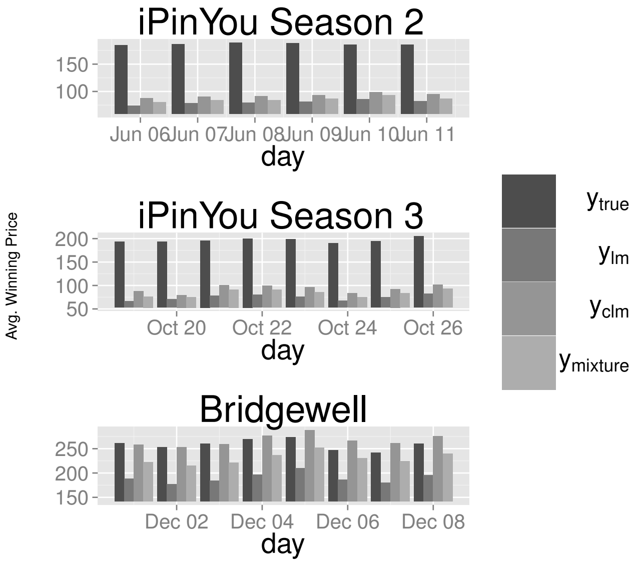

Figure 5: The average winning price estimated by different models on the losing bids.

iments, the bias, which is observed via the difference of the average predicted winning price and average true winning price as shown in Fig. 5, of β_{mix} on the censored data is less than β_{lm}. The bias of the β_{mix} exceeds β_{lm} only if the β_{clm} extremely overestimates the winning price on the censored data. It is not likely to happen in the real case because β_{clm} also learns from the observed data whose winning prices are lower. As shown in our experiments, there is no β_{clm} which extremely overestimates the winning price on censored data.

Although the characteristic of the dataset and the performance of the predicted winning rate, which are shown in Table 1, Table 2, and Table 3, are different, the results of our model are still robust compared to the linear regression, i.e. the model ignoring the censored data. If we change the ratio of the bidding price at the beginning of the simulation, β_{mix} consistently outperforms β_{lm} as shown in Fig. 9.

In summary, our experiments show that β_{mix} outperforms β_{lm}. We therefore conclude that the proposed framework indeed enhance the prediction of the winning price. The main reason is that the β_{lm} underestimates the winning price on the censored data and the proposed model lifts the estimation on the censored data properly. In addition, when compared to the censored regression, our model is much more robust.

5. RELATED WORK

While many existing reports study good bidding strategies for maximizing the revenues of either DSPs or SSPs [11, 6, 5], the issue of predicting the winning price of a bid has not been well explored yet. Zhang et al. [20] proposed a framework to optimize the bidding strategy for DSPs, of which the cost of a bid, usually the same as the winning price, is a required input. As the winning price is not available in many cases, they used the bidding price as the upper bound of the cost instead. Ghosh et al. [7] proposed adaptive bidding algorithms under a fixed budget with user-specified constraints. They considered both the settings with fully and partially observed winning price information. For the latter case, they simply assumed the winning price was drawn i.i.d. from a CDF $\mathbb{P}$, as their main goal is to win more bids when controlling the cost. All the above reports justify that knowing the winning price in RTB is important in practice. In this study we aim to further provide a systematic method based on machine learning and statistics to predict this value.

Cui et al. [4] studied the prediction of winning price, but on the seller side. Unlike Ghosh et al. [7] modeling the winning price by i.i.d. CDF $\mathbb{P}$, they modeled the winning price with the mixture-of-log-normal distribution on various targeting attributes. Since their study target is for SSPs, they did not address challenges of data censoring. Also, they paid attention to the global distribution of the winning price while we focus on the prediction of the winning price conditioned to the features, i.e., the case when only some features are available for DSPs.

Reisinger and Driscoll [14] studied the *publisher-advertiser information asymmetry* for SSPs. Their future work includes using the censored regression to construct the full bid distribution for SSPs because the floor price can be treated as dynamic *left-censoring*. Our work can be regarded as similar to their future work, but we are studying the bid distribution conditioned on the partially observed features and designing the prediction method for DSPs explicitly.

To our knowledge, we are the first to predict the winning price on the DSP side and with the censored data. We study the censored regression and extend it heuristically for the RTB data. We show that the accuracy of the winning price model can be enhanced prominently by cooperating with the censored information.

