Approaches For Secure Sales Lift Measurement

Jessica Hwang, Jerome Friedman, Jim Koehler, Yunxiao Li, David Chan

May 22, 2021

Abstract

We propose a number of potential approaches to enable Sales Lift measurement in a privacy-safe and secure manner. We discuss these approaches in the context of digital publisher media channels, but in theory these approaches can be extended to most types of media channels that has the ability to link media exposure with outcomes. We discuss Sales Lift measurement both in the context of single publisher and multi-publisher scenarios and weigh the trade-off of the different solutions in terms of utility, privacy, security, and computation costs.

1 Introduction

There is an advertiser-led industry initiative to develop a neutral, cross-media measurement system that will address advertiser needs, provide a global framework for consistency and allow for local market flexibility to address market-level needs. Called the cross-media measurement framework (XMM), this initiative is being developed under the auspices of the World Federation of Advertisers (WFA, (14)), in combination with national advertiser associations and global advertisers. See (13) for a more detailed overview of this initiative.

XMM is intended to provide a framework from which the publisher data, advertiser and/or third party outcomes data can be bought together in a privacy-safe and secure manner, to enable measurement of ads effectiveness. There is an effort to develop XMM solutions for brand measurement, such as Reach/Frequency. (8) outlines a privacy-centric approach for cross-publisher Reach and Frequency estimation. Aside from Brand measurement, there is also outcomes-based measurement, examples of which include multi-touch attribution, Sales Lift and Media Mix Models. See for example (7), which proposes an approach for measuring lift in a privacy-safe manner, based on A/B experiments.

Offline sales based outcomes is of particular interest to advertisers. Often, advertisers would like to measure the incremental offline sales that occurs due to running an online ad campaign. This is a challenge as the offline sales data needs to be linked with the online ad campaign data, which is a non-trivial task.

If the offline sales occurs in retailer channels the advertiser does not own, then gaining access to the sales data itself is another challenge. A prominent example of this is Consumer Package Goods advertisers, where the sales of such products can occur in channels such gas stations, supermarkets and pharmacies. Such advertisers rely on third party data aggregators to provide the sales data and/or to also provide the sales lift measurement. Such third party data providers are referred to as TDP in this paper.

One approach to sales lift measurement is geo-based experiments (i.e. (15)). An advantage of this approach is that it only relies upon aggregate sales data being provided by the third-party. However, there is setup complexity and costs and also opportunity costs with running geo-based experiments.

This paper presents an outline of different potential approaches to enable a secure sales lift solution within the XMM framework. However, the potential approaches outlined could also be applicable for lift measurement outside of the XMM framework. For example, when a publisher and TDP would want to work together directly to enable other types of lift measurement. Section 2 will briefly go through an overview of the XMM framework, including the privacy and security considerations. Section 3 provides a short introduction to sales lift measurement and some common solutions currently used by the industry today. Section 4 goes through the notation that will be used through-out the paper. Section 5 then walks through the proposed solutions for the scenario where we have one publisher and one TDP. Section 6 goes through the more complex scenario where we have multiple publishers and one TDP. We conclude with a short remarks in Section 7.

2 Overview of Cross-Media Measurement Initiative

The goals of the XMM initiative is to connect data from advertisers, publishers and other third-parties on a standardized platform that will be used as the basis for cross-media measurement. We refer to WFA as the organization responsible for administrating the XMM framework. WFA will develop standards on data and methods, thus making cross-media measurement more comparable. Although WFA is primarily addressing cross-media measurement, there are benefits to using it also for single media measurement, on a platform where data and methods can be standardized.

Partners will each contribute their assets to a central system but will still retain control and ownership of their own assets. WFA will maintain a common ID framework that will allow data from the various partners to be connected together for measurement. The likelihood of re-identification of users increases when partners datasets are combined, and technical measures, such as multi-party computation (6), are put in place to limit such re-identification.

Additional methods such as encryption, k-anonymity and differential privacy will be further employed to ensure measurement is done in a privacy safe and secure manner and no data in the clear is leaked across partners. Depending on the measurement use-case, different technical and infrastructural execution will be needed.

3 Sales Lift Measurement

In Sales Lift measurement, the advertiser is interested in measuring the incremental sales that occur due to an ad campaign that was placed on a digital publisher's web property. In this paper, we focus on the situation where the advertiser's products are sold through various retail channels and the advertiser relies upon a third-party to provide their sales data. Consumer Packaged Goods manufacturers are one set of advertisers where this is the case. We henceforth refer to such third-party data providers as TDPs.

Current Sales Lift solutions depend on the publisher sending in their ad impression data into a so called "clean room". The TDP also sends their sales data into the secure clean room. We assume that there is sufficient match rates between the publisher's data and the TDP's data to make measurement feasible. Within the clean room, both the ad exposure data and sales outcome data are operated on in the clear and various lift estimators can be used. Standard lift estimators that are used included matched ANCOVA, Doubly Robust Inverse Propensity Weighted (DRIPW) and Targeted Maximum Likelihood (TMLE). See (5) for a comparison of various lift estimators that are used in Sales Lift measurement.

In the solution described above, the data is allowed to be in the clear within the clean room. However, due to security considerations with the XMM framework, even within a clean room, the data is not allowed to be in the clear and must have some form of differential privacy and/or encryption applied to it. This makes for a more secure and but more challenging measurement environment.

We limit the scope of this paper to looking at only one type of lift estimator, that is, the Doubly-robust Inverse Propensity Weighted estimator, or DRIPW. We don't consider matching based estimators such as matched ANCOVA, as the need to add differential privacy, greatly degrades the performance of such estimators. Match-based methods are also not as efficient as non match-based methods and do not have the doubly robust property.

We also assume that there is a mechanism provided as part of the WFA framework that would allow us to connect the ad exposure data with the sales outcome data.

Lastly, we note that in CPG Sales Lift measurement, the sales data provided by the TDP is typically collected at the household level, whereas the ad exposure data is at an ID or user level. In order to do measurement here, the ID-level exposure data needs to be rolled up to the household level and a household is considered exposed, if any of its household members is exposed. In such situations, we assumed that we have a scheme to go from ID level ad exposure statuses to household level ad exposure statuses in a private and secure manner.

There may be contamination in terms of households being identified as unexposed when in reality it was exposed, or vice-versa due to incomplete or poor match quality. Such contamination may have a large effect, but we don't concern ourselves with it within the scope of this paper.

4 Notation and Setup

In this section, we describe the setup and notation used throughout the rest of the paper. For IDs $i = 1, \ldots, n$, the publisher's data consist of exposure indicators $T_i \in \{0, 1\}$, where $T_i = 1$ if ID i is exposed to the ad campaign

and $T_i = 0$ otherwise, and covariates X_i related to ID i's demographics and and media-related activity, such as impressions on desktop and mobile platforms.

The data owned by the TDP consist of Y_i, a binary or continuous sales outcome of interest, and covariates Z_i related to the ID's prior purchase activity or other factors that could affect the ID's purchasing behavior such as demographics. There could be an overlap in the set of covariates in X and Z. An example of a binary sales outcome Y_i would be the indicator of whether ID i purchased a product in a given period of time after the ad campaign. An example of a continuous sales outcome Y_i would be the dollar amount spent on a product.

The DRIPW estimator to estimate the sales lift of the ad campaign on the outcome of interest is defined as follows:

$$\widehat{\tau} = \sum_{i=1}^{n} \widehat{w}_i \widehat{\tau}_i$$

where

$$\widehat{\tau}_i = \frac{T_i \cdot (Y_i - \widehat{m}(Z_i, 1))}{\widehat{b}(X_i)} - \frac{(1 - T_i) \cdot (Y_i - \widehat{m}(Z_i, 0))}{1 - \widehat{b}(X_i)} + \widehat{m}(Z_i, 1) - \widehat{m}(Z_i, 0).$$

Function $\widehat{m}$ approximates the conditional expected outcome function $m(z, t) = \mathbf{E}[Y \mid Z = z, T = t]$ and $\widehat{b}(X_i)$ is the estimated propensity function which models the conditional probabilities of exposure to the ad campaign.

For the Average Treatment Effect (ATE), we have $\widehat{w}_i = 1/n$ and for the Average Effect on the Treated (ATT), we have $\widehat{w}_i = \widehat{b}(X_i) / \sum_{j=1}^{n} \widehat{b}(X_j)$.

5 Single publisher Scenario

We first consider the scenario with a single publisher and a single TDP. The extent of a publisher or TDP's universe of IDs can also be sensitive data, so we further break out this scenario into two sub-scenarios. Whether or not it would be alright to expose the list of IDs from one party to another party. Even if all data from the sender is privatized, the need to match IDs can potentially expose the universe of IDs from the sender to the receiver.

We propose a statistical-privacy approach for the scenario where the publisher is allowed to send its data to the TDP, exposing its list of IDs. We propose a private-computing approach (see (6) for details) for the scenario where the partners would not want to expose its list of IDs.

5.1 ID list can be exposed

5.1.1 Solution

Our proposed solution in this setting proceeds in three steps. First, the publisher sends its exposure indicators to the TDP with differential private noise added. To do this, the publisher randomly flips the exposure bits T_i with probability q, independently, producing privatized exposure indicators $\widetilde{T}_i$. These are sent to the TDP in unencrypted form, thus exposing the publisher's ID list to the TDP, but not the exact values of the exposure indicators.

Second, the TDP fits an outcome model to predict the outcome Y_i given T_i and Z_i. For example, when Y_i is binary, the outcome model may be a logistic regression: $Y_i \sim$ Bernoulli(p_i) where

$$\text{logit}(p_i) = \alpha + \beta T_i + \gamma^\top Z_i.$$

When Y_i is continuous, the outcome model may be a linear regression:

$$Y_i = \alpha + \beta T_i + \gamma^\top Z_i + \epsilon_i \qquad \epsilon_i \sim N(0,1).$$

Since the TDP does not have access to T_i but rather to the noisy indicators $\widetilde{T}_i$, this is an errors-in-variables regression problem (see (9), (11)), which can be fitted using maximum likelihood (10) or pseudolikelihood approaches (2). After fitting the outcome model, the TDP plugs in $T_i = 1$ and $T_i = 0$ for each user to produce a pair of predicted counterfactual outcomes, $\hat{m}(Z_i, 1)$ and $\hat{m}(Z_i, 0)$, which are needed to compute the DRIPW estimator.

Finally, given estimates of the outcome model parameters, the TDP and publisher jointly compute the DRIPW estimator. Noting that the DRIPW estimator consists of inner products between data from the publisher and data from the TDP, the estimator can be computed securely via homomorphic encryption. (See (1)). This allows the final estimate to be computed only on the encrypted data from the publisher and the TDP.

5.1.2 Estimation of the outcome model

Estimation of the outcome model parameters can proceed by maximum likelihood (ML) (10) or pseudolikelihood (2). To maximize the full likelihood, we propose using the expectation maximization (EM) algorithm (4). To maximize a pseudolikelihood, we propose a regression calibration (RC) approach (see (3)). Further details can be found in the Appendix.

