

Efficient Large-Scale Internet Media Selection  
Optimization for Online Display Advertising  
Forthcoming at *Journal of Marketing Research*

COURTNEY PAULSON<sup>1</sup>, LAN LUO<sup>2</sup>, AND GARETH M. JAMES<sup>3</sup>

June 15, 2018

<sup>1</sup>Robert H. Smith School of Business, University of Maryland.

<sup>2</sup>Marshall School of Business, University of Southern California.

<sup>3</sup>Marshall School of Business, University of Southern California.

## *Abstract*

In today's digital market, the number of websites available for advertising has ballooned into the millions. Consequently, firms often turn to ad agencies and demand-side platforms (DSPs) to decide how to allocate their Internet display advertising budgets. Nevertheless, most extant DSP algorithms are rule-based and strictly proprietary. This paper is among the first efforts in marketing to develop a non-proprietary algorithm for optimal budget allocation of Internet display ads within the context of programmatic advertising. Unlike many DSP algorithms that treat each ad impression independently, our method explicitly accounts for viewership correlations across websites. Consequently, campaign managers can make optimal bidding decisions over the entire set of advertising opportunities. Therefore, they can avoid overbidding for impressions from high cost publishers, unless such sites reach an otherwise unreachable audience. Our method can also be used as a budget-setting tool, because it readily provides optimal bidding guidelines for a range of campaign budgets. Finally, we can accommodate several practical considerations including consumer targeting, target frequency of ad exposure, and mandatory media coverage to matched content websites.

*Keywords:* Online Display Advertising; Internet Media Selection; Programmatic Advertising; Optimal Internet Advertising Budget Allocation; Real time Bidding; Constrained Convex Optimization; Coordinate Descent Gradient Search

With the increased role of Internet use in the U.S. economy, Internet advertising is becoming vital for company survival. In 2012, U.S. digital advertising spending (including display, search, and video advertising) totaled 37 billion dollars, of which Internet display advertising accounted for 40%. This percentage is expected to continue to grow, outpacing paid search ad spending (eMarketer, 2012). Such an increasing trend in Internet display advertising is related to a wide range of benefits offered by this advertising format, including building awareness and recognition, forming attitudes, and generating direct responses such as website visits and downstream purchases (Danaher et al., 2010; Hoban and Bucklin, 2015; Manchanda et al., 2006).

Nevertheless, firms face considerable challenges in deciding where to place their online display ads. By 2012, the number of websites available for advertising had reached the millions (HBR, 2015). As the number of ad buying opportunities proliferates, advertising automation has become inevitable. Indeed, many believe that the future of online display advertising will be largely programmatic (MarketingLand, 2013; HBR, 2015).

Despite the increasing popularity of Internet display ads and the rise of programmatic advertising, very few marketing papers have examined optimal budget allocation of Internet displays ads within the context of programmatic advertising. To date, Danaher’s Sarmanov-based model (Danaher et al., 2010) was among the first and most successful attempts in the nonprogrammatic setting. This method has proven to work well for budget allocation in settings on the order of 10 websites, under the assumption that the cost to advertise at each website (i.e., CPM) is fixed and known. Danaher and Smith (2011) also proposed a Copula-based approach to efficiently model page view distributions of 45 websites. Nevertheless, under either the Sarmanov- or Copula-based framework, the objective function is highly nonconvex. Therefore, the consideration of each additional website increases the complexity of the budget allocation optimization exponentially, an inevitable limitation given the huge volume of Internet websites on which firms could potentially choose to advertise over.

This paper is among the first efforts in marketing to develop a non-proprietary algorithm for optimal budget allocation of Internet display ads within the context of programmatic advertising. Our method allows campaign managers to develop specific bidding guidelines

over a range of budgets. Under our approach, campaign managers can maximize several key performance indicators (KPIs), including reach (the probability that an Internet user views the ad at least once during a campaign), frequency (the average number of exposures among those reached), Gross Rating Points (frequency multiplied by reach), and effective frequency or frequency capping (the proportion of consumers who view the ad within a range of frequencies).

Given that the programmatic market with RTB (real time bidding) currently comprises about half of the U.S. display ad market, we mainly focus our discussion on the applicability of our method in the context of programmatic ad buying with RTB. Nevertheless, our approach is also directly applicable to nonprogrammatic and programmatic direct display ad markets.<sup>1</sup>

Under RTB, when an ad space becomes available, the ad agency or DSP has only milliseconds to place a bid (VentureBeat, 2012). Because bidding decisions must be made in real-time, most DSP algorithms rely on predefined bidding guidelines to decide whether and how much to bid for available impressions. We follow a similar approach by proposing a method with which campaign managers can calibrate a set of prescriptive bidding guidelines prior to placing bids. Based on results from our optimization, campaign managers can then decide on which websites to place bids, as well as the price at which they should bid in a real-time setting.

In practice, most extant DSP algorithms are strictly proprietary, are often rule-based and consider each ad impression singularly (G2Crowd, 2016; Adelphic, 2015; Wang et al., 2016). Given the massive number of available impressions and varying information contained in any given impression, it is challenging to develop an algorithm that specifically accounts for potential correlations across all ad impressions. As a result, existing bidding algorithms often follow some sort of simple rubric for making the bidding decision, rather than optimizing over the entire set of advertising opportunities (Adelphic, 2015).

---

<sup>1</sup>The U.S. display ad market is expected to be 17% nonprogrammatic, 31% programmatic direct, and 52% programmatic with RTB in 2017 (eMarketer, 2013). We outline in Web Appendix A how our method can be used for nonprogrammatic and programmatic direct display ad markets. For completeness, we also provide a comparison between our method and the smaller-scale optimization proposed by Danaher (Danaher, 2007; Danaher et al., 2010) in this appendix.

In contrast, our approach explicitly takes into account viewership correlations across all websites under consideration. Consequently, campaign managers can avoid overbidding for impressions from high cost publishers, unless such sites reach an otherwise unreachable audience. For example, the 2011 comScore Media Metrix data show that there is up to a 95% correlation in the viewership of Businessweek.com and Reuters.com. In such cases, firms can benefit greatly from strategically placing more bids for impressions at the more cost-effective website of the two, since both largely attract the same viewers. When a large number of websites is under consideration, this advantage will be considerably amplified. Consequently, we propose a publisher-centric method as a viable alternative to complement existing rule-based algorithms used by most DSPs.

One reason for the difficulty of optimizing budget allocations over a large number of websites is that the problem of choosing the subset of websites on which to bid is generally NP-hard. In a setting involving ad inventory from  $p$  potential websites, each of the  $2^p$  possible website subsets must be considered, leading to a computationally infeasible problem. While no feasible solution to this problem currently exists in the digital ad marketplace, the Internet media selection problem is analogous to the classic linear regression variable selection problem involving selecting a subset from a large number of independent variables.

A common solution adopted in the statistical literature involves optimizing a constrained convex loss function, a relaxed version of the NP-hard variable selection problem. Some recent papers include the Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), the elastic net (Zou and Hastie, 2005), the adaptive Lasso (Zou, 2006), CAP (Zhao et al., 2009), the Dantzig selector (Candes and Tao, 2007), the relaxed Lasso (Meinshausen, 2007), and VISA (Radchenko and James, 2008). We extend this concept to the website selection setting by developing a constrained convex optimization, which can be effectively applied to settings with a very large number of websites. Our method is related to the Lasso formulation (Tibshirani, 1996) but diverges in that our optimization does not involve a quadratic loss function. We also leverage the bid landscape literature (Cui et al., 2011; Iyer et al., 2011; Wang et al., 2016; AdWords, 2016) by incorporating the probability of winning a bid when deciding how much to bid for impressions from each website.

Our empirical results demonstrate that our method can effectively develop prescriptive bidding guidelines in online display ad campaigns involving a large number of websites. In addition, because our optimization runs efficiently over a range of budgets, campaign managers can use our method to determine an optimal campaign budget. We also show that the proposed method is flexible enough to accommodate common Internet display advertising considerations, such as consumer targeting, target frequency of ad exposure, and mandatory media coverage to matched content websites. Hence, we believe that the proposed optimization method will be of considerable value in Internet display ad campaigns.

The remainder of the paper is structured as follows. First, we discuss the Internet display ad market and the real-world applicability of the proposed method. Second, we describe details of our optimization function and a coordinate descent algorithm to solve the optimization. Third, we provide a conceptual demonstration of our approach using two simulated illustrations. We then present two case studies (a McDonald’s McRib Advertising Campaign and a Norwegian Cruise Lines Wave Season Advertising Campaign) using 2011 comScore Media Metrix data to demonstrate the proposed method in a real-world setting. Lastly, we conclude the paper with a summary of our findings, contributions, and avenues for future work.

## INSTITUTIONAL SETTING AND MANAGERIAL USAGE

### *Programmatic Advertising with RTB for Internet Display Ads*

A common goal for an Internet display ad campaign is to allocate the firm’s budget to maximize some KPI metric. In today’s digital ad marketplace, pay-per-impression (CPM buying) is the most common form of purchase in display ad campaigns, with pay-per-click (CPC buying) being the most popular for search ad campaigns (PromiseMedia, 2015; iThinkMedia, 2016). While a small percentage of advertisers are still pushing for pay-per-acquisition (CPA buying) campaigns, this advertising model is far less popular in practice because it places all risk and responsibility on publishers, while putting none on advertisers (PromiseMedia, 2015).

The primary focus of this paper is on Internet display ads, so we emphasize pay-per-impression campaigns where the KPIs are gauged by exposure-based metrics such as reach, frequency, and Gross Rating Points (GRPs).<sup>2</sup> In this setting, our approach works best for brand-oriented campaigns that emphasize factors such as building awareness and recognition or forming attitudes, which in turn may also generate downstream purchases. Additionally, we focus on the scenario of private marketplace auctions (the most popular form of programmatic auction) where only select websites and advertisers are allowed to engage in the auction (DNB, 2015).<sup>3</sup> Within this context, campaign managers need to determine whether and how much to bid for ad impressions from all websites under consideration.