6. CONCLUSION AND FUTURE WORK

In this paper, we studied the prediction of winning price for DSPs in the RTB process when only partial features and only the winning price of historical winning bids were

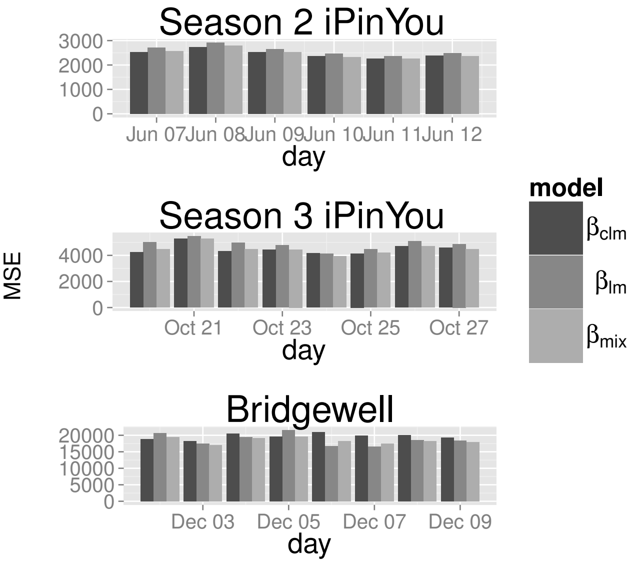

Figure 6: Mean squared error of the predicted winning price.

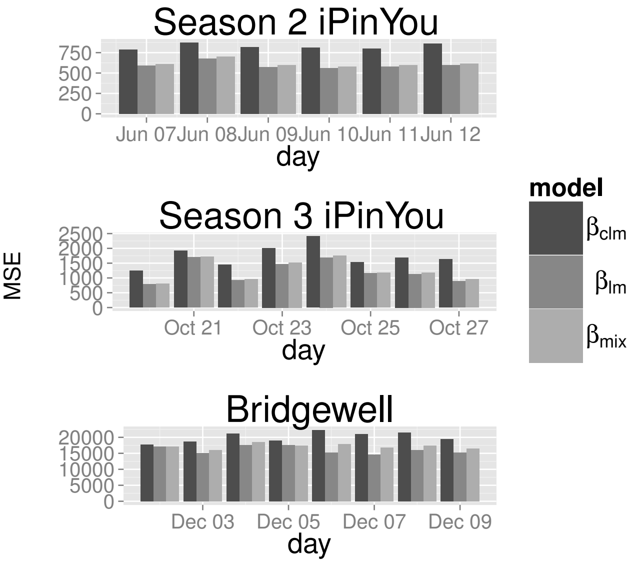

Figure 7: Mean squared error of the predicted winning price on the winning bids.

able to be observed. We showed that the prediction performance was improved by taking the censored information into account. Because the patterns of the winning price on the observed data and unobserved data are different in RTB, which violates the assumption of censored regression, we further proposed a mixture model to combine the power of linear regression on observed data and censored regression on both observed and censored data. Since the linear regression model works better on the winning bids in general and the censored regression model takes the censored information into account, we proposed to weight the two models according to the winning rate, which shows the probability that a DSP wins a bid. If the winning rate was higher, the linear regression model accounted for more percentage in the mixture model, and vice versa. From experiments on three real RTB datasets we showed that the proposed mixture model outperformed the model with observed data only.

For future work, we will seek finer and more sophisticated models that can better fit the censored winning price. Also, we would like to study the effect of introducing winning price prediction on real applications of RTB and online advertising such as designing bidding strategy, exploring the potential of new campaigns, and estimating the CTR and the conversion rate.

7. ACKNOWLEDGMENTS

This study was supported in part by the Ministry of Science and Technology of Taiwan, R.O.C., under Contracts MOST103-2221-E-001-006-MY2 and MOST103-3111-Y-001-025. The authors would also like to thank Bridgewell Inc. for providing the RTB logs in this work. All opinions, findings, conclusions and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

8. REFERENCES

[1] T. Brambor, W. R. Clark, and M. Golder. Understanding interaction models. improving empirical analyses. *Political Analysis*, 14:63–82, 2006.

[2] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208, Sept. 1995.

[3] O. Chapelle. Modeling delayed feedback in display advertising. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1097–1105, New York, NY, USA, 2014. ACM.

[4] Y. Cui, R. Zhang, W. Li, and J. Mao. Bid landscape forecasting in online ad exchange marketplace. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 265–273, New York, NY, USA, 2011. ACM.

[5] J. Feldman, M. Henzinger, N. Korula, V. S. Mirrokni, and C. Stein. Online stochastic packing applied to display ad allocation. In M. de Berg and U. Meyer, editors, *ESA (1)*, volume 6346 of *Lecture Notes in Computer Science*, pages 182–194. Springer, 2010.