5.1.3 Bootstrap Confidence Interval

Confidence intervals are helpful for decision makers to understand the level of uncertainty in the sales lift estimates. We propose to compute nonparametric confidence intervals of both outcome model coefficients and the doubly robust estimator of ATE through bootstrap resampling. Though bootstrapping could be computationally expensive, we found that with the scope of this paper bootstrap confidence intervals: (i) are generally

applicable to various types of outcome models; (ii) enjoy high reliability, for instance, demonstrating desired coverage probability. In our implementation, a confidence interval is computed with 500 bootstrap resampling replicates.

5.1.4 Simulation Results

In this section, we show numerical results on simulated data to demonstrate the performances of two candidate methods. More details about the data generating process for the simulated data can be found in the Appendix. Holding the same data generating process, the magnitude of DP-noise added to exposure indicators varies from 5% to 40%. We consider three variants of applying either ML or RC estimation procedure. These variants only differ in (i) whether the TDP would include propensity score as a covariate in outcome modeling; and (ii) whether the publisher would send an exact or privatized version of propensity scores to the TDP. First, a variant is called ”standard” if the propensity score is excluded from the outcome modeling. Between the remaining two variants that take propensity score into outcome modeling, the ”exact” variant represents the option that publisher sends the exact propensity scores estimated with the real exposure indicators and covariate data, while the ”private” variant means sending propensity scores estimated with randomly flipped exposures (i.e., real exposures adding DP noise). For convenience, we will denote all these six approaches with abbreviations ML-std, ML-exact, ML-private, RC-std, RC-exact, and RC-private. It is unlikely that the ”exact” variant would be used in practise, but is included here for comparison purposes. Simulation results for continuous sales outcome are shown in Fig 1, Fig 2 and Fig 3. Simulation results for binary sales outcome are shown in Fig 4, Fig 5 and Fig 6.

To compare the performances of the six candidate approaches, we focus on three metrics in the evaluation: (i) bias in the estimator of treatment effect, (ii) actual coverage probability of confidence interval, and (iii) the average width of the confidence interval. A candidate method performs reasonably well only if minimal or no bias can be observed in ATE. Confidence interval provides a plausible range of the ground-truth lift value. Any decision upon this range is statistically reliable only if the actual coverage probability of these intervals is greater than or equal to the nominal level (e.g, 90%).

Under these simulation scenarios, all candidate methods show none or minimal biases in estimating the treatment effect. When the magnitude of DP noise increases to as high as 0.4, the three RC approaches seem to perform more robustly in terms of reducing biases. This is possibly because the RC approach is usually not impacted by computational issues such as EM algorithm convergence that may occur in running the ML procedure. The actual coverage probabilities of these bootstrap confidence intervals approximately align with the desired coverage values. This implies that we can draw valid statistical decisions on sales lift estimates even though the exposure indicators are privatized with a substantial amount of DP noises. An increased CI width is expected because DP noise brings a higher level of uncertainty to estimating the outcome model. It is also expected that the level of widening is monotonically increasing with the magnitude of DP noise. However, excluding the two standard variants, we only observe at most 20% relative increase in CI width compared to the one that is computed without flipping exposure bits. It is worth noting that the CI with the variant exact or private is consistently narrower than the width with variant standard where no propensity score information is utilized in outcome model. It is likely that estimating an outcome model with propensity scores enhances the likelihood of correctly imputing a true exposure give the positive correlation between publisher propensity scores and the true exposure indicators.

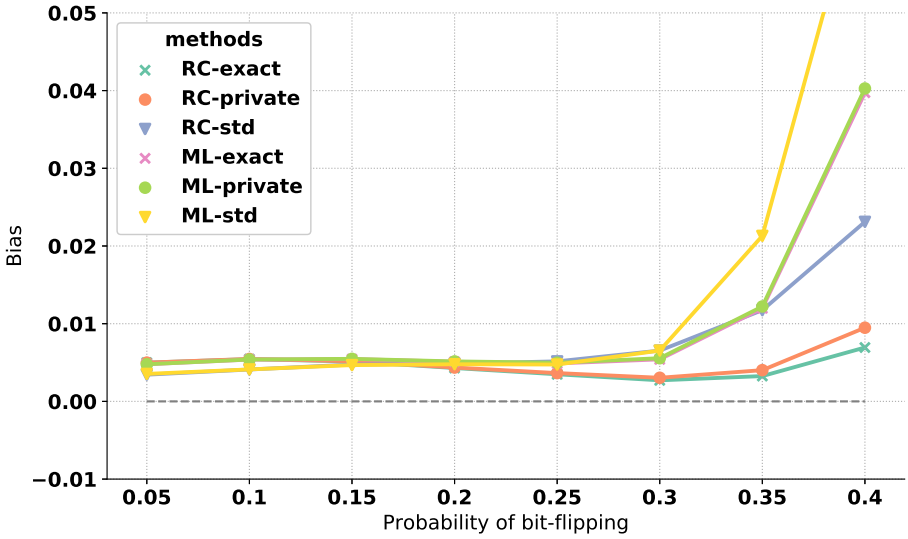

Figure 1: Biases of ATE estimator under linear outcome scenario.

Another notable observation is that the amount of CI increase is much smaller in the treatment effect compared to the CI increase in any coefficients of the outcome model (listed in Appendix). This is possibly due to the doubly-robust property of the DRIPW estimator, but further investigations would need to be carried out to confirm this.

Lastly, we note that sending exact or privatized propensity scores would not impact these three metrics. This is likely due to a positive correlation between the two sets of propensity scores.

5.2 ID list can't be exposed

If the ID list can't be exposed, then a secure multi-party computation approach such as "Private Join and Compute" (see (12)) can be used to join the data from the publisher and the TPD. The result of this approach is a joined dataset consisting of only the rows from the intersection of the data from the publisher and the TDP. In order for this to work, the TDP's covariates needs to be privatized, and/or coarsened in such a way that makes difficult to re-identify a user based on the covariates alone.

The only necessary information from the publisher is the exposure status, which is easily privatized. For the TDP's covariates, which can consist of a mix of numeric and categorical data, it can be a little more challenging to privatize the data without losing too much information. Coarsening is one possible approach. Another approach would be to transform data, say by applying Principal Components and only taking the first X principal components. The exact impact of various types of privatization on the covariates and the trade-off between privacy and model prediction accuracy is still an open research topic. It would be expected that ML-based models would be better able to work with such privatized covariates and recover the salient information for the model fitting.

At this point, with the joined dataset, the outcome models can be fitted by the TDP, following a similar process

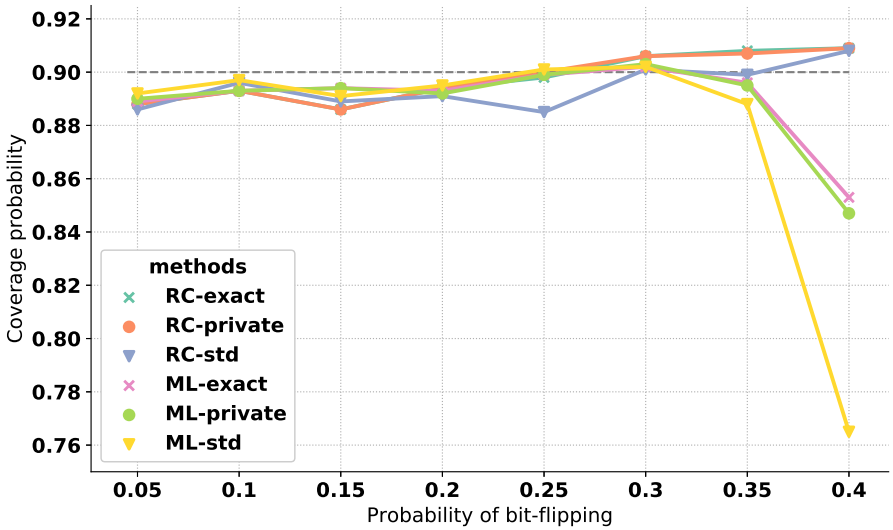

Figure 2: Coverage probabilities of ATE confidence interval under linear outcome scenario.

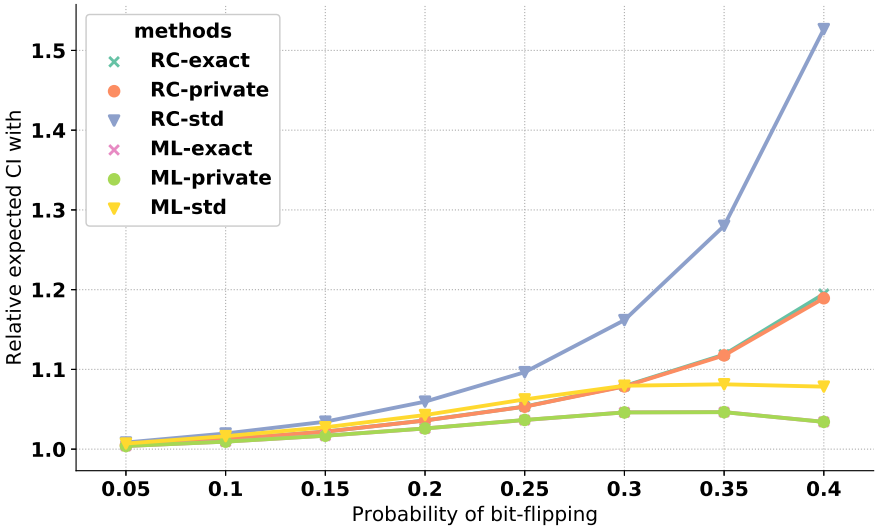

Figure 3: Relative widths of ATE confidence interval under linear outcome scenario. The relative width is defined as the average width of ATE CI when estimating outcome model with noisy exposures $\widetilde{T}$ divided by the average width of ATE CI when estimating outcome model without DP-noise.

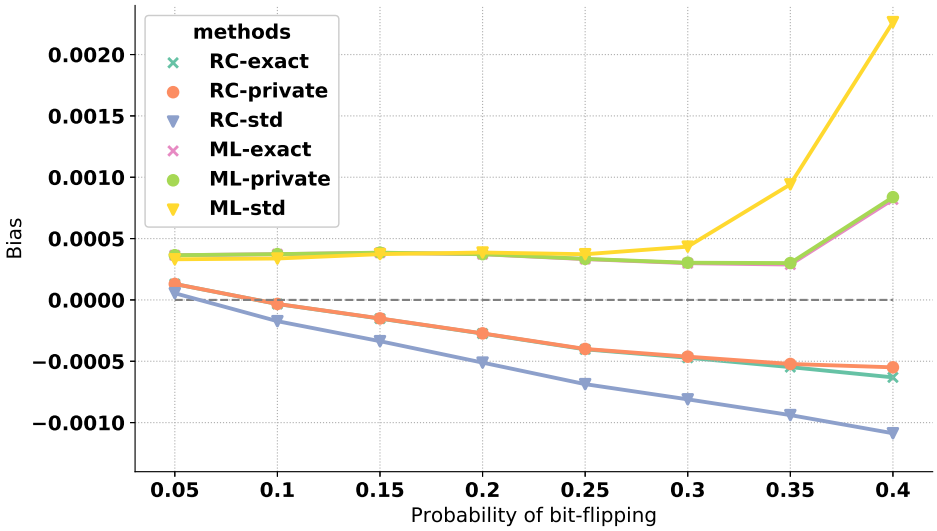

Figure 4: Biases of ATE estimator under logistic outcome scenario.