In our setting, one practical consideration involves ascertaining the likelihood of winning the impression for any particular bid at a given website. In practice, such relationships are often revealed in a process called bid landscaping (AdWords, 2016; Iyer et al., 2011). By varying bids and observing the corresponding changes in the probability of winning that bid at a given website, advertisers can make more informed decisions about their bidding choices.<sup>4</sup> Some advertising agencies even supply “bid simulators”, giving advertisers an interactive tool to see how changing their bids affects their returns (AdWords, 2016). Further, if campaign managers have certain targeted consumer groups, bid landscapes can be generated by bidding for ad impressions from those target groups to create specific demographic-based curves.

Based on this bid landscape procedure, campaign managers can plot *cost curves* that show the relationship between the bid price for an ad impression and the likelihood of winning that bid. Figure 1 provides an example cost curve which shows the DSP has a 50% probability to win the auction when bidding \$3, 62.5% when bidding \$4, but only 35.1% with a bid of \$2. The expected number of winning bids is therefore the product of the probability of winning

---

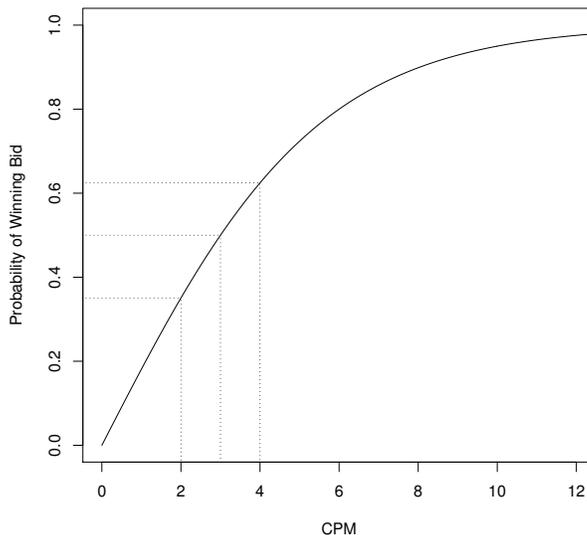
<sup>2</sup>In the Methodology Section we also discuss how our optimization could be modified to maximize non-exposure-based KPIs such as clicks or purchases if such data were available.

<sup>3</sup>A key advantage of private marketplace auctions is that advertisers don’t have to worry about their ads popping up on less reputable sites, and publishers can also assert some control over the quality of advertisers who can place ads on their websites.

<sup>4</sup>According to our conversations with industry experts, these bid landscapes are often generated during an initial “burn-in” advertising period prior to an ad campaign. During this period, a wide range of bids are placed at various websites in order to track how often bids are accepted at those websites at a variety of price points.

the bid and the expected number of page views at this website. In the example above, with a bid of \$3, the DSP can expect to win 5,000 bids if 10,000 impressions were available.

**Figure 1: Example Cost Curve**



In practice, bid landscapes might fluctuate over time, possibly due to short-term changes in supply and demand of desired ad impressions or changes in website visitation. As a result, campaign managers often monitor the DSP’s performance on an interim basis (e.g. daily) so that they can recalibrate the bidding guidelines if needed.

### *How Our Method Can be Used in Practice*

Here we discuss the real-world applicability of our method for RTB and how it can be used as a budget-setting tool for campaign managers. As discussed earlier, a DSP only has milliseconds to determine whether and how much to bid under RTB. Therefore, we suggest, like many extant DSP algorithms, the campaign manager runs the proposed method prior to the ad campaign in order to determine optimal bidding prices for each website (including placing zero bids on certain websites). As discussed above, cost curves may fluctuate over time but this can be easily handled by regularly re-running the optimization, as is common for other DSP algorithms.

Because our optimization method is computationally efficient, running the algorithm over a range of budgets on a daily (or even hourly) basis is highly feasible. Consider a moderately-sized setting involving 500 websites. Running the method for 100 budgets with a relatively small maximum budget of \$20,000 takes approximately 0.11 seconds per budget calculation. The average per budget time increases to 0.49 seconds with a maximum budget of \$250,000, and 2.56 seconds for a maximum budget of \$1,000,000.<sup>5</sup> Thus, for most realistic budget allocations, campaign managers can, if desired, regularly recalibrate the optimization. In this regard, our method is similar in spirit to sequential learning algorithms whose purpose is to combine new incoming data with past data to improve optimization going forward (Lewis and Gale, 1994; Yingwei et al., 1997). Campaign managers can update cost curves based on recently accepted and rejected bids at each website, which in turn updates the bidding guidelines once the method is rerun. Additionally, because our optimization relies on an iterative approach to find the optimal solution, using previous estimates as “warm starts” can greatly improve computation time when cost curves and other parameters change.

It is also worth noting that, although for illustrative purposes we currently demonstrate our method using shifted logistic cost curves as seen in Figure 1, the proposed approach is well-behaved and computationally efficient for any differentiable concave cost curves (see proof in Web Appendix B). This is a reasonable requirement, since bid landscapes are frequently smoothed (Wang et al., 2016) and practical bidding data often result in concave bid landscapes (Feldman and Muthukrishnan, 2008; Zhang et al., 2014; Zhang and Wang, 2015).

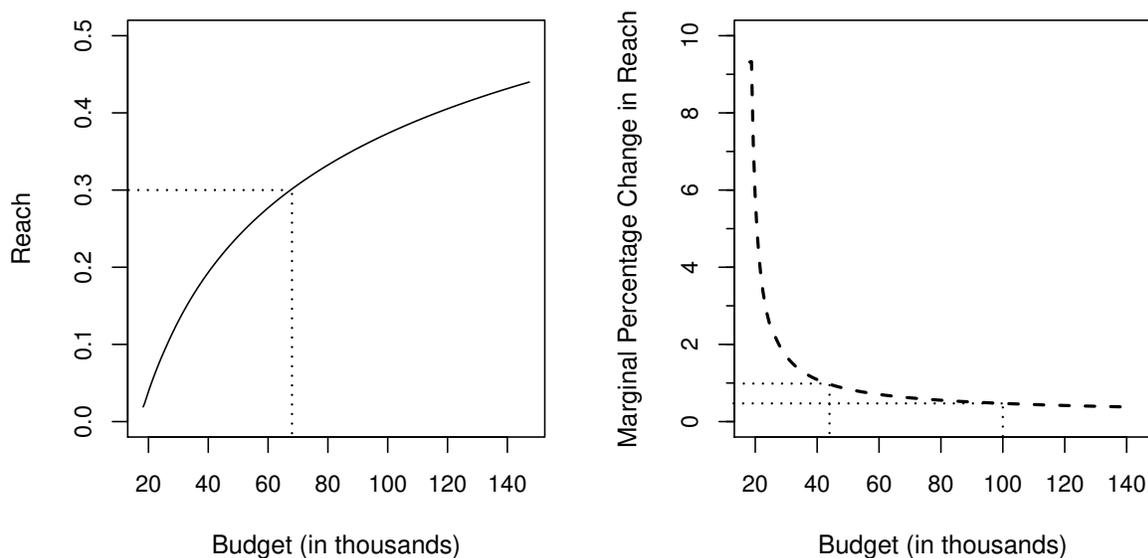
Due to the efficiency of our method, campaign managers can readily use marginal KPI curves to determine optimal budget allocation. Figure 2 provides an example. The left panel illustrates the change in a given KPI (in this case reach), with different budget allocations, for a campaign of up to \$150,000. Clearly there are diminishing marginal returns in reach as

---

<sup>5</sup>The reported computing times are based on R code run on a computer with a 3.40 GHz processor and a Windows 10 operating system, using the cost curve functional forms employed in our paper. Computation time would improve considerably using a more computationally efficient language such as C or C++, or a system with higher computing power. The computation time increases with budget because with small budgets most websites end up with a zero expenditure allocation, which can be determined very quickly. For larger budgets there are more non-zero websites, each of which requires more effort to determine the exact allocation.

the budget increases. The curve in the right panel shows the marginal change in reach, i.e., the percentage change in reach for a 1% change in budget. For example, marginal percentage change in reach at a budget of \$100,000 is 0.5%, i.e., as budget increases from \$100,000 to \$101,000 (a 1% increase), the percentage change in reach is  $(37.5\% - 37.3\%)/37.3\% = 0.5\%$ .

**Figure 2: Reach (left) and the corresponding marginal percentage change in reach (right)**



Campaign managers can use such curves as a budget-setting tool. For example, the firm's managers may have originally designated \$60,000 in order to reach 30% of their customers, but discover that a budget of \$70,000 is required (left panel of Figure 2) and adjust accordingly. Additionally, the firm might find it advantageous to utilize a budget where the marginal returns do not fall below a given value, say 1%. By examining the marginal curve, the firm will discover this occurs at approximately \$45,000 rather than \$60,000 (right panel of Figure 2). Consequently, our method results in a budget landscape, showing advertisers how changes in budget affect outcomes of the campaign.

## METHODOLOGY

In this section we present our method. We first outline the derivation of our optimization function and a coordinate descent algorithm to solve the optimization. The convex structure of the function over which we optimize and the efficient algorithm result in a highly computationally tractable approach, even for thousands of websites. We then discuss a number of extensions including consumer targeting, alternative KPIs, and mandatory media coverage to matched content websites.

### *Model Formulation*

Consider a firm that has a budget  $B$  for a campaign that is to be run over a particular time span (e.g., one month or one quarter). A common goal for such a campaign would be to allocate the firm’s budget across a set of  $p$  possible websites to maximize some KPI measure. In what follows, we first demonstrate our method using *reach* as the KPI of the ad campaign. Later in this section we describe the extension to other exposure-based KPI measures.

Let  $X_j$  represent the number of times an ad is served to a random potential customer at website  $j$  during the course of the ad campaign, where  $j = 1, \dots, p$ . Hence,  $Y = \sum_{j=1}^p X_j$  corresponds to the total number of times the ad is served to this random customer over all websites under consideration. Within this context, our problem can be formulated as a fairly common marketing scenario: given that we are constrained by a budget  $B$ , how do we allocate that budget to maximize reach during our Internet display ad campaign?

Let  $s_j$  represent the probability we win a bid (or correspondingly, the probability that a random customer is served our ad) at website  $j$ , given that we bid  $c_j$  (the bid price per 1000 impressions). An example of such a cost curve is illustrated in Figure 1. Additionally, let  $\tau_j$  represent the number of available impressions (in thousands) at the  $j$ th website during the campaign period. Then the expected total number of ads served at the  $j$ th website will be  $\tau_j s_j$  thousand at a total cost of  $c_j \tau_j s_j$  dollars.<sup>6</sup>

---

<sup>6</sup>Note that although  $s_j$  depends on  $c_j$ , for notational simplicity, we refer to the  $s_j(c_j)$  function as  $s_j$ .