[6] J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pages 117–126, Washington, DC, USA, 2009. IEEE Computer Society.

[7] A. Ghosh, B. I. Rubinstein, S. Vassilvitskii, and M. Zinkevich. Adaptive bidding for display advertising. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 251–260, New York, NY, USA, 2009. ACM.

[8] Google. The arrival of real-time bidding. 2011.

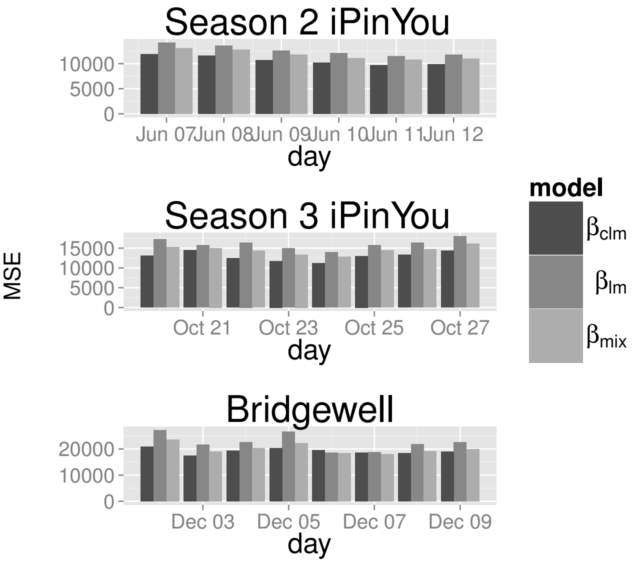

Figure 8: Mean squared error of the predicted winning price on the losing bids.

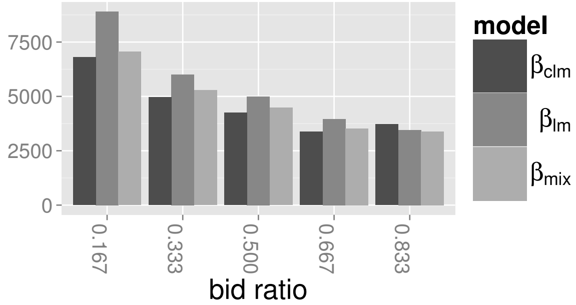

Figure 9: Mean square error of the predicted winning prices with different winning/losing bid ratio on the iPinYou dataset on 2013-10-20.

[9] W. H. Greene. Censored data and truncated distributions. Working papers, New York University, Leonard N. Stern School of Business, Department of Economics, 2005.

[10] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, and J. Q. Candela. Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising, ADKDD 2014, August 24, 2014, New York City, New York, USA*, pages 1–9, 2014.

[11] R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the Twenty-second Annual ACM Symposium on Theory of Computing*, STOC '90, pages 352–358, New York, NY, USA, 1990. ACM.

[12] J. Klein and M. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, 2003.

[13] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, S. Chikkerur, D. Liu, M. Wattenberg, A. M. Hrafnkelsson, T. Boulos, and J. Kubica. Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2013.

[14] J. Reisinger and M. Driscoll. Pricing externalities in real-time bidding markets. In *NIPS 2010 Workshop: Machine Learning in Online ADvertising*, 2010.

[15] W. Schnedler. Likelihood estimation for censored random vectors. Working Papers 0417, University of Heidelberg, Department of Economics, Feb. 2005.

[16] J. Tobin. Estimation of relationships for limited dependent variables. Cowles Foundation Discussion Papers 3R, Cowles Foundation for Research in Economics, Yale University, 1956.

[17] J. Wang, S. Shen, and S. Seljan. Real-time bidding: A new frontier of computational advertising research. Workshop in CIKM, 11 2013.

[18] K. Weinberger, A. Dasgupta, J. Langford, A. Smola, and J. Attenberg. Feature hashing for large scale multitask learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1113–1120, New York, NY, USA, 2009. ACM.

[19] S. Yuan, J. Wang, and X. Zhao. Real-time bidding for online advertising: Measurement and analysis. *CoRR*, abs/1306.6542, 2013.

[20] W. Zhang, S. Yuan, and J. Wang. Optimal real-time bidding for display advertising. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 1077–1086, 2014.

[21] W. Zhang, S. Yuan, J. Wang, and X. Shen. Real-time bidding benchmarking with ipinyou dataset. *arXiv preprint arXiv:1407.7073*, 2014.