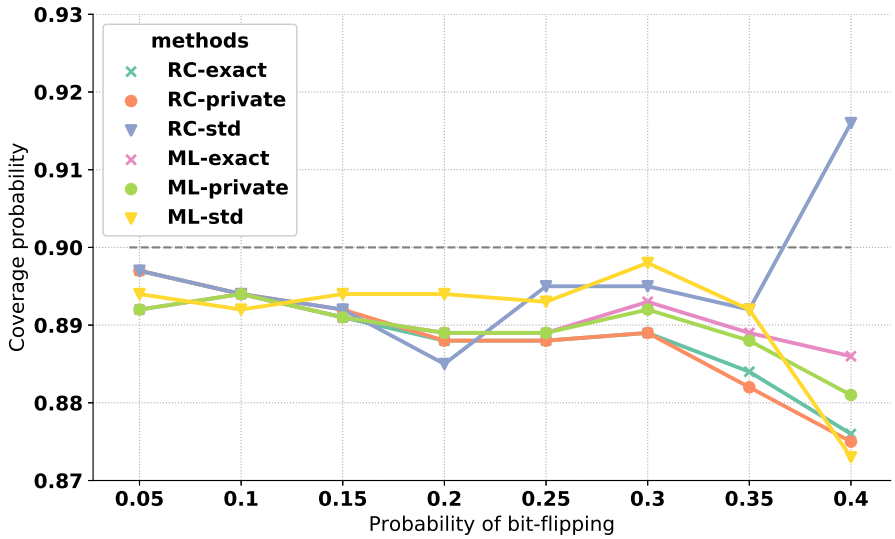

Figure 5: Coverage probabilities of ATE confidence interval under logistic outcome scenario.

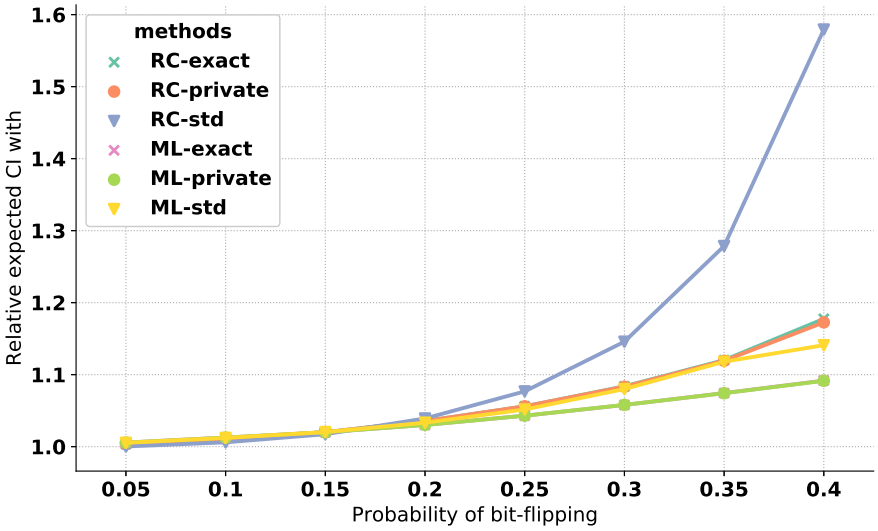

Figure 6: Relative widths of ATE confidence interval under logistic outcome scenario. The relative width is defined as the average width of ATE interval when estimating outcome model with noisy exposures $\widetilde{T}$ divided by the average width of ATE when estimating outcome model without DP-noise.

as in Section 5.1.

6 Multi-publisher Scenario

In the multi-publisher setting, we again have a single TDP which owns the sales-related predictors and outcome data, but there are now multiple publishers, each with their own ad exposure data. We will continue to assume a common ID space on which the publishers and TDP can join their data.

Estimation of marginal treatment effects for each individual publisher can proceed as in the previous section, but interaction effects are now of interest as well. Assuming two publishers A and B for concreteness, we may want to estimate the effect of exposure to publisher A *and* publisher B, the effect of exposure to publisher A *or* publisher B, or the effect of exposure to publisher A *but not* publisher B. Sample size considerations will determine which interaction effects can be estimated.

An issue that arises in the multi-publisher setting is the estimation of joint propensity scores. Let T_i^A and T_i^B be the indicators of exposure to publisher A and publisher B, respectively, for user i, and let $b(X_i^A)$ and $b(X_i^B)$ be the corresponding propensity scores, obtained by regressing T_i^A on predictors X_i^A and regressing T_i^B on predictors X_i^B. The indicator of exposure to both publishers is $T_i^{A\cap B} \equiv T_i^A T_i^B$. Under the restrictive assumption of independence, the propensity score for $T_i^{A\cap B}$ is the product of the marginal propensity scores. However, absent the independence assumption, it is no longer sufficient for each publisher to compute a marginal propensity score. Instead, we must fit a model that predicts $T_i^{A\cap B}$ from X_i^A and X_i^B.

We propose that each publisher will send noisy exposure indicators $\tilde{T}_i^A$ and $\tilde{T}_i^B$, which have been randomly

flipped with some known probability, to the TDP or to a semi-trusted party managed by WFA, along with predictors X_i^A and X_i^B. To preserve privacy, the predictors should have some form of privatization applied to it. This could include dimension reduction, hashing, or other forms of aggregation.

The TDP or WFA will then fit a regression model to estimate the true joint propensity score, $P(T_i^{A\cap B} = 1|X_i^A, X_i^B)$. However, the TDP or WFA does not have access to $T_i^{A\cap B}$, only $\tilde{T}_i^{A\cap B} = \tilde{T}_i^A\tilde{T}_i^B$. Therefore, the TDP or WFA must fit a propensity score model using noisy exposure indicators.

In the following section, we outline a method for accomplishing this task: estimating a propensity score when the exposure indicator (in this case, $T_i^{A\cap B}$), has been flipped with some probability. We show that by applying a simple adjustment factor, it is possible to accurately estimate the true propensity score even when a substantial amount of noise has been added to the exposure indicators.

6.1 Propensity score estimation from noisy exposure indicators

For a binary exposure indicator $T \in \{0, 1\}$, the goal is to estimate the probability that $T = 1$ as a function of predictor variables X. The exposures are randomly changed to their complementary values $T = 0/1 \to \tilde{T} = 1/0$ with probability $0 < q < 1/2$ to obscure their original values. The obscured values $\{\tilde{T}_i\}_{i=1}^n$ are then used to train the predictive model.

Given an original y-value with probability $p = \Pr(T = 1 \mid X)$ and flip probability q, one has

$$\tilde{p} = \Pr(\tilde{y} = 1) = p(1 - q) + q(1 - p)$$

with corresponding odds

$$\tilde{o} = \frac{\tilde{p}}{1 - \tilde{p}} = \frac{o + q - qo}{1 - q + qo} \tag{1}$$

and $o = p/(1 - p)$ being the odds of the corresponding unperturbed probability. Inverting (1) one obtains

$$o = \frac{\tilde{o}(1 - q) - q}{1 - q(1 + \tilde{o})} \tag{2}$$

Thus, for any perturbed value $\tilde{o}$ in the range

$$\frac{q}{1 - q} < \tilde{o} < \frac{1 - q}{q} \tag{3}$$

one can use (2) to determine its corresponding original unperturbed value o. A first-order approximation to (2) is

$$\log(o) = \frac{\log(\tilde{o})}{1 - 2q}. \tag{4}$$

Knowing the value of $\log(\tilde{o})$, one can directly calculate $\log(o)$ from (2). However the value of $\log(\tilde{o})$ is never known. It is estimated from the (perturbed) training data by some machine learning procedure. Errors in such estimates of $\log(\tilde{o})$ are translated into errors in $\log(o)$ through (2).

This is illustrated in Figure 7. One thousand equispaced $\log(o)$ values were generated in the range $-5 < \log(o) < 5$. These values were transformed to perturbed $\log(\tilde{o})$ values using (1). In order to simulate estimation uncer-

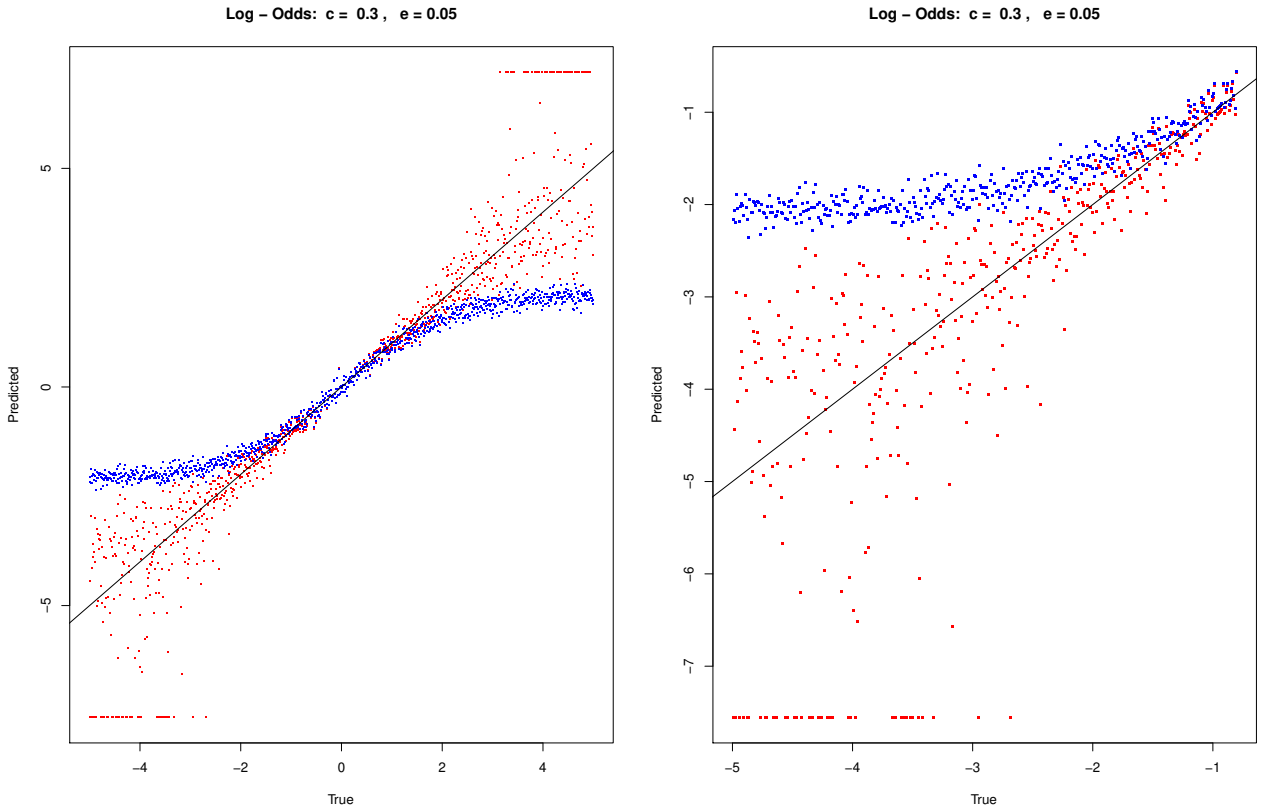

Figure 7: Transformed perturbed log-odds with uncertainty versus true log odds, for flip probability $q = 0.3$. Red points represent exact transformation. Blue points represent first order approximation. (a) Full range of log odds. (b) Zoomed-in view for extreme log odds.

tainty, errors were then added $\log(\tilde{o}) \to \log(\tilde{o}) + 0.05 \cdot \varepsilon$ with ε generated from a standard normal distribution. The result was transformed back to unperturbed values via (2). Note that log-odds is the natural scale for adding errors since that is what most machine learning procedures directly estimate.