Within this setup, maximizing reach is equivalent to minimizing the probability a random customer is not served the ad at any of the websites under consideration. Hence, our question is equivalent to the following optimization problem:

$$\min_{\mathbf{c}} P(Y = 0|\mathbf{c}) \quad \text{subject to} \quad \sum_{j=1}^p c_j s_j \tau_j \leq B, \quad \text{and} \quad c_j \geq 0, \quad j = 1, \dots, p, \quad (1)$$

where  $\mathbf{c} = (c_1, \dots, c_p)$  denotes the bid prices for impressions from the  $p$  websites. It is challenging to solve Equation 1 because  $p$  may be in the thousands, which means this is an extremely high-dimensional optimization problem. Additionally, the optimal solution to Equation 1 should be able to accommodate corner solutions (i.e., the solution should allow  $c_j = 0$  to arise as an optimal solution for certain websites, so the DSP does not waste time bidding at undesirable websites). We discuss our solution to both challenges below.

We first note that, by the law of iterated expectations,  $P(Y = 0|\mathbf{c}) = E_{\mathbf{Z}} \{P(Y = 0|\mathbf{Z}, \mathbf{c})\}$ , where  $\mathbf{Z} = (Z_1, \dots, Z_p)$  with  $Z_j$  representing the number of page views at website  $j$  by a random customer. In practice the expectation over  $Z$  is difficult to calculate, so we approximate it via

$$P(Y = 0|\mathbf{c}) \approx \frac{1}{n} \sum_{i=1}^n P(Y = 0|\mathbf{Z} = \mathbf{z}_i, \mathbf{c}), \quad (2)$$

where  $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})$  are the page views at websites  $j = 1, \dots, p$ , for the  $i$ th person from a random sample of  $n$  customers. Such data are readily available from commercial browsing-tracking companies (e.g., the comScore Media Metrix data) and other similar data management platforms.<sup>7</sup>

The decomposition above is important because  $\mathbf{Z}$  captures the correlation in viewership of our ad among different websites. In particular, we note that  $X_j$  and  $X_k$  are clearly not independent since a customer who visits website  $j$  may well be more (or less) likely to also visit website  $k$ . However, once we condition on  $\mathbf{Z}$  and  $\mathbf{c}$ , it is reasonable to assume that

---

<sup>7</sup>comScore data aggregate page views by domain names due to privacy considerations. For example, if a user reads 2 articles on Yahoo Sports, then reads 3 Yahoo News articles, comScore would enter the session as 1 visit to Yahoo.com with 5 webpages viewed. Because we use comScore data in our empirical investigation, we follow methodology proposed by Danaher (Danaher, 2007; Danaher et al., 2010) by using page view matrix  $\mathbf{z}_i$  to capture viewership correlations across websites. Nevertheless, if more fine-grained data were available (e.g., page visits in the Yahoo example above were logged as 2 page views at Yahoo Sports and 3 page views at Yahoo News), due to its computational efficiency, our method can also readily accommodate within-website correlation by treating each sub-domain page view as a separate column in the  $\mathbf{z}_i$  matrix.

$X_j$  and  $X_k$  are conditionally independent random variables. In other words, once the bid price and the number of page views at a particular website are fixed, we assume that ads are served randomly to a given customer at website  $j$  with probability  $s_j$ , independently of  $X_k$ ,  $k \neq j$ . For example, again consider Businessweek.com and Reuters.com. It is highly likely that  $X_1$  and  $X_2$  would be positively correlated, because there is up to 95% correlation in the viewership of these two websites. However, conditional on the number of page views a customer has at each website, say  $z_1 = 10$  and  $z_2 = 15$ , and the bid prices at both sites, say  $c_1 = \$5$  and  $c_2 = \$10$ , with winning probabilities of  $s_1 = 0.5$  and  $s_2 = 0.8$ , then  $X_1$  and  $X_2$  become 10 and 15 independent random coin flips with respective probability of being served the ad at 0.5 and 0.8. Under such a setting, knowing that (say)  $X_1 = 5$  provides no new information about  $X_2$  (which has an expected value of 12 in this example).

Hence, we model  $X_j$ , conditional on  $z_{ij}$  and  $c_j$ , as independent Poisson random variables, i.e.,  $X_j|z_{ij}, c_j \sim \text{Pois}(\gamma_{ij})$ , or equivalently,

$$P(X_j = x|Z_j = z_{ij}, c_j) = \frac{e^{-\gamma_{ij}} \gamma_{ij}^x}{x!}. \quad (3)$$

The expected number of ad appearances  $\gamma_{ij}$  is given by the probability of an ad served on a random page view at the  $j$ th website ( $s_j$ ) multiplied by the number of page views ( $z_{ij}$ ), i.e.,  $\gamma_{ij} = s_j z_{ij}$ . For example, if at bidding price  $c_j$  there is a 20% chance of winning bids for an ad at a particular website and a consumer views ten webpages at that site,  $\gamma_{ij} = 0.2 \times 10 = 2$ . So on average, we expect the consumer to be served the ad twice during the ten page views. With this setup, correlations in viewership among the  $p$  websites are directly captured in the  $z_{ij}$  terms, which in turn carry into  $\gamma_{ij}$ . Namely, the  $\gamma_{ij}$  terms are unconditionally correlated but are conditionally (on  $\mathbf{z}_i$  and  $\mathbf{c}$ ) independent.

Thus, conditional on  $\mathbf{z}_i$  and  $\mathbf{c}$ ,  $Y = \sum_{j=1}^p X_j$  also has a Poisson distribution with expected value  $\gamma_i = \sum_{j=1}^p \gamma_{ij} = \sum_{j=1}^p s_j z_{ij}$ , i.e.,

$$P(Y = y|\mathbf{Z} = \mathbf{z}_i, \mathbf{c}) = \frac{e^{-\gamma_i} \gamma_i^y}{y!}, \quad (4)$$

and in particular,  $P(Y = 0|\mathbf{Z} = \mathbf{z}_i, \mathbf{c}) = e^{-\gamma_i}$ .<sup>8</sup> For ease of exposition, we present our method with  $\gamma_i$  being a deterministic function of  $\mathbf{c}$  and  $\mathbf{z}_i$ , i.e.,  $\gamma_i = \sum_j s_j z_{ij}$ . In Web Appendix C we describe an extension where  $\gamma_i$  comes from a random distribution with  $E(\gamma_i) = \sum_j s_j z_{ij}$ , which would account for unobserved heterogeneity (such as a user’s likelihood of seeing an ad).

Combining Equation 4 with our original Equation 1 and the sample approximation in Equation 2, we obtain the following optimization problem:

$$\min_{\mathbf{c}} \frac{1}{n} \sum_{i=1}^n e^{-\gamma_i} \quad \text{subject to} \quad \sum_{j=1}^p c_j s_j \tau_j \leq B \quad \text{and} \quad c_j \geq 0, \quad j = 1, \dots, p. \quad (5)$$

The optimization function in Equation 5 can be viewed as an approximation to  $E_{\mathbf{Z}} \{e^{-\gamma}\}$ . We prove in Web Appendix D that, under suitable conditions, our optimization function converges to the population level function as  $n$  and  $p$  approach infinity.

Equation 5 has the following appealing properties. First, although not immediately obvious, the form of Equation 5 encourages the corner solutions we desired. Similar to the well-known Lasso formation, the constraint region in our setting has the form of a multi-dimensional rhomboid, i.e., it has many “sharp points” on the axes where  $c_j = 0$ . Hence, the solution within the constraint region which minimizes the optimization function often falls on one of these points in a similar way to Lasso (see Web Appendix B for further intuition of this phenomenon). As budget  $B$  decreases, the constraint becomes confined to a smaller region around the origin and, hence, all of the  $c_j$ ’s are shrunk towards zero, producing a sparse solution. Therefore, while our optimization function is not sum of squares as in Lasso, the form of our constraint region is similar to that for Lasso (with the region being constrained to be positive in our setting since spending cannot be negative). Second, the objective function in Equation 5 is well-behaved and convex as long as the cost curve functions  $s_1, \dots, s_p$  are concave (see Web Appendix B for proof). This in turn makes our method highly computationally efficient. Lastly, as discussed above, our optimization function directly incorporates correlations among viewerships. In particular, we show in Web

---

<sup>8</sup>Note that although modeling  $X_{ij}$  using a Poisson distribution works well in practice and provides a simple expression for  $P(Y = 0|\mathbf{Z} = \mathbf{z}_i, \mathbf{c})$ , the same approach could be applied with other distributions. For example, in Web Appendix A we provide an alternative implementation where the Poisson distribution is replaced by a Negative Binomial distribution.

Appendix E that our definition of reach can be expressed as a term assuming independence plus an adjustment for the covariance among the  $Z_j$  terms. We also provide an analytic calculation of change in reach as a function of budget  $B$  in Web Appendix F.

### *The Optimization Algorithm*

A common approach in statistics, economics, and operations research for solving a constrained optimization is to re-express it as an unconstrained optimization via a Lagrangian formulation. We take that approach with Equation 5, reformulating it as:

$$\min_{\mathbf{c}} \frac{1}{n} \sum_{i=1}^n e^{-\gamma_i} + \frac{\lambda}{n} \left( \sum_{j=1}^p c_j s_j \tau_j - B \right) \quad \text{subject to} \quad c_j \geq 0, \quad j = 1, \dots, p, \quad (6)$$

where  $\lambda > 0$  is the Lagrangian multiplier.<sup>9</sup> Note  $\lambda$  must be greater than zero in our setting given that budget must always be finite. It is evident that, for each given budget, there is a corresponding Lagrangian multiplier  $\lambda$ . For a given number of websites, as budget increases,  $\lambda$  decreases, and the algorithm allocates more budget to more websites. As budget decreases,  $\lambda$  increases, and we get a sparser solution.

Let  $w_j = c_j s_j \tau_j$  represent the amount spent at website  $j$ . Note that, since  $\tau_j$  is given and  $s_j$  is a function of  $c_j$ , there is a one-to-one correspondence between  $w_j$  and  $c_j$ . Hence, optimizing over  $w_j$  will also solve for  $c_j$ . Using this change of variable, Equation 6 can be simplified as follows:

$$\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n e^{-\gamma_i} + \frac{\lambda}{n} \sum_{j=1}^p w_j \quad \text{subject to} \quad w_j \geq 0, \quad j = 1, \dots, p. \quad (7)$$

where  $\gamma_i = \sum_{j=1}^p s_j(w_j) z_{ij}$  is a function of  $\mathbf{w}$ , and  $B$  drops out since it is a constant which does not affect the solution.

Although there is no direct closed form solution to Equation 7, similar problems have been extensively studied in recent statistics literature (e.g., Efron et al. 2004; Friedman et al. 2010; Goeman 2010; Hesterberg et al. 2008; Rosset and Zhu 2007; Schmidt et al. 2007). As

---

<sup>9</sup>In statistics, this is commonly referred to as a penalized optimization equation. The  $\frac{\lambda}{n} \sum_j c_j s_j \tau_j$  penalty would frequently be written as an  $\ell_1$  penalty rather than a summation term. However, for our setup, these two formulations are identical, since we have the condition  $c_j \geq 0$  for all  $j$ .

---

**Algorithm 1 Coordinate Descent Algorithm for Budget Optimization**


---

1. Specify a maximum budget,  $B_{\max}$ .
  2. Initialize algorithm with  $\mathbf{w}_0 = \mathbf{0}$ , and  $\lambda$  corresponding to  $B = 0$ .
  3. For  $j$  in 1 to  $p$ ,
    - (a) Marginally optimize Equation 7 over a single  $w_j$ , keeping  $w_1, w_2, \dots, w_{j-1}, w_{j+1}, \dots, w_p$  fixed.
    - (b) Iterate until convergence.
  4. Increase budget by incrementally decreasing  $\lambda$  over a grid of values, with each  $\lambda$  corresponding to a budget, and repeat Step 3 until reaching  $B_{\max}$ .
- 

a result, there exist highly efficient algorithms for solving such problems. In this paper, we utilize one such algorithm known as coordinate descent to solve Equation 7 over a grid of values for  $\lambda$ , which in turn provides optimal allocations for a range of possible campaign budgets. While certainly not the only candidate algorithm for our optimization, coordinate descent has proved to be both simple and highly efficient for problems similar to ours (e.g., Wu and Lange 2008; Friedman et al. 2010; Breheny and Huang 2011).

Essentially, the coordinate descent algorithm simplifies our optimization to a series of one-dimensional optimizations as described in Algorithm 1 (see Web Appendix B for more details of the algorithm). What makes this approach so efficient is that each update step is fast to compute and typically not many iterations are required to reach convergence. This efficiency means that it is feasible to solve the optimization function for a large range of budgets in a relatively short period of time. Indeed, there are additional computational savings when the optimization function is iteratively solved over multiple budgets. Once we compute a solution for any particular budget, the optimal allocation for any similar budgets can be rapidly found by initializing  $\mathbf{w}_0$  in Algorithm 1 with the allocation from the previous budget. Since the new solution will be close to the old one, only a few iterations will generally be required to converge to the new allocation. Note that our coordinate descent algorithm is guaranteed to converge to a global optimum provided the optimization function in Equation 7 is convex (Luo and Tseng, 1992). We prove in Web Appendix B that a sufficient condition for convexity is that each  $s_j(w_j)$  is concave.

We use a second-order Taylor approximation to implement step 3a. Specifically, let  $\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_p$  represent the current estimates for  $\mathbf{w}$ . Then, using standard calculations it is not hard to show that Step 3a can be approximately solved by optimizing:

$$\min_{w_j} \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \eta_{ij} (w_j - \tilde{w}_j)^2 - \theta_{ij} (w_j - \tilde{w}_j) \right) + \frac{\lambda}{n} w_j \quad \text{subject to} \quad w_j \geq 0. \quad (8)$$

where  $\theta_{ij} = e^{-\tilde{\gamma}_i} s'_j(\tilde{w}_j) z_{ij}$ ,  $\eta_{ij} = e^{-\tilde{\gamma}_i} z_{ij} (z_{ij} s'_j(\tilde{w}_j)^2 - s''_j(\tilde{w}_j))$ , and  $\tilde{\gamma}_i = \sum_{j=1}^p s_j(\tilde{w}_j) z_{ij}$ . We show in Web Appendix B that the solution to Equation 8 is given by

$$w_j = \begin{cases} \tilde{w}_j + \frac{\sum_{i=1}^n \theta_{ij} - \lambda}{\sum_{i=1}^n \eta_{ij}} & \text{for } H_j > \lambda \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where  $H_j = \sum_{i=1}^n (\tilde{w}_j \eta_{ij} + \theta_{ij})$  (note that  $H_j$  is always positive here). Equation 9 incorporates the  $w_j \geq 0$  condition by testing if the  $w_j$  coefficient has been forced below zero by the update. If it has, we set that coefficient to 0, the minimum value allowed (since budget cannot be negative). This equation can be computed quite efficiently so Equation 7 can be solved by iteratively computing Equation 9 for  $j$  from 1 to  $p$  and repeating until convergence.<sup>10</sup>

## Extensions

In this section we consider several extensions to the underlying methodology.

*Consumer Targeting.* In practice, campaign managers often wish to target particular consumer groups based on users' demographic characteristics and/or their past purchase/browsing behavior. The former is referred to as demographic targeting and the latter is referred to as behavioral targeting. In what follows we discuss how demographic and behavioral targeting can be readily incorporated into the proposed model above.

In the case of demographic targeting, for example, the McRib campaign may be geared toward households with children and lower income. Our method can easily accommodate

---

<sup>10</sup>Because we employ a Taylor approximation in our algorithm, we also empirically verify the convergence of the approximation. We ran our algorithm with numerous initialization points to determine if the optimization had converged to a global optimum. In all cases, we obtained identical solutions regardless of initialization points, and the convergence was achieved under very few iterations. For further testing, we ran a complete enumeration comparison to verify whether our solution resulted in a global optimum. For the results of this complete enumeration approach, please refer to Web Appendix G.

such needs by allocating, say 80% of the budget, to the targeted consumer demographic group and using the remaining budget to attract other potential consumers. Since different consumer groups often involve differentiated costs, the campaign manager can run the method separately for each consumer demographic group and its corresponding cost curves. For example, to target the group comprised of lower income households with children, the McRib campaign manager would run the algorithm first on the subset of lower income families, and second on the remaining consumers, each with their corresponding cost curves. This ultimately gives two KPI metric curves, but since the consumer groups are distinct, there is no overlap. Managers can decide how they wish to combine the curves to get an overall blueprint of the campaign.

When it comes to behavioral targeting, ads are aimed at certain website visitors based on their past purchase/browsing behavior. For example, Norwegian Cruise Line (NCL) may want to implement an ad campaign targeting users who visited at least one aggregate travel site in the previous month. In such cases, if DSPs have access to behavioral-tracking tools, a similar approach could be used to accommodate behavioral targeting. Within this setting, once the high-value consumers are identified (for example, previous aggregate travel site users), they become a new demographic to target. Assuming it is more costly to re-target a user, these consumers will now have new cost curves compared to users who have not yet been tracked. Instead of each website having a cost curve that holds for all users, now all re-targeted users have a new cost curve for each subsequent website they visit. In such cases, the campaign manager can simply run our method just on the subset of targeted users (e.g. targeting last month’s travel site users) with their corresponding cost curves.

Due to inherent similarities between these two types of consumer targeting, we provide an example of demographic targeting in the McRib Advertising Campaign case study. An illustration of behavioral targeting is available from the authors upon request.

*Alternative Exposure-Based Key Performance Indicators.* Consider the more general form of Equation 1:

$$\min_{\mathbf{c}} -g(\mathbf{c}) \quad \text{subject to} \quad \sum_j c_j s_j \tau_j \leq B \quad \text{and} \quad c_j \geq 0, \quad j = 1, \dots, p, \quad (10)$$

where  $g(\cdot)$  represents a given key performance indicator (KPI) of interest. So far we have concentrated on reach, i.e.,  $g(\mathbf{c}) = P(Y > 0|\mathbf{c})$ . However, there are several alternative exposure-based KPIs that are commonly used in Internet display ad campaigns. In what follows we describe some of the most commonly applied KPIs and how our method can be adapted to these alternative metrics:<sup>11</sup>

- *Frequency*, the average number of exposures among those reached:  $g(\mathbf{c}) = E_Y \{Y|Y > 0, \mathbf{c}\}$ . Under our modeling setup, frequency is given by  $E_{\mathbf{Z}} \left\{ \frac{\gamma}{1-e^{-\gamma}} \right\}$  which we approximate by  $\frac{1}{n} \sum_{i=1}^n \frac{\gamma_i}{1-e^{-\gamma_i}}$ .
- *GRPs*, frequency multiplied by reach: It is not hard to show that GRP corresponds to the average number of exposures:  $g(\mathbf{c}) = E_Y \{Y|\mathbf{c}\}$  (Danaher, 2008). Under our modeling setup, GRP is given by  $E_{\mathbf{Z}} \{\gamma|\mathbf{c}\}$  which we approximate by  $\frac{1}{n} \sum_{i=1}^n \gamma_i$ .
- *Effective frequency* and *frequency capping*<sup>12</sup>, the proportion of customers who view the ad between  $k_a$  and  $k_b$  times, where  $k_a < k_b$  respectively represent lower and upper bounds on ad exposures (e.g., Krugman, 1972; Naples, 1979; Danaher et al., 2010). For example, sales conversions and profits from online display ads might be highest when the consumer is served an ad within a certain range of frequencies (e.g., one to three times) during the duration of the ad campaign. In this setting,  $g(\mathbf{c}) = P(k_a \leq Y \leq k_b|\mathbf{c})$  which is given by  $E_{\mathbf{Z}} \left\{ \sum_{y=k_a}^{k_b} \frac{e^{-\gamma} \gamma^y}{y!} \right\}$ . We can approximate this expectation using  $\frac{1}{n} \sum_{i=1}^n \sum_{y=k_a}^{k_b} e^{-\gamma_i} \gamma_i^y / y!$ . So, using the example of  $1 \leq Y_i \leq 3$ , our problem would involve maximizing  $\frac{1}{n} \sum_{i=1}^n e^{-\gamma_i} (\gamma_i + \frac{1}{2} \gamma_i^2 + \frac{1}{6} \gamma_i^3)$ .