Figure 7(a) shows a plot of estimated log-odds versus true (original) log-odds on the perturbed scale for the 1000 points. Red points represent using the exact transformation (2) whereas the blue points use the first order approximation (4). The black line represents equality. Both approaches are seen to fail for extreme log-odds $|\log(o)| > 2$. The former (blue) has low variance everywhere but is highly biased at the extremes. The latter (red) is roughly unbiased everywhere but exhibits high variance at the same extreme estimated values. The truncated predicted values shown at the vertical extremes represent $\tilde{o}$ values estimated to be outside the range (3).

In the sales lift setting, we are likely to see severe class imbalance, since the number of users exposed to the intersection of multiple publishers is dwarfed by the number of unexposed users. Figure 7(b) zooms in on $\log(o) < -1.39$ ($p < 0.2$). Neither the exact (2) nor approximate (4) transformation provide satisfactory results in this case.

Both the exact transformation and first order approximation provide relatively accurate predictions in the central log-odds range $-2 \lesssim \log(\tilde{o}) \lesssim 2$. This suggests a strategy of centering the log-odds before perturbation. This can be arranged by selecting equal numbers of each class in the training sample or by weighting the observations so that the sum of weights in each class are the same. Training is performed on the balanced perturbed sample and the resulting estimates transformed using (2) or (4). Finally the transformed estimates are uncentered using the known original imbalance.

Figure 8 shows the result of this strategy on the same data shown in Figure 7. The result is seen to be dramatically improved accuracy even at the most extreme log-odds.

6.2 Simulation studies

Simulation experiments were conducted to ascertain the loss of accuracy due to perturbation of the exposure indicators, and the effectiveness of the centering strategy to help mitigate it in the presence of highly unbalanced training samples. In all experiments the data had $K = 10$ predictor variables generated from a joint normal distribution with covariance matrix elements $c_{ij} = 1 - |i - j|/K$. The training data sample size was $N = 30K$. The sample size of 30K was chosen to illustrate the removal of bias, and not be representative of a typical campaign size. The label change probability was $q = 0.3$. All models were evaluated on an independent test data set of $30,000$ observations. All plots show a random subsample of $1,500$ test observations so as to see details more clearly.

6.2.1 Linear model

In this study the true log-odds were taken to be a linear function of $\mathbf{x}$

$$\log \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \beta_0 + \sum_{j=1}^{K} \beta_j x_j$$

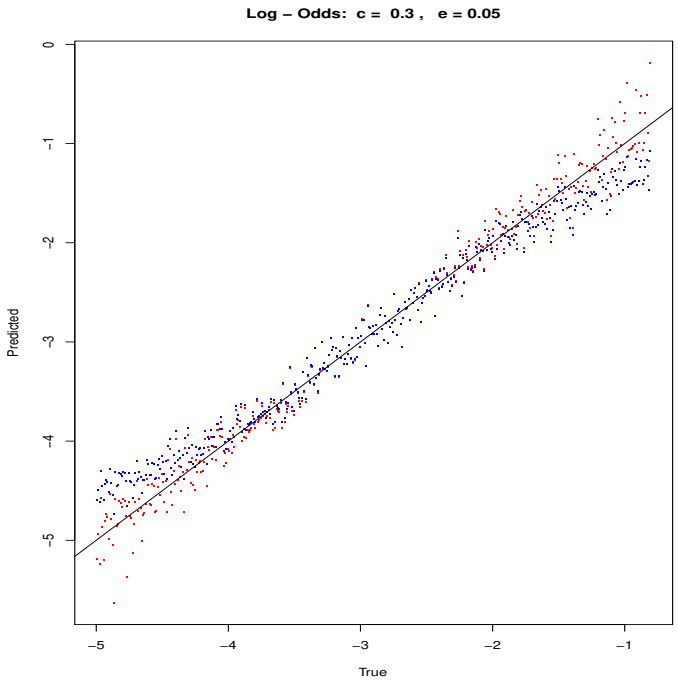

Figure 8: Transformed perturbed log-odds with uncertainty versus true log odds for asymmetric class counts after class centering. Red points represent exact transformation. Blue points represent first order approximation.

with coefficients randomly generated from a uniform distribution $\beta_j \sim U(-1, 1)$. The value of the intercept was taken to be $\beta_0 = \log(10)$ resulting in a global marginal class imbalance of $\Pr(y = 1)/\Pr(y = 0) = 8.91$.

The left frame of Figure 9 plots the true probability $p(\mathbf{x})$ versus estimated probability $\hat{p}(\mathbf{x})$ using linear logistic regression (GLM). The red line represents equality. The estimation process consists of applying GLM to perturbed the outcomes ($q = 0.3$) and then using (2) to recover estimates corresponding to the original scale. The right frame shows corresponding results using the class centering strategy (Figure 9). GLM is trained on a perturbed sample with equal numbers of each of the two classes. Resulting estimates are transformed back to the unperturbed scale (2) and then re-centered using the known class proportions in the original training sample. The displayed average absolute error

$$AAE = \frac{1}{N}\sum_{i=1}^{N} |\, p(\mathbf{x}_i) - \hat{p}(\mathbf{x}_i)\,| \tag{5}$$

is computed on a $N = 30K$ test data set. Note that the axes in Figure 9 are plotted on the probability rather than the log-odds scale so the large errors at the extremes seen in Figure 9 are less evident. The centering strategy is seen here to reduce *probability* prediction error by almost a factor of two.

Figure 10 shows analogous results using gradient boosting (GBM) to estimate log-odds on the perturbed data. As expected these estimates are somewhat less accurate than those of the correctly specified model (GLM) seen in Figure 9. The centering strategy is still seen to provide some improvement.

Finally, Figure 11 shows corresponding results for random forest probability prediction. Both centered and non-

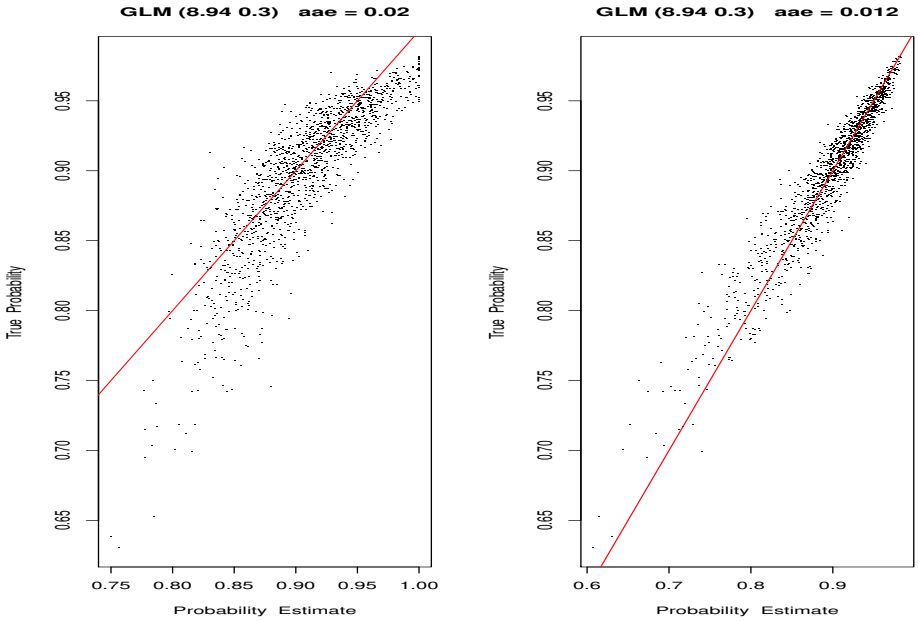

Figure 9: True probability versus estimated probability using linear logistic regression (GLM) on perterbed ($q = 0.3$) linear log-odds model data. Left: no class centering. Right: with class centering.

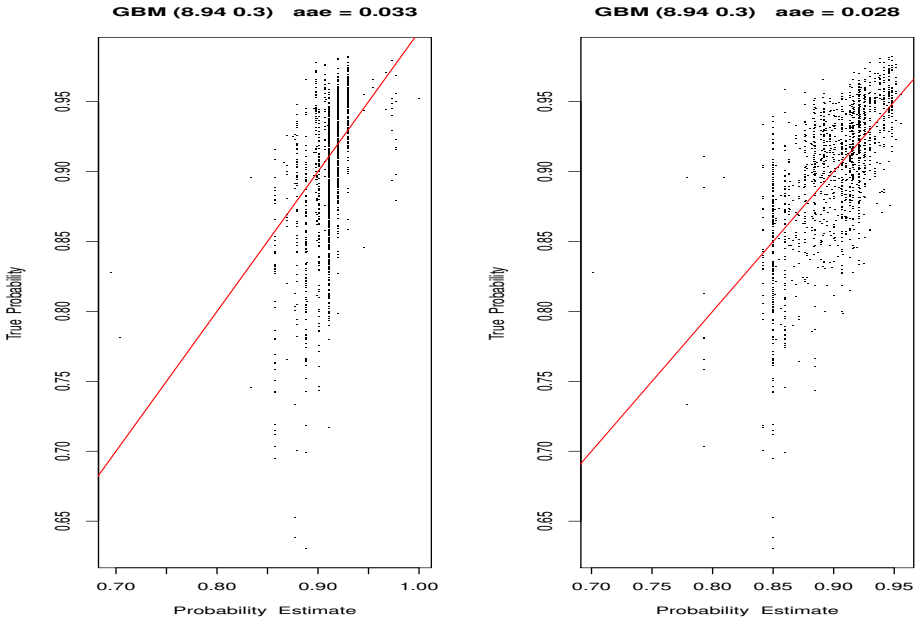

Figure 10: True probability versus estimated probability using gradient boosting regression (GBM) on perturbed ($q = 0.3$) linear log-odds model data. Left: no class centering. Right: with class centering.