The proposed method can be readily modified to accommodate all these KPIs. Again we take a second-order Taylor expansion, resulting in equations with a similar form to Equations 8 and 9. The only difference becomes new values for  $\theta_{ij}$  and  $\eta_{ij}$  in each case. When optimizing

<sup>11</sup>Firms could also readily modify our approach to maximize alternative KPIs such as clicks or purchases. For example, as long as the advertiser has information on click-through rates (or conversion rates) from different websites, say  $q_j$ , we can directly accommodate such needs by defining  $\gamma_i$  as the expected number of clicks (or conversions) and setting  $\gamma_i = \sum_j q_j s_j z_{ij}$ . Because  $q_j, j = 1, \dots, p$ , is known, the optimization can then be carried out in a similar fashion to the approach already described.

<sup>12</sup>This latter term typically refers to the setting where the upper limit  $k_b$  is set to a particular value so users are not exposed to an ad more than a given number of times.

frequency we apply Equations 8 and 9 with

$$\theta_{ij} = \frac{s'_j z_{ij}}{1 - e^{-\gamma_i}} - \frac{s_j s'_j z_{ij}^2 e^{-\gamma_i}}{(1 - e^{-\gamma_i})^2}, \quad \eta_{ij} = \frac{-s''_j z_{ij}}{1 - e^{-\gamma_i}} + \frac{z_{ij}^2 e^{-\gamma_i} (s_j'^2 (2 - z_{ij}) + s_j s_j'')}{(1 - e^{-\gamma_i})^2} - \frac{2z_{ij}^3 e^{-2\gamma_i} s_j s_j'^2}{(1 - e^{-\gamma_i})^3}.$$

Alternatively, the corresponding values for *GRPs* are  $\theta_{ij} = s'_j z_{ij}$  and  $\eta_{ij} = -s''_j z_{ij}$ , while those for *effective frequency* (with  $k_a = 1$  and  $k_b = 3$ ) are given by

$$\theta_{ij} = s'_j z_{ij} e^{-\gamma_i} (1 - \gamma_i^3/6), \quad \eta_{ij} = z_{ij} e^{-\gamma_i} [\gamma_i^2 s_j'^2 z_{ij}/2 - (s_j'' - z_{ij} s_j'^2)(1 - \gamma_i^3/6)].$$

As one application of an alternative KPI to reach, we demonstrate the effective frequency metric in the McRib Advertising Campaign case study.

*Mandatory Media Coverage to Matched Content Websites.* Aside from consumer targeting, a firm might wish to impose mandatory media coverage to certain subsets of websites. For example, when planning the online advertising campaign for its annual “wave season,” Norwegian Cruise Lines (NCL) may want to specify that the ad agency allocates a certain minimum budget to advertising on aggregate travel sites such as Orbitz or Expedia in addition to other websites.

In this example, Equation 7 can be adjusted to constrain  $w_j$  above a certain threshold, say  $w_j \geq \min_j$ , to ensure that a minimum budget is allocated to each aggregate travel website  $j$ . Using the same approach as for optimizing Equation 7, we can show that the new optimization is accomplished by setting the “otherwise” condition in Equation 9 to a minimum non-zero amount. We would then replace Equation 9 with the following. An example of this extension is provided in the NCL Wave Season Online Advertising Campaign case study.

$$w_j = \begin{cases} \tilde{w}_j + \frac{\sum_{i=1}^n \theta_{ij} - \lambda}{\sum_{i=1}^n \eta_{ij}} & \text{for } H_j - \lambda > \min_j \\ \min_j & \text{otherwise.} \end{cases} \quad (11)$$

## SIMULATION STUDIES

In this section, we demonstrate our method in two simulated settings. We first present a conceptual demonstration of our method with a simple five-website example. We then show

how our method can be used for optimal budget allocation when the number of websites under consideration is very large (e.g., 5000 websites).<sup>13</sup>

The cost curves,  $s_j$ , are an integral part of our methodology. Although in practice these curves can be assumed known to the campaign manager, in this paper we demonstrate our method using a shifted logistic curve (Figure 1). A standard logistic curve follows the equation  $s_j(c_j) = \frac{e^{c_j a_j + b_j}}{1 + e^{c_j a_j + b_j}}$ , where  $c_j$  is the bidded CPM for website  $j$ ,  $a_j$  represents the steepness of the curve, and  $b_j$  is the standard position parameter for the logistic curve. For simplicity, we assume  $b_j = 0$  for all  $j$ , although this is not crucial for the methodology. Further, since our cost curves should satisfy the constraint  $s_j(0) = 0$ , we shift the logistic curve as follows  $s_j(c_j) = 2\left(\frac{e^{c_j a_j}}{1 + e^{c_j a_j}} - \frac{1}{2}\right) = \frac{e^{c_j a_j} - 1}{e^{c_j a_j} + 1}$ . In our setting, the more important parameter is  $a_j$ , which controls the steepness of the cost curve. We set  $a_j$  at the value which gives bidders a 50% chance of winning a bid at an average CPM value. For example, if the average CPM on a given website is \$5.00, the  $a_j$  for that website is found by solving  $s_j(5) = 0.5$ , i.e.,  $a_j = \frac{1}{5} \log 3$ .

### ***Conceptual Demonstration with 5 Websites***

We first consider a five website setting so features of our method can be represented graphically. Table 1 gives the average CPM,  $c_0$  (that is, the value at which there is a 50% chance of winning the bid), expected total number of page views,  $\tau_j$ , and the total number of unique visitors for each website. We kept the number of unique visitors constant so the only factors influencing budget allocation across the five websites are  $c_0$  and  $\tau_j$ . These website examples correspond to moderately- to highly-visited websites, generated to mimic similar website and page view distributions observed in the real comScore data used in our case studies.

Figure 3 shows the five cost curves on the left, and the number of impressions bought relative to the bidded CPM on the right. Here, Website 1 (black solid) has the lowest  $c_0$  (\$3.00), through to Website 5 (dashed purple) which has the highest  $c_0$  (\$7.00). Consider

---

<sup>13</sup>For simplicity, the simulation data in both examples are generated independently without correlations. Since the proposed method is designed to leverage correlations across sites, this setup provides a lower bound with respect to advantages from our approach. Additionally, we chose to simulate the dataset for the 5,000 website example because data cleaning in comScore for this many websites would be prohibitively time consuming.

**Table 1: Average CPM and expected number of page views for five-website example**

	Avg. CPM ( $c_0$ )	Page Views (in mil)	Number of Unique Visitors
Website 1	3.0	30	10,000
Website 2	4.0	20	10,000
Website 3	5.0	50	10,000
Website 4	6.0	10	10,000
Website 5	7.0	40	10,000

a bid of CPM = \$5.00. For Website 1 ( $c_0 = \$3.00$ ),  $s_1(5) = 0.723$  so a \$5.00 bid results in approximately  $s_1(5)\tau_1 = 0.723 \times 30 = 21.7$  million impressions bought during the campaign. By contrast, for Website 5 ( $c_0 = \$7.00$ ),  $s_5(5) = 0.373$  so a \$5.00 bid results in only  $s_5(5)\tau_5 = 0.373 \times 40 = 14.9$  million impressions, even though  $\tau_5 > \tau_1$ .

Figure 4 demonstrates how budget is being allocated at the five websites at three particular budget points: \$100,000 (\*), \$500,000 ( $\Delta$ ), and \$2,000,000 ( $\boxtimes$ ). The left-hand plot shows the total reach achieved across the five websites for budgets up to approximately \$2,000,000. On the right-hand side, the five website cost curves show how budget was allocated across the five websites by the algorithm. The location of each symbol represents the bid CPM at that website for a given budget. Our method first allocates the majority of the budget to the lower cost, but higher expected page view, websites (One and Three). Then, as viewership at those websites is maxed out, it begins to allocate to the higher cost or lower page view websites to reach different viewers; demonstrating on a small scale the way in which the optimization allocates budget across the considered websites.

### ***Simulated Large-Scale Problem: 5000 Websites***

In practice, most Internet media selection problems involve far more than a handful of websites. Here we illustrate the proposed method in a simulated setting involving thousands of websites. Specifically we simulate an Internet usage matrix of 50,000 people over 5,000 websites. To mimic the observed comScore data (which had many zero entries) we simulate page views to each website by generating standard normal random variables, which are rounded and converted to absolute values, and then multiplied by a random integer from

Figure 3: Example cost curves for the five simulated websites (left) and number of impressions bought at each CPM (right)

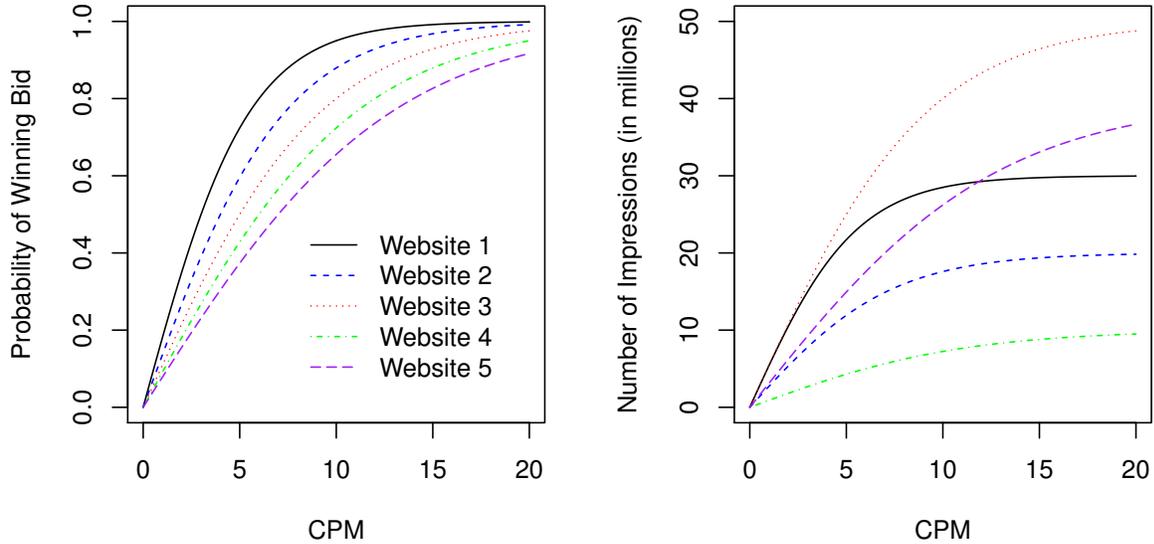
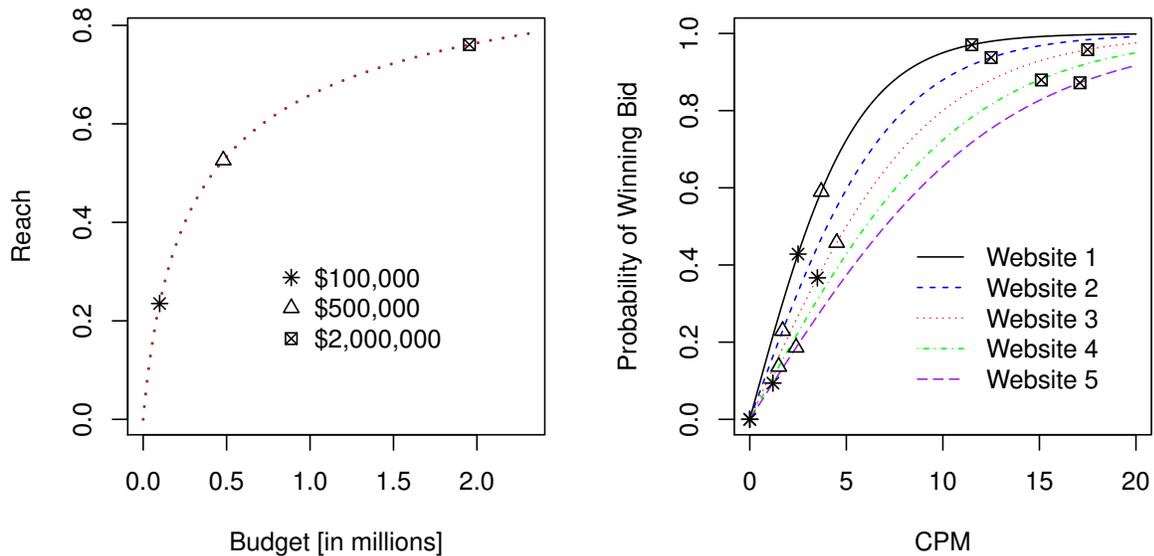


Figure 4: Total reach achieved using the five websites given budget (left) and the allocation of budget by bidded CPM at three points (right)



zero to ten with higher weight on a value of zero. The average CPMs of these websites are randomly generated, chosen from 0.25 to 8.00 in increments of 0.25.

We compare the proposed method to the following benchmark approaches: 1) equal allocation over all 5,000 websites; 2) equal allocation over the 1,000 websites with lowest average CPM; 3) cost-adjusted (i.e., average page views/average CPM for each website) allocation over the most visited 25, 50, and 100 websites; and 4) a benchmark greedy algorithm that sequentially allocates budget to one website at a time. To implement approaches 1) through 3), we select  $c_j$  so that  $w_j = s_j \tau_j c_j$  is allocated evenly (in the case of the first two approaches) or by cost-adjustment (in the case of the third approach) across all websites.

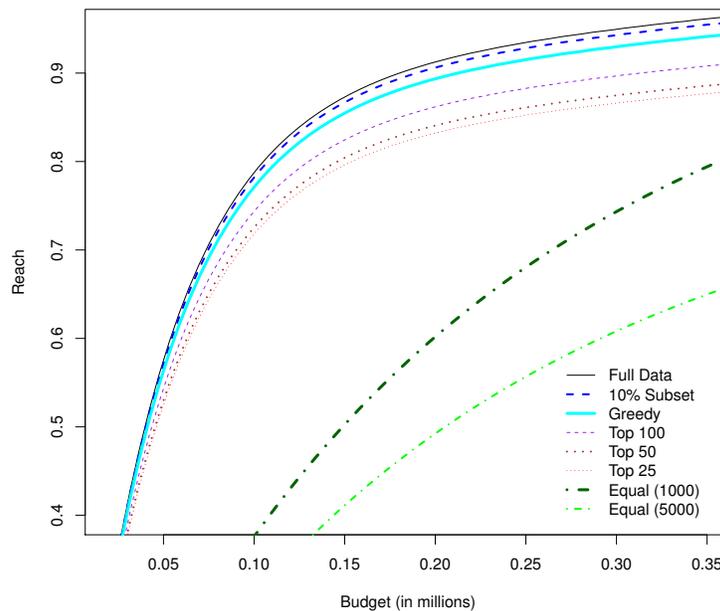
Our benchmark greedy algorithm follows the description in Danaher (2008). Similar approaches have been well-represented in media allocation problems and are a popular choice in high-dimensional or large-scale problems, when it is impractical to simultaneously optimize over all  $p$  variables (e.g. Zhang et al. 2000; Parsons et al. 2004; Hatano et al. 2015; Miyuchi et al. 2015; Zhang et al. 2017). The benchmark greedy approach is a simple, but less effective, way to optimize Equation 5. As with all greedy methods, it only looks one step ahead and increases the allocation to the currently best website, defined as the website that causes the largest single increase in reach. As we increase the allocation to this website, other websites will become more competitive in terms of increasing reach and eventually one of them will pass the first website and become the new currently best website. At that point the greedy method freezes the first website at its current allocation and shifts additional budget allocations to the new “best” website. We continue with this process, sequentially increasing the allocation to a single website, until  $\sum_j w_j = B$  so the entire budget has been allocated.

The benchmark greedy algorithm increases reach with the addition of each new website but does not fully optimize Equation 5 because it only adjusts  $w_j$  once and then freezes budget spent on that site after the next website is added. By comparison, our approach fully optimizes Equation 5, since we do not fix  $w_j$  at its current value as new websites are added. Instead, we iteratively update all  $w_j$  (either up or down) until all available budget has been allocated. See Web Appendix H for further discussions on conditions under which the advantages of our approach over this benchmark greedy approach are amplified or lessened. In general, our approach exhibits distinct advantages over the benchmark greedy approach

when 1) the websites under consideration are similar in terms of cost and/or visitation; and/or 2) websites exhibit overlap in viewership.

Figure 5 shows the resulting comparisons with the proposed method represented as a blue dashed line. To estimate reach, we first calibrate the proposed and benchmark greedy methods using a 10% subset of the 50,000 users as the training data set, then report reach on the 90% holdout data. A similar approach is used in all reach curve comparisons in this paper. As a sensitivity analysis, we ran these results with 20 different randomly chosen 10% subsets. Figure 5 is presented as the average reach across the 20 runs at each budget point. It is important to note the relative rankings of the reach performances did not change in any of the 20 runs.

**Figure 5: Simulated Data-5000 Websites (reach is calculated using 90% holdout data)**



We compare the holdout reach calculated from the 10% subset to the results we would achieve over an optimization calibrated on the entire 50,000-user data set (black solid). Even with only a ten percent subset of the data, the proposed method yields reach estimates very similar to those calibrated on the entire dataset. In addition, the proposed method outperforms all benchmark approaches. We find that the equal allocation approaches are clearly an inefficient budget utilization. The cost-adjusted approaches show significant improvement,

but they still perform worse than our method. As one would expect (given the more sophisticated approach and its incorporation of the cost curves), the benchmark greedy algorithm performs better than all other benchmark approaches but still not as well as our proposed method. Overall, we show that the proposed method can be used to effectively and efficiently allocate advertising budget across a very large set of websites.

## EMPIRICAL INVESTIGATION

In this section, we discuss two case studies where we use the proposed method and its extensions for McDonald’s McRib and Norwegian Cruise Lines Wave Season online display advertising campaigns. These empirical illustrations are based on the 2011 comScore Media Metrix data, which comes from the Wharton Research Data Service ([www.wrds.upenn.edu](http://www.wrds.upenn.edu)). comScore uses proprietary software to record daily webpage usage information from a panel of 100,000 Internet users (recorded anonymously by individual computer). Therefore, the comScore data can be used to construct a matrix of all websites visited and the number of page views each computer had at each website during a particular time period. A number of prior studies in marketing have utilized the comScore data (e.g., Danaher, 2007; Liaukonyte et al., 2015; Montgomery et al., 2004; Park and Fader, 2004).

### *Case Study 1: McDonald’s McRib Sandwich Online Advertising Campaign*

In our first case study, we consider a yearly promotion for McDonald’s McRib Sandwich, which is only available for approximately one month each year. Because the McRib is often offered in or around December (AdAge, 2012), we consider the comScore data from December 2011 to approximate an actual campaign. In particular, we manually identified the 500 most visited websites that also supported Internet display ads. Our data then contains a record of every computer that visited at least one of these 500 websites at least once (56,666 users). Thus,  $Z$  is a 56,666 by 500 matrix. We then separate our full data set into a ten percent training data set (5667 users), and a ninety percent holdout data set. As before, we use the training data to fit the method, and calculate reach on the holdout data. As a sensitivity

analysis, we ran these results with 100 different randomly chosen 10% subsets. For clarity, we report results for only one run of the data in Tables 3 and 4 below. However, all reach curves in the figures that follow represent the average reach across the 100 runs. Relative rankings of the reach performances did not change in any of the 100 runs.

Table 2 provides the makeup for the sixteen categories among the 500 websites we consider in this application. The Total Websites column lists the total number of websites in each category, while Total Impressions provides the total number of webpages viewed and thus the total number of ad impressions available (in millions).<sup>14</sup> For simplicity, the average CPM values for each website are based on average costs of the website categories provided by comScore Inc.’s Media Metrix data from May 2010 (Lipsman, 2010). As discussed in the simulation study section, we then calculate shifted logistic cost curves for each category of websites based on a  $c_0$  of these average CPMs.<sup>15</sup> Table 2 shows that Entertainment and Gaming are by far the largest categories by the number of websites (with 92 and 77 websites out of 500, respectively), Portals are the largest category by the number of total impressions, and Sports, Newspaper, and General News are the most expensive categories at which to advertise (all over \$6.00 on average). Additionally, advertising costs vary considerably across these website categories. In Web Appendix I (Table A2), we also provide an overview of viewership correlations within and across each of the sixteen website categories. Finally Table 2 shows the number of websites in each category which fall in the top 10, 25, and 50 most-visited websites.