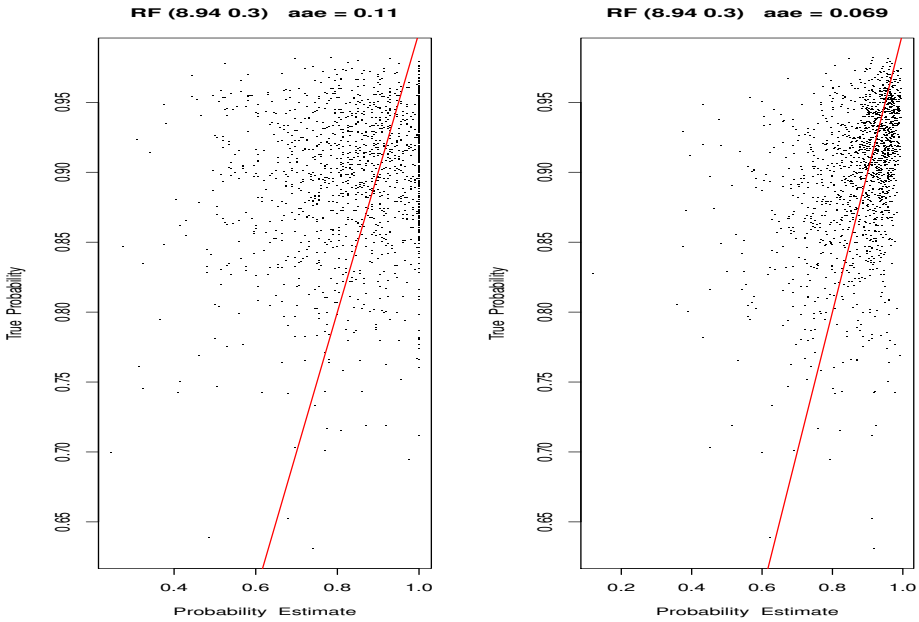

Figure 11: True probability versus estimated probability using random forest (RF) on perturbed ($q = 0.3$) linear log-odds model data. Left: no class centering. Right: with class centering.

centered estimates are substantially less accurate that those of the other methods but centering still provides substantial improvement.

6.2.2 Non-linear model

In this section the analysis of Section 6.2.1 is repeated on data simulated from a non-linear log-odds model. All aspects are the same as above except that the true log-odds are a highly nonlinear function of the predictor variables $\mathbf{x}$. The simulated log-odds function is specified as

$$\log \frac{p(\mathbf{x})}{1 - p(\mathbf{x})} = \sum_{j=1}^{10} c_j \, B_j(x_j) \, / \, std_{x_j}(B_j(x_j)) \tag{6}$$

with the value of each coefficient c_j being randomly drawn from a standard normal distribution. Each basis function takes the form

$$B_j(x_j) = sign(x_j) \, | \, x_j \, |^{r_j} \tag{7}$$

with each exponent r_j being separately drawn from a uniform distribution $r_j \sim U(0, 2)$. The denominator in each term of (6) prevents the suppression of the influence of highly nonlinear terms in defining log-odds.

Figures 12, 13 and 14 show the results for this nonlinear log-odds model analogous to Figs. 9, 10 and 11 respectively for the linear model. In all cases the centering strategy substantially improved performance. Here none of the estimation methods is correctly specified. As might be expected gradient boosting (GBM) out performed the linear log-odds model (GLM). Random forest probability prediction (RF) under achieved here.

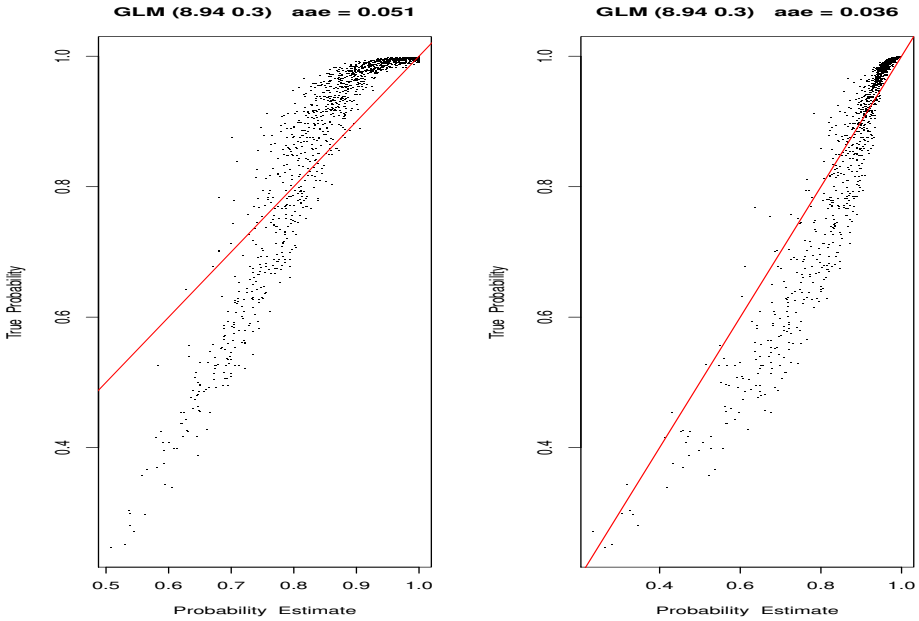

Figure 12: True probability versus estimated probability using linear logistic regression (GLM) on perterbed ($q = 0.3$) non linear log-odds model data. Left: no class centering. Right: with class centering.

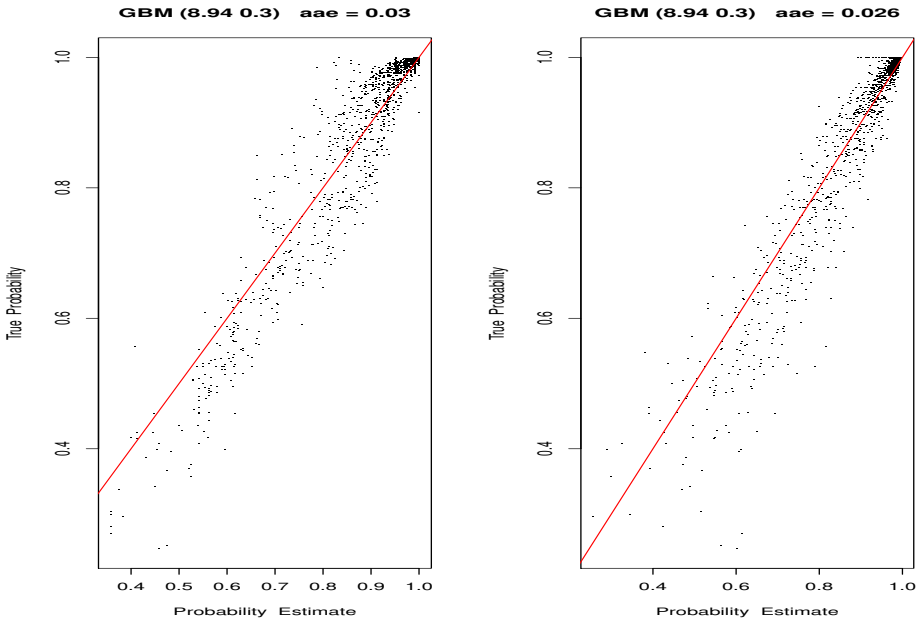

Figure 13: True probability versus estimated probability using gradient boosting (GBM) on perterbed ($q = 0.3$) non linear log-odds model data. Left: no class centering. Right: with class centering.

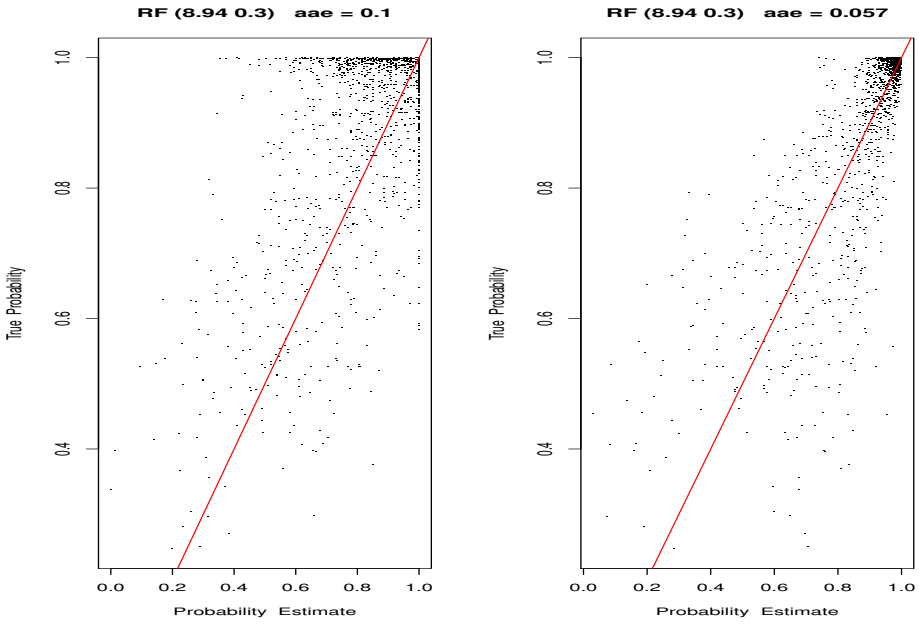

Figure 14: True probability versus estimated probability using random forest (RF) on perturbed ($q = 0.3$) non linear log-odds model data. Left: no class centering. Right: with class centering.

7 Remarks

There are many avenues of additional research identified in this paper that would need to be pursue in order to build a fully operational secure sales lift solution under the XMM framework. For example, appropriate techniques to privatize covariate data that would be used to build a joint propensity model and the trade-off between privacy and model accuracy. The paper also makes many assumptions about the availability of functionality that is yet to be built, for example, the common ID framework. However, it is hoped that enough of an outline of potential approaches has been presented in this paper, to demonstrate that a secure sales lift measurement solution is potentially feasible and within the range of statistical and computational techniques available today.

8 Acknowledgements

The authors would like to thank the many colleagues at Google for their helpful comments and discussions. In particular the authors would like to acknowledge Karn Seth and Mariana Raykova for all their input around secure computation techniques.

9 Appendix

9.1 Expectation Maximization

The expectation-maximization (EM) algorithm obtains maximum likelihood estimates of parameters in a likelihood model with incomplete data. The algorithm iterates between computing the expected complete-data likelihood at the current parameter estimates (the E step) and maximizing the expected complete-data likelihood to update the parameter estimates (the M step). In this section we explain how the EM algorithm proceeds for secure sales lift when the outcome model is a linear regression or logistic regression.

First, suppose a continuous outcome model: $Y_i = \alpha + \beta T_i + \gamma^\top Z_i + \epsilon_i$. The complete likelihood model for $(Y_i, T_i, \tilde{T}_i)$ given Z_i is determined by:

- the conditional distribution of Y_i given T_i and Z_i, specified by the regression model,
- the conditional distribution of $\tilde{T}_i$ given T_i, specified by the bit-flipping noise model,
- the conditional distribution of T_i given Z_i, for which we assume a logistic regression model: logit$(P(T_i = 1|Z_i)) = \eta^\top Z_i$.