In what follows we describe results from the proposed method over three different scenarios: 1) the original approach that maximizes overall reach, 2) our extension to maximize reach among targeted consumer demographics, and 3) our extension to maximize effective frequency with an upper and lower limit of ad exposures.

*McRib Campaign: Maximizing Overall Reach.* In this subsection, we assume McDonald’s is attempting to reach as many users as possible. Because McDonald’s is a large company which already enjoys significant brand awareness, the goal of its McRib campaign is to remind

---

<sup>14</sup>To get these total impressions values, we follow the procedure of Danaher et al. (2010) by extrapolating the comScore data to the effective U.S. Internet population size in Dec. 2011.

<sup>15</sup>As noted previously, in practice DSPs would apply actual, individualized cost curves for all sites in such an optimization, based on the results of their own empirical evaluations.

**Table 2: Website Categories in McRib Data**

Category	Total Websites	Total Imp. (in mil)	Avg. CPM	Top 10	Top 25	Top 50
Community	23	3322.5	2.10	0	1	4
E-mail	7	4044.7	0.94	1	3	5
Entertainment	92	7511.6	4.75	0	0	0
Fileshare	28	3226.2	1.08	0	2	7
Gaming	77	6404.2	2.68	0	0	1
General News	12	1468.6	6.14	0	0	0
Information	47	5994.9	2.52	1	1	3
Newspaper	27	1926.1	6.99	0	0	0
Online Shop	29	3346.5	2.52	0	1	1
Photos	9	770.4	1.08	0	0	2
Portal	30	30407.7	2.60	3	5	7
Retail	57	9895.1	2.52	1	2	7
Service	18	4574.4	2.52	1	2	2
Social Network	17	5417.1	0.56	3	8	11
Sports	17	1637.2	6.29	0	0	0
Travel	10	529.7	2.52	0	0	0

consumers who already know about McDonald’s or the McRib that the seasonal sandwich is returning.

Tables 3 and 4 report the total number of websites chosen in each category and the average bidding price of chosen sites under two budgets (\$100,000 and \$1,000,000, respectively) for this standard setup (Original), as well as for the two extensions (Targeted Consumers and Targeted Frequency) considered later in this section.<sup>16</sup> Table 3 shows the results based on a relatively small budget, with an overall expected reach of just over 10%. Thus here, budget is allocated to the most valuable websites and website categories, generally chosen by viewership and cost. In contrast, Table 4 presents results from a higher budget of \$1,000,000, where reach is approximately 40%. At this budget the allocation is more diversified, and budget is allocated across a greater number and variety of websites.

<sup>16</sup>Detailed budget allocation results for bid price at each website are available from authors upon request.

**Table 3: Website Categories Chosen by Method at a budget of \$100,000, McRib**

Category	Total Sites	Avg CPM	Original		Targeted Consumers		Targeted Frequency	
			Chosen Sites	Avg Bid Price	Chosen Sites	Avg Bid Price	Chosen Sites	Avg Bid Price
Community	23	2.10	0	0	0	0	0	0
E-mail	7	0.94	5	0.61	4	0.59	7	0.48
Entertainment	92	4.75	0	0	3	2.02	0	0
Fileshare	28	1.08	9	0.54	9	0.52	15	0.37
Gaming	77	2.68	5	0.77	20	1.35	25	0.69
General News	12	6.14	0	0	0	0	0	0
Information	47	2.52	0	0	0	0	0	0
Newspaper	27	6.99	0	0	0	0	0	0
Online Shop	29	2.52	2	1.11	4	1.06	9	0.89
Photos	9	1.08	2	0.61	1	0.9	4	0.58
Portal	30	2.60	0	0	0	0	0	0
Retail	57	2.52	3	1.01	3	0.98	11	0.73
Service	18	2.52	0	0	0	0	0	0
Social Network	17	0.56	16	0.28	16	0.31	17	0.21
Sports	17	6.29	0	0	0	0	0	0
Travel	10	2.52	0	0	0	0	0	0

Despite their differences in available budget, both tables show some general trends. For example, not surprisingly we do not bid on many websites in relatively expensive categories such as Sports, Newspaper, and General News. Advertising at a more expensive website is only desirable when that website can reach an otherwise unreachable audience. In this case, other websites, such as Social Networking, offer a relatively inexpensive way to reach consumers who are visiting other websites as well. Table A2 in Web Appendix I shows social networking sites have relatively high correlations in viewership across other site categories, with the only exceptions being e-mail and gaming sites. Consequently, the optimization ultimately bids heavily across the entire Social Networking category (bidding at almost all 17 websites) and leaves out the expensive categories where reach would be duplicated. It is also worth noting that at the \$1,000,000 budget our method places bids at all websites in the

**Table 4: Website Categories Chosen by Method at a budget of \$1,000,000, McRib**

Category	Total Sites	Avg CPM	Original		Targeted Consumers		Targeted Frequency	
			Chosen Sites	Avg Bid Price	Chosen Sites	Avg Bid Price	Chosen Sites	Avg Bid Price
Community	23	2.10	7	2.05	9	2.01	9	1.99
E-mail	7	0.94	7	0.94	7	0.96	7	0.91
Entertainment	92	4.75	3	3.11	6	3.42	7	3.03
Fileshare	28	1.08	22	1.15	20	1.11	24	1.07
Gaming	77	2.68	22	2.51	35	3.51	30	2.48
General News	12	6.14	0	0	0	0	0	0
Information	47	2.52	8	2.38	7	2.40	11	2.35
Newspaper	27	6.99	0	0	0	0	0	0
Online Shop	29	2.52	6	2.53	11	2.39	9	2.52
Photos	9	1.08	8	1.04	9	1.02	9	1.00
Portal	30	2.60	3	2.19	2	2.17	8	2.14
Retail	57	2.52	15	2.41	18	2.37	22	2.36
Service	18	2.52	11	2.38	9	2.26	13	2.37
Social Network	17	0.56	17	0.63	17	0.61	17	0.59
Sports	17	6.29	1	4.31	0	0	0	0
Travel	10	2.52	4	2.50	2	2.48	5	2.44

e-mail category. In addition to the relative lower cost of advertising on these websites, there is a very low within-category correlation in viewership among e-mail sites (0.01 absolute average correlation). This indicates the same consumer often does not visit more than one e-mail site, so bidding for more e-mail website impressions can result in a larger increase in reach.

Figure 6 shows the reach results from the proposed method, along with the following benchmark methods: equal allocation over all 500 websites; equal allocation across all websites in the five lowest average CPM categories (Community, E-mail, Fileshare, Photos, and Social Network, which total 84 websites among them); cost-adjusted allocation across the top 10, 25, and 50 most visited websites; and the benchmark greedy algorithm described

previously. The proposed method again performs well with ten percent calibration data. The reach estimates based on the ten percent calibration data are very close to those from the calibration based on the entire data set. Additionally, the reach estimates from the naive approaches are significantly below both, while the benchmark greedy algorithm provides a more sophisticated, but still not optimal, middle-level solution. Compared with the proposed method, the benchmark greedy algorithm generally bids on more websites with a lower average bid per website, but the overall categories and trends remain similar. Table A1 in Web Appendix I provides further details on reach comparisons between the proposed and the benchmark greedy methods at various percentiles across the 100 runs at selected budget points. As expected, the proposed approach outperforms the benchmark greedy method across all 100 runs, even at a confidence level of 99.9%. One interesting observation is that the equal allocation across the lowest average CPM website categories outperforms the other naive methods. This is not too surprising given our method chose many of these same websites; these categories do reach a majority of users as compared to other benchmark equal allocation methods.

**Figure 6: McRib Campaign, Maximizing Reach (calculated using 90% holdout data)**

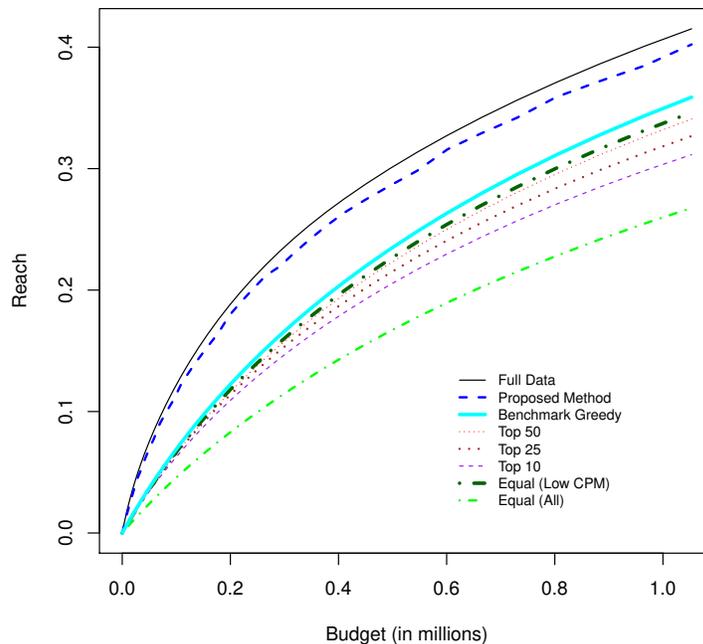
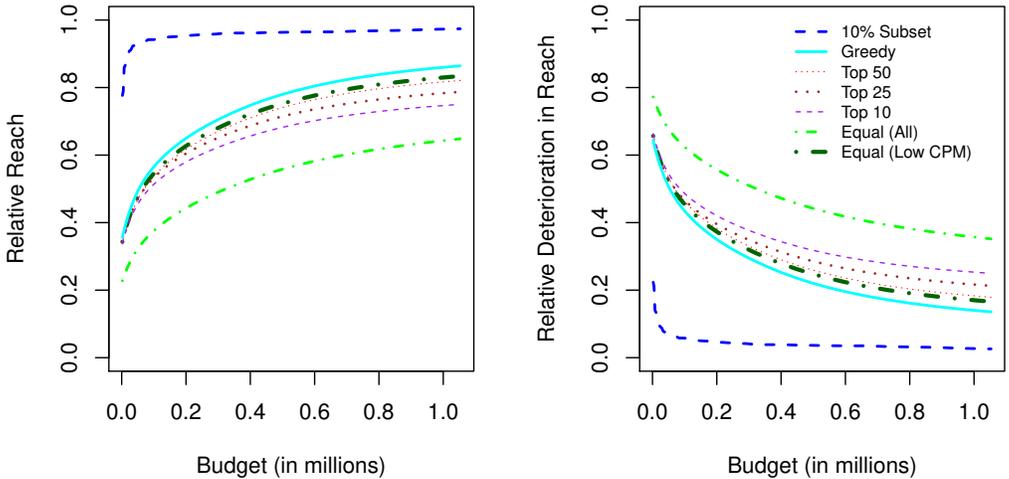


Figure 7 compares relative reach for the proposed method, as well as each of the benchmark methods, relative to the full data set reach. The left-hand plot shows the reach of each method (i.e., estimated reach) divided by the maximum reach value obtained from the full data calibration (i.e., full reach) i.e., estimated reach/full reach. Here, 1.0 would be an exact match. As the plot shows, the proposed method easily outpaces its competitors, reaching an asymptote of 1.0. The right-hand plot shows the relative deterioration in achieved reach, i.e., (full reach - estimated reach)/full reach. A perfect match would thus mean a relative deterioration of zero, which only the proposed method approaches.

**Figure 7: McRib Campaign, Relative Reach and Deterioration in Reach Estimates**



*McRib Campaign with Targeted Consumer Demographics.* In practice, companies often have specific target demographics in mind when running online display ads. In this subsection we demonstrate how our method can be readily modified to accommodate such needs. We consider a setting involving two demographic variables (children and income level), because McDonald’s has historically targeted families with children (Intel, 2014) and fast food in general tends to target lower-income households (Drewnowski and Darmon, 2005). Hence, we illustrate our approach in a scenario where the McRib campaign has identified a key target demographic: lower-income households with children.

We assume an ad served to our target demographic costs twice that of the same ad served to other viewers. For example, if an ad on a Community website has an average CPM of

\$2.10, an ad served to a target demographic member from this site would have a CPM of \$4.20 on average. We further assume McDonald’s wishes to allocate 80% of the total advertising budget to lower income families with children, while reserving the remaining budget for all other users. Following the procedure outlined in the Methodology Section, the method is run twice: first, on the target demographic, using the higher cost curves; and second, on the remaining users, using the original cost curves. These metrics are then combined to create the values in Tables 3 and 4 (Targeted Consumers), where “Chosen Websites” indicates the total number of sites chosen in a given category by either optimization, and “Average Bid Price” likewise reflects the average bid including both groups.

In this example, targeting these demographics does not drastically change the types of websites chosen. Families with children and lower-income households did not represent a significant deviation from the overall population in terms of their Internet browsing behavior. However, we do observe some interesting changes. Most notably Tables 3 and 4 show a marked increase in the number of websites targeted, and the average bid price, within the Gaming and Entertainment categories. Most of the gaming websites contain online flash-based games, which primarily target young players (360i, 2008). Hence, it is likely that proportionally more of the targeted family-based consumers frequently visit such sites. Similar patterns can be observed in the entertainment websites, with the target demographic visiting the Entertainment category websites at a slightly higher rate than the general population.

*McRib Campaign with Target Frequency of Ad Exposure.* In this subsection we demonstrate a case in which McDonald’s wishes to allocate its ad budget such that each individual is exposed to the ad no more than three times during the course of the McRib campaign. For simplicity, we use the data set without demographic targeting, although both approaches could readily be used together. In this case, the “effective frequency” is the value of the function  $e^{-\gamma}(\gamma + \frac{1}{2}\gamma^2 + \frac{1}{6}\gamma^3)$ .

Again, Tables 3 and 4 show the optimization allocation across website categories for this extension (targeted frequency). In general, in this setting, our method bids at lower prices across more website categories, reducing the probability that an ad will appear to a particular

viewer more than three times. For example, more Gaming websites are selected because these sites have many repeat visitors, but low visitation correlation within the Gaming category. Our approach chooses to bid lower prices at more Gaming sites, which gives consumers a low probability of being served the ad on any particular visit but will ultimately reach different consumers with each ad appearance. Similarly, we choose more Entertainment website impressions in the high budget example. Although this category is more expensive than others, we observe low repeat visitation for Entertainment websites. The websites seem to be more universally visited, so advertising on an Entertainment website results in more different people being served the ad. Overall, the method less often includes websites with high repeat visitation, ensuring that a consumer is not served the ad more times than desired.

### ***Case Study 2: Norwegian Cruise Lines Wave Season Online Advertising Campaign***

Each January, the cruise industry begins advertising for its annual “wave season,” which runs from January through March. Norwegian Cruise Lines (NCL) is among the cruise operators that participate heavily in wave season (Satchell, 2011). Consumers who are interested in booking a cruise often use travel aggregation sites, like Orbitz and Priceline, to compare offerings across multiple cruise lines. Hence, we demonstrate the extension in which NCL wants to allocate a minimum amount of budget to a set of major aggregate travel websites. While this is a hypothetical example, it is realistic and can be readily applied to similar scenarios.

Our method handles such scenarios using the extension described in the Methodology Section. Imagine NCL wants to allocate a certain percentage of its advertising budget to eight major aggregate websites (CheapTickets.com, Expedia.com, Hotwire.com, Kayak.com, Orbitz.com, Priceline.com, Travelocity.com, and TripAdvisor.com). We consider two scenarios: first, where our method is restricted to place at least 10% of the budget at each of these eight sites (a total of 80% of the budget going to the aggregate sites), and second, where we place at least 2.5% of the budget at each of these eight sites (a total of 20% of the budget going to the aggregate sites). We follow the same procedure as in the previous case

study to obtain the 500 most visited websites in January 2011 that supported online display advertisements. 48,628 users visited at least one of these 500 websites during January 2011, meaning our  $Z$  matrix is 48,628 by 500. We again divide this data into a 10% subset (4,863 users) of calibration data and use the remaining 90% as holdout data.<sup>17</sup>

Figure 8 demonstrates our reach curves under this extension, with the constrained optimization (mandatory media coverage of aggregate travel sites) in dash blue, and the unconstrained optimization in solid black. We also include reach curves from equal allocation across the eight aggregate sites, allocation across the eight aggregate sites based on the Danaher approach, assuming fixed costs, as demonstrated in Web Appendix A, and the benchmark greedy algorithm as described previously (first, with 80% of the budget allocated to the aggregate travel websites and second with 20%). Figure 8 shows these allocations with the 80% allocation in the top row and the 20% allocation in the bottom row.

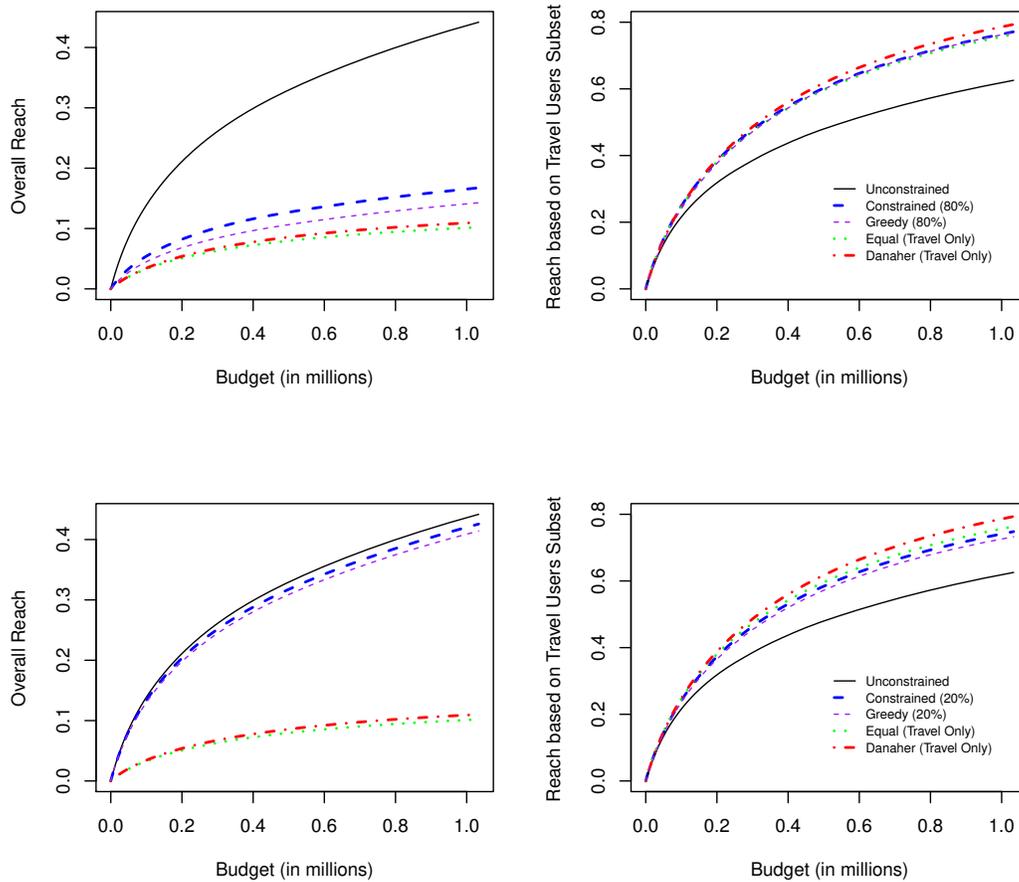
The left hand plots show the overall reach on the holdout data. As expected, the unconstrained curve outperforms the constrained curve, since we cannot improve overall reach by constraining our optimization. As observed previously, the benchmark greedy algorithm performed worse than the constrained method but better than the other benchmarks. As expected, the reach when allocating 20% of the budget to the aggregate sites is much closer to the unconstrained optimization than is the 80% allocation. When the constrained and the benchmark greedy methods are forced to allocate 80% of their budget to the eight travel sites, which only comprise a small proportion