Due to the built-in conditional independences of our problem setting, the joint distribution factors as $f(Y_i, T_i, \tilde{T}_i|Z_i) = f(Y_i|T_i, Z_i)f(\tilde{T}_i|T_i)f(T_i|Z_i)$, so the likelihood is fully specified. The second term, $f(\tilde{T}_i|T_i)$, can be ignored in parameter fitting because it does not involve any unknown parameters.

The complete-data log likelihood is

$$\begin{aligned}
&\ell(\alpha, \beta, \gamma, \sigma^2, \eta|T_i, \tilde{T}_i, Y_i, Z_i) = \\
&\quad -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{1}{2\sigma^2}(Y_i - \alpha - \beta T_i - \gamma^\top Z_i)^2 \\
&\quad + \sum_{i=1}^{n} T_i \log\left(\frac{\exp(\eta^\top Z_i)}{1+\exp(\eta^\top Z_i)}\right) + \sum_{i=1}^{n}(1-T_i)\log\left(\frac{1}{1+\exp(\eta^\top Z_i)}\right)
\end{aligned}$$

E step: Let θ denote the unknown parameters, $\theta = (\alpha, \beta, \gamma, \eta)$. The expectation of the above likelihood at current parameter estimates $\theta^{(t)}$, conditional on the observed data $(Y_i, \tilde{T}_i)$, is

$$\begin{aligned}
&E(\ell(\alpha, \beta, \gamma, \sigma^2, \eta|T_i, \tilde{T}_i, Y_i)|\tilde{T}_i, Y_i, \theta^{(t)}) = \\
&\quad -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\frac{1}{2\sigma^2}(Y_i - \alpha - \beta - \gamma^\top Z_i)^2(\mu_{1i}(\theta^{(t)})\tilde{T}_i + \mu_{0i}(\theta^{(t)})(1-\tilde{T}_i)) \\
&\quad - \sum_{i=1}^{n}\frac{1}{2\sigma^2}(Y_i - \alpha - \gamma^\top Z_i)^2((1-\mu_{1i}(\theta^{(t)}))\tilde{T}_i + (1-\mu_{0i}(\theta^{(t)}))(1-\tilde{T}_i)) \\
&\quad + \sum_{i=1}^{n}(\mu_{1i}(\theta^{(t)})\tilde{T}_i + \mu_{0i}(\theta^{(t)})(1-\tilde{T}_i))\log\left(\frac{\exp(\eta^\top Z_i)}{1+\exp(\eta^\top Z_i)}\right) \\
&\quad + \sum_{i=1}^{n}((1-\mu_{1i}(\theta^{(t)}))\tilde{T}_i + (1-\mu_{0i}(\theta^{(t)}))(1-\tilde{T}_i))\log\left(\frac{1}{1+\exp(\eta^\top Z_i)}\right),
\end{aligned}$$

where

$$\mu_{1i}(\theta) = \frac{p \cdot N(Y_i|\alpha+\beta+\gamma^\top Z_i, \sigma^2) \cdot \text{expit}(\eta^\top Z_i)}{p \cdot N(Y_i|\alpha+\beta+\gamma^\top Z_i, \sigma^2) \cdot \text{expit}(\eta^\top Z_i) + (1-p) \cdot N(Y_i|\alpha+\gamma^\top Z_i, \sigma^2) \cdot (1-\text{expit}(\eta^\top Z_i))},$$
$$\mu_{0i}(\theta) = \frac{(1-p) \cdot N(Y_i|\alpha+\beta+\gamma^\top Z_i, \sigma^2) \cdot \text{expit}(\eta^\top Z_i)}{(1-p) \cdot N(Y_i|\alpha+\beta+\gamma^\top Z_i, \sigma^2) \cdot \text{expit}(\eta^\top Z_i) + p \cdot N(Y_i|\alpha+\gamma^\top Z_i, \sigma^2) \cdot (1-\text{expit}(\eta^\top Z_i))}.$$

M step: Maximizing the expected complete-data log likelihood amounts to fitting two regressions. To obtain $\alpha^{(t+1)}, \beta^{(t+1)}), \gamma^{(t+1)}$, we perform a weighted least squares regression with $2n$ data points:

- The weights are $\mu_{1i}(\theta^{(t)})\tilde{T}_i + \mu_{0i}(\theta^{(t)})(1-\tilde{T}_i)$ for $i = 1, \ldots, n$, concatenated with $(1-\mu_{1i}(\theta^{(t)}))\tilde{T}_i + (1-\mu_{0i}(\theta^{(t)}))(1-\tilde{T}_i)$ for $i = 1, \ldots, n$.
- The covariates are an intercept of length $2n$, a binary covariate consisting of n ones followed by n zeroes, and two stacked copies of Z_i for $i = 1, \ldots, n$.
- The outcome consists of two stacked copies of Y_i for $i = 1, \ldots, n$.

To obtain $\eta^{(t+1)}$, we perform either a "pseudo" logistic regression with fractional outcomes $\mu_{1i}(\theta^{(t)})\tilde{T}_i + \mu_{0i}(\theta^{(t)})(1-\tilde{T}_i)$ and covariates Z_i, or equivalently, a weighted logistic regression with $2n$ data points, where the weights are the same as above, the outcome consists of n ones concatenated with n zeroes, and the covariates are Z_i stacked on itself.

Second, suppose a binary outcome model: $Y_i \sim \text{Bernoulli}(p_i)$ where $\text{logit}(p_i) = \alpha + \beta T_i + \gamma^\top Z_i$.

9.2 Regression Calibration Details

Regression calibration is one of the popular statistical methods to estimate a regression model which involves measurement error in its covariates. It is known that the regression calibration produces approximately unbiased estimates of regression coefficients. Measurement error in our application corresponds to DP noise that is added to exposure bits in TDP outcome model.

We present the key components of applying regression calibration to secure sales lift problems by focusing on the estimation procedure when TDP outcome model is linear. For other types of GLM, similar procedure should still apply. The underlying outcome model on the i-th unit without noisy exposure bits is that:

$$\mathrm{E}[Y_i] = \alpha + \beta T_i + \gamma Z_i + \lambda Z_i \times T_i, \tag{8}$$

recalling that T_i is real exposure indicator and Z_i represents TDP covariates. The goal is to estimate all parameters α, β, γ, given that real exposure bits T_i cannot be observed on the TDP side. Take the expectation of Y_i conditioning on Z_i and $\widetilde{T}_i$:

$$\begin{aligned}\mathrm{E}[Y_i \mid Z_i, \widetilde{T}_i] = \alpha + \beta \Pr[T_i = 1 \mid Z_i, \widetilde{T}_i] + \gamma Z_i + \\ \lambda Z_i \times \Pr[T_i = 1 \mid Z_i, \widetilde{T}_i].\end{aligned} \tag{9}$$

This model above is called the calibrated outcome model. The regression calibration estimates are the least-square estimates that solves calibrated model. Notice that the key difference between equation 8 and 9 is that

the real exposure bits T_i is substituted by $\Pr[T_i = 1 \mid Z_i, \widetilde{T}_i]$ which can be viewed as an "imputed" value of T_i by virtue of the correlation between T_i and the tuple $(Z_i, \widetilde{T}_i)$. We will discuss how to estimate this quantity in the next section. In fact, for a variety of secure sales lift applications, we can utilize the following two-step procedure to estimate TDP outcome model:

1. Build a prediction model that predicts T with tuple $(Z_i, \widetilde{T}_i)$. Find the quantity $\Pr[T_i = 1 \mid Z_i, \widetilde{T}_i]$.
2. In outcome model without exposure bits (e.g., equation 8), replace T_i with the quantity $\Pr[T_i = 1 \mid Z_i, \widetilde{T}_i]$. In calibrated model (e.g., equation 9), proceed with the common estimation procedure in a GLM (e.g., the least-square estimates for linear regression, maximum-likelihood estimation for logistic regression, and etc.).

9.2.1 Estimation of $\Pr(T_i = 1 \mid Z_i, \widetilde{T}_i)$

We assume that TDP covariates Z_i are sufficiently informative to predict the real exposures T_i. We also assume publisher propensity score PS_i as one dimension within Z_i if the publisher sends propensity scores to TDP. This is because publisher propensity scores are usually strongly correlated with T_i. Under the most extreme case where Z_i is strongly linked with T_i, we are capable of imputing the values of exact exposures almost perfectly. In our implementation, we model the relationship between Z_i and T_i through a logistic regression in which $\Pr(T_i = 1 \mid Z_i) = \text{expit}(\eta_0 + \eta_1 Z_i)$. For convenience, let us pretend to include an additional constant 1 inside the covariates Z_i, so that we can simplify this formula: $\Pr(T_i = 1 \mid Z_i) = \text{expit}(\eta Z_i)$.

When the probability of flipping exposure bits is q, the "sensitivity" probability $\Pr(\widetilde{T}_i = 1 \mid T_i = 1) = 1 - q$; and the "specificity" probability $\Pr(\widetilde{T}_i = 0 \mid T_i = 0) = 1 - q$. Therefore, we should have $\Pr(\widetilde{T}_i = 1 \mid Z_i) = \Pr(\widetilde{T}_i = 1 \mid T_i = 1, Z_i) \Pr(T_i = 1 \mid Z_i) + \Pr(\widetilde{T}_i = 1 \mid T_i = 0, Z_i) \Pr(T_i = 0 \mid Z_i)$ by integrating out T_i. Moreover, the bit-flipping DP-noise is independent from covariates Z_i. This further yields:

$$\Pr(\widetilde{T}_i = 1 \mid Z_i) = (1 - q)\text{expit}(\eta Z_i) + q(1 - \text{expit}(\eta Z_i)). \tag{10}$$

We can now find out the maximum-likelihood estimators . The log-likelihood L is:

$$\begin{aligned} &\sum_{i=1}^{n} \widetilde{T}_i \log((1 - q)\text{expit}(\eta Z_i) + q(1 - \text{expit}(\eta Z_i))) + \\ &\quad (1 - \widetilde{T}_i) \log(q \times \text{expit}(\eta Z_i) + (1 - q)(1 - \text{expit}(\eta Z_i))). \end{aligned} \tag{11}$$

In addition, $\partial L / \partial \eta$, the first order derivative w.r.t η equals:

$$\sum_{i=1}^{n} Z_i (\text{expit}(-\eta Z_i) - \frac{q\widetilde{T}_i}{q + (1 - q)\text{expit}(\eta Z_i)} - \frac{(1 - q)(1 - \widetilde{T}_i)}{1 - q + q \times \text{expit}(\eta Z_i)}). \tag{12}$$

We will obtain MLE $\hat{\eta}$ by finding the root to equation $\partial L / \partial \eta = 0$. In our implementation, we utilize an optimization approach that relies on both first and second order derivatives.

According to the Bayes' rule, after finding MLE, we will have the following estimators:

$$\begin{aligned}\widehat{\Pr}(T_i = 1 \mid Z_i, \widetilde{T}_i = 1) &= \frac{(1-q)\text{expit}(\hat{\eta} Z_i)}{(1-q)\text{expit}(\hat{\eta} Z_i) + q(1-\text{expit}(\hat{\eta} Z_i))} \\ \widehat{\Pr}(T_i = 1 \mid Z_i, \widetilde{T}_i = 0) &= \frac{q \times \text{expit}(\hat{\eta} Z_i)}{q \times \text{expit}(\hat{\eta} Z_i) + (1-q)(1-\text{expit}(\hat{\eta} Z_i))}.\end{aligned} \tag{13}$$

9.2.2 Validity of the regression calibration approach

In this section, we briefly introduce why this simple correction procedure yields approximately unbiased results when we estimate an outcome model with noisy exposures.

First of all, the regression calibration estimator can be viewed as an MLE to an approximated likelihood function. Suppose that the real likelihood contribution for the i-th unit can be written as:

$$\Pi(Y_i, \widetilde{T}_i \mid Z_i; \theta, q, \eta) = \int \pi(Y_i \mid T_i, Z_i; \theta) \pi(\widetilde{T}_i \mid T_i; q) \pi(T_i \mid Z_i; \eta) dT_i. \tag{14}$$

We use notation to represent all parameters $(\alpha, \beta, \gamma, \lambda)$ in outcome model 8.

In the first step of regression calibration, we rewrite $\pi(\widetilde{T}_i \mid T_i; q)\pi(T_i \mid Z_i; \eta)$ as $\pi(T_i \mid Z_i, \widetilde{T}_i; q, \eta) \times \pi(\widetilde{T}_i \mid Z_i; q, \eta)$. Then the likelihood becomes

$$\begin{aligned}&\int \pi(Y_i \mid T_i, Z_i; \theta) \pi(T_i \mid Z_i, \widetilde{T}_i; q, \eta) \pi(\widetilde{T}_i \mid Z_i; q, \eta) \\ &= \pi(\widetilde{T}_i \mid Z_i; q, \eta) \times \int \pi(Y_i \mid T_i, Z_i; \theta) \pi(T_i \mid Z_i, \widetilde{T}_i; q, \eta) dT_i,\end{aligned} \tag{15}$$

in which the first term in the decomposition $\pi(\widetilde{T}_i \mid Z_i; q, \eta)$ no longer involves integral over T_i. Summing up this term over i, we find $\hat{\eta}$, the MLE to parameter η, based on the likelihood $\sum_i \pi(\widetilde{T}_i \mid Z_i; q, \eta)$.

In the second step of regression calibration, we want to find MLE to θ based on the likelihood $\int \pi(Y_i \mid T_i, Z_i; \theta)\pi(T_i \mid Z_i, T_i; q, \eta)dTi$. This means the integral of $\pi(Y_i \mid T_i, Z_i; \theta)$ based on posterior density $\pi(T_i \mid Z_i, \widetilde{T}_i; q, \eta)$. This quantity can be approximated by $\pi(Y_i \mid T_i, Z_i; \theta)$ in which we plug in $\hat{T}_i$, the posterior mean based on the posterior density. The accuracy of the regression calibration estimator relies on a close approximation between the quantity $\sum_i \pi(Y_i \mid \hat{T}_i, Z_i; \theta)$ and $\sum_i \int \pi(Y_i \mid T_i, Z_i; \theta)\pi(T_i \mid Z_i, \widetilde{T}_i; q, \eta)dT_i$. Under various settings of GLM, (2) summarized the impact of this approximation and concluded that regression calibration is most useful when:

- The true effects of the covariates measured with error (in our case, $\widetilde{T}$) are moderate;
- Measurement error variance (in our case, magnitude of DP noise q) is small.

In fact, in many cases this approximation works well. The MLE to θ will then follow the standard analysis after we substitute these noisy T with the posterior mean of T.

10 Simulation Appendix

10.1 Simulated Data Generating Process

Each simulated dataset consists of 100,000 (0.1 million) users. Data attributes of every individual user can be represented as a tuple of the following elements:

- X: user's online activity data that function as the predictors in publisher propensity model;
- T: real exposure indicator (1 indicates an exposed user and 0 indicates an unexposed user) that publisher observes;
- $\widetilde{T}$: noisy exposure indicator which differs from T by adding DP noise;
- PS: propensity scores that publisher sends to third-party advertisers;
- Z: third-party measurements on users which serve as the covariates in outcome model;
- Y: the sales outcome (e.g., dollar spending of each user, conversion rate, and etc.) variable in outcome model.

Among all, publisher possesses $(X, T, \widetilde{T}, PS)$ and decides the manner to send these data to third-party advertiser. Therefore, we call these data publisher data. The process of generating publisher data is discussed in section 10.1.1 Similarly, the remaining items are called third-party data. The corresponding data generating process is discussed in section 10.1.2.

10.1.1 Publisher data

Online activity data on the i-th unit, X_i, are drawn from an independent 3 dimensional Gaussian distribution with zero mean and heterogeneous variances:

$$X_i = (X_{i,1}, X_{i,2}, X_{i,3}) \sim \mathrm{N}((0,0,0),\ \mathrm{diag}(1,1,3)). \tag{16}$$

The probability of a given user receiving ads exposure, $\Pr(T_i = 1)$, is generated by a logistic model:

$$\Pr(T_i = 1) = \mathrm{sigmoid}(c_0 + c_1 X_{i,1} + c_2 X_{i,2} + c_3 X_{i,3}), \tag{17}$$

where $\mathrm{sigmoid}(x) = (1+e^{-x})^{-1}$. To ensure that our simulated data mimic real applications, we deliberately pick $c_0 = \log(2/23), c_1 = 0.5, c_2 = 0.3$, and $c_3 = -0.2$. Under this scenario, approximately 10% users are assigned to the exposed group while the remaining 90% are unexposed users.

Noisy exposure indicator $\widetilde{T}_i$ is generated from real exposure T_i by randomly flipping the exposure status with probability q, which is equivalent to the following transition rule:

$$\begin{pmatrix} \Pr(\widetilde{T}_i = 0 \mid T_i = 0) & \Pr(\widetilde{T}_i = 1 \mid T_i = 0) \\ \Pr(\widetilde{T}_i = 0 \mid T_i = 1) & \Pr(\widetilde{T}_i = 1 \mid T_i = 1) \end{pmatrix} = \begin{pmatrix} 1-q & q \\ q & 1-q \end{pmatrix}. \tag{18}$$

In this equation, probability q quantifies the magnitude of DP noise added to the exposure indicator. In simulation studies, we select 8 q values: $5\%, 10\%, 15\%, 20\%, 25\%, 30\%, 35\%$ and 40%. Although the scenarios that correspond to $q = 35\%$ ($\epsilon = 0.62$) and 40% ($\epsilon = 0.41$) could be practically infeasible, we include them to examine the robustness of candidate methods.

Propensity score is the last component in publisher data. We consider two variants of propensity score. First, the exact score is the predicted value of $\Pr(T_i = 1)$, after fitting a logistic regression model with T_i and X_i. The other variant, private score, corresponds to the predicted value of $\Pr(\widetilde{T}_i = 1)$ after fitting a logistic regression model with $\widetilde{T}_i$ and X_i. Based on empirical study, we observed a significant amount of deviation between the two variants, but also a positive correlation between them in general.

10.1.2 Third-party data

In outcome model, covariates on the i-th unit $Z_i = (Z_{i,1}, Z_{i,2}, Z_{i,3})$ are drawn separately from an independent 3 dimensional Gaussian distribution with unit variance, depending on whether this user receives treatment or not. That is:

$$\begin{aligned} Z_i \mid T_i = 0 &\sim \mathrm{N}((0,0,0),\ \mathrm{diag}(1,1,1)) \\ Z_i \mid T_i = 1 &\sim \mathrm{N}((0.2,0.1,-0.1),\ \mathrm{diag}(1,1,1)) \end{aligned} \tag{19}$$

Suppose that $Y_i(1)$ and $Y_i(0)$ denote outcome value under the exposed and unexposed condition respectively. Although in reality a variety of outcome models could be useful in estimating sales lift, we focus on two particular types of outcome models: (i) linear model in which outcome is a continuous variable, such as consumer spending; and (ii) logistic model in which outcome is a binary variable, such as whether or not customer purchases a specific item.

In linear outcome model, the conditional expected exposed outcome

$$\mathrm{E}[Y_i(1) \mid Z_i] = 11 + 2Z_{i,1} + Z_{i,2}; \tag{20}$$

and the conditional expected unexposed outcome

$$\mathrm{E}[Y_i(0) \mid Z_i] = 10 + Z_{i,1} + Z_{i,2} + Z_{i,3}. \tag{21}$$

It is easy to see that the average exposed outcome is 11.5 while the average unexposed outcome is 10. Meanwhile, the treatment effect at a specific user $\tau(Z_i) = \mathrm{E}[Y_i(1) \mid Z_i] - \mathrm{E}[Y_i(0) \mid Z_i]$. Following the definition of ATE, the expected sales lift ATE can be written as

$$\begin{aligned} &\Pr(T_i = 1)\mathrm{E}[\tau(Z_i) \mid T_i = 1] + \Pr(T_i = 0)\mathrm{E}[\tau(Z_i) \mid T_i = 0] \\ &= 0.1 \times (1 + \mathrm{E}[Z_{i,1} \mid T_i = 1] - \mathrm{E}[Z_{i,3} \mid T_i = 1]) + 0.9 \times 1 = 1.03 \end{aligned} \tag{22}$$

Finally, the observed outcome variable Y_i is generated by the process:

$$Y_i \mid T_i = t, Z_i \sim \mathrm{N}(\mathrm{E}[Y_i(t) \mid Z_i], 5^2), \quad \text{for } t = 0, 1, \tag{23}$$

where the definitions of conditional exposed and unexposed outcome are in equations 20 and 21.

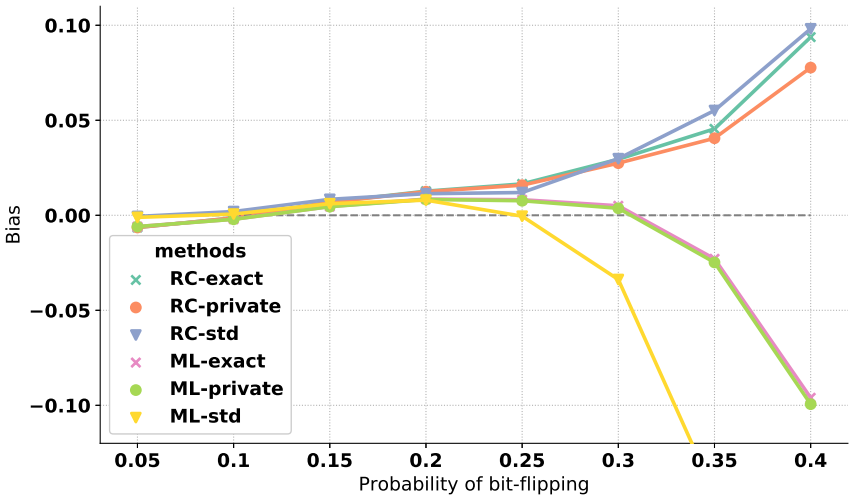

Figure 15: Biases of estimator for β (the model coefficient that corresponds to exposure bits) under linear outcome scenario.

In logistic outcome model, for convenience, let us assume that the outcome variable means whether to purchase an item. Then the conditional probability of making a purchase for an exposed user

$$\Pr(Y_i(1) = 1 \mid Z_i) = \text{sigmoid}(-2.3 + 0.4Z_{i,1} + 0.2Z_{i,2}); \tag{24}$$

and the conditional probability of making a purchase for an unexposed user

$$\Pr(Y_i(0) = 1 \mid Z_i) = \text{sigmoid}(-2.5 + 0.2Z_{i,1} + 0.2Z_{i,2} + 0.2Z_{i,3}). \tag{25}$$

Treatment effect at a specific user $\tau(Z_i) = \Pr(Y_i(1) = 1 \mid Z_i) - \Pr(Y_i(0) = 1 \mid Z_i)$. Following the same arguments in equation 22, the sales lift ATE based on this logistic model is approximately 2%.

10.2 Additional Simulation Results

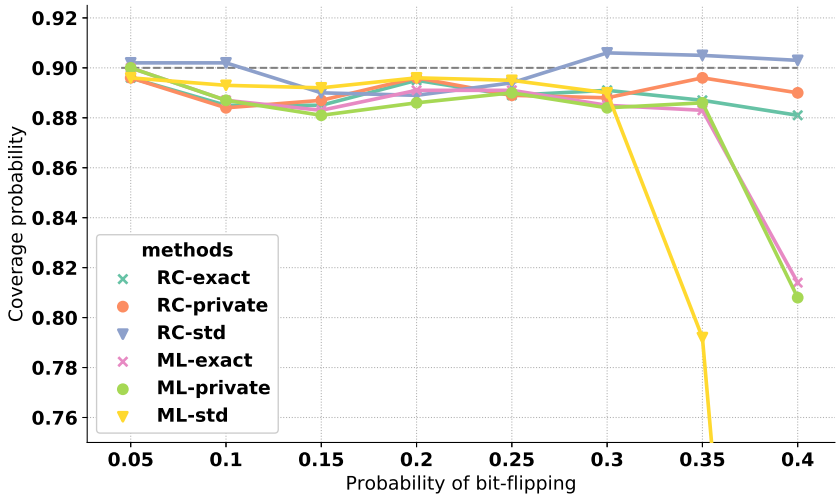

Figure 16: Coverage probabilities of confidence interval for β (the model coefficient that corresponds to exposure bits) under linear outcome scenario.

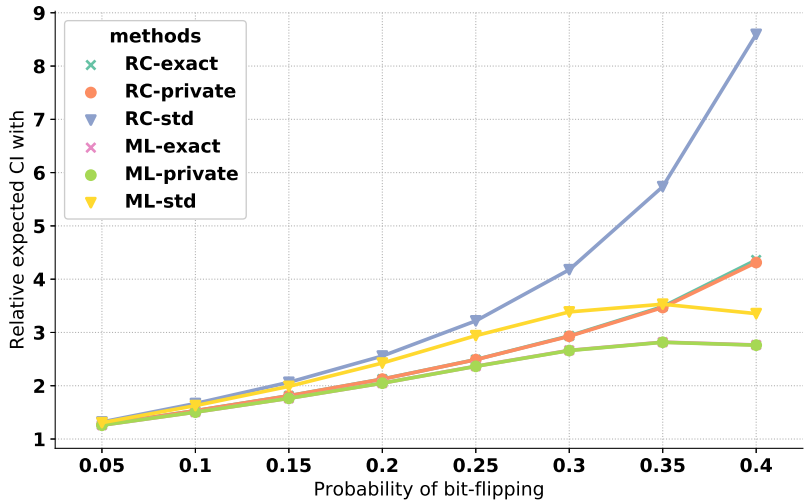

Figure 17: Relative widths of confidence interval for β (the model coefficient that corresponds to exposure bits) under linear outcome scenario. The relative width is defined as the average interval width when estimating outcome model with noisy exposures $\widetilde{T}$ divided by the average interval width when estimating outcome model without DP-noise.

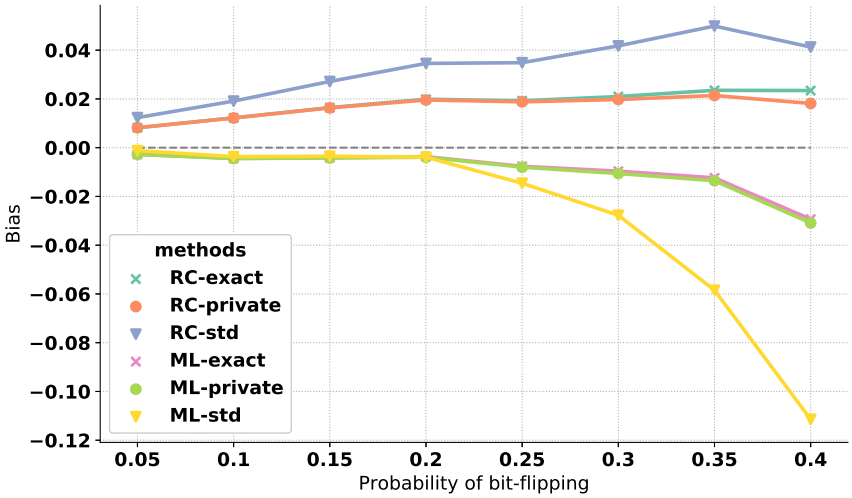

Figure 18: Biases of estimator for β (the model coefficient that corresponds to exposure bits) under the logistic outcome scenario.

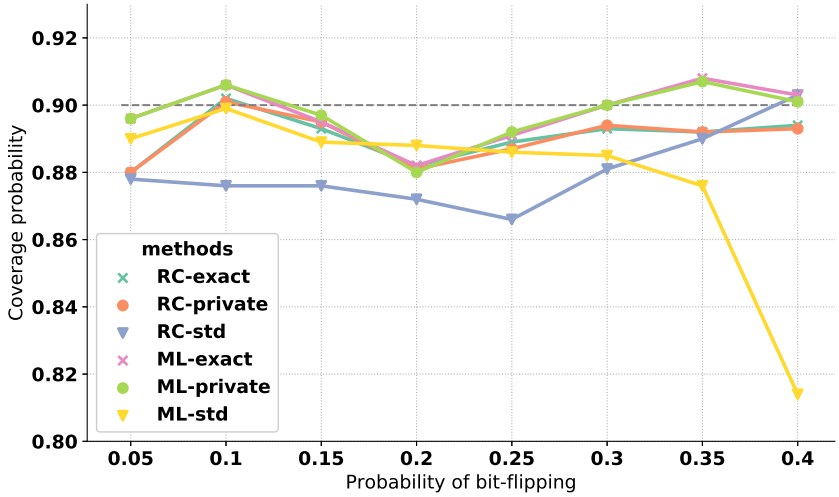

Figure 19: Coverage probabilities of confidence interval for $beta$ (the model coefficient that corresponds to exposure bits) under logistic outcome scenario.

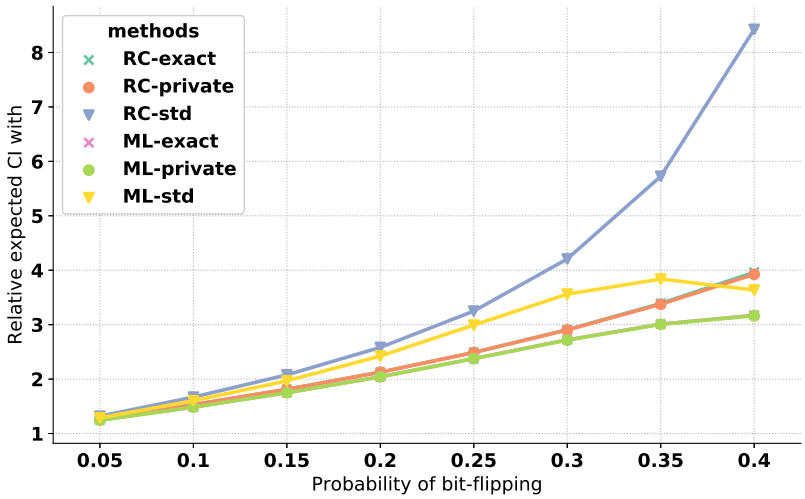

Figure 20: Relative widths of confidence interval for β (the model coefficient that corresponds to exposure bits) under logistic outcome scenario. The relative width is defined as the average interval width interval when estimating outcome model with noisy exposures $\widetilde{T}$ divided by the average interval width when estimating outcome model without DP-noise.

References

[1] Acar, A., Aksu, H., Uluagac, A. S., and Conti, M. A survey on homomorphic encryption schemes: Theory and implementation, 2017. https://arxiv.org/abs/1704.03578.

[2] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. *Measurement error in nonlinear models: a modern perspective*. CRC press, 2006.

[3] Carroll, R. J., and Stefanski, L. A. Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association 85*, 411 (1990), 652–663.

[4] Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*, 1 (1977), 1–22.

[5] Hankin, M., Chan, D., and Perry, M. A comparison of causal inference methods for estimating sales lift. Tech. rep., Google Inc., 2020.

[6] Ion, M., Kreuter, B., Nergiz, A. E., Patel, S., Raykova, M., Saxena, S., Seth, K., Shanahan, D., and Yung, M. On deploying secure computing: Private intersection-sum-with-cardinality. Cryptology ePrint Archive, Report 2019/723, 2019. https://eprint.iacr.org/2019/723.

[7] Movahedi, M., Case, B. M., Knox, A., Li, L., Li, Y. P., Saravanan, S., Sengupta, S., and Taubeneck, E. Private randomized controlled trials: A protocol for industry scale deployment, 2021.

[8] Peng, J., Schneider, S., Knightbrook, J., Book, L., Ma, S., Huang, X., Daub, M., Yang, Y., Frye, J., Wright, C., Skvortsov, E., Liu, Y., and Koehler, J. Privacy-centric cross-publisher reach and frequency estimation via vector of counts. Tech. rep., Google Inc., 2021.

[9] Raymond J. Carroll, Clifford H. Spiegelman, G. K. K. L. K. T. B., and Abbott, R. D. On errors-in-variables for binary regression models. *Biometrika 71*, 1 (04 1984), 19–25.

[10] Schafer, D. W., and Purdy, K. G. Likelihood analysis for errors-in-variables regression with replicate measurements. *Biometrika 83*, 4 (1996), 813–824.

[11] Stefanski, L. A., and Carroll, R. J. Covariate measurement error in logistic regression. *The Annals of Statistics 13*, 4 (1985), 1335 – 1351.

[12] Google Inc. Helping organizations do more without collecting more data. https://security.googleblog.com/2019/06/helping-organizations-do-more-without-collecting-more-data.html.

[13] World Federation Of Advertisers. WFA CMM framework. https://wfanet.org/l/library/download/urn:uuid:ea16e189-7592-416e-be8e-063bd674de9e/wfa+industry+framework+for+xmm.pdf.

[14] World Federation Of Advertisers. World federation of advertisers. https://wfanet.org/.

[15] Vaver, J., and Koehler, J. Measuring ad effectiveness using geo experiments. Tech. rep., Google Inc., 2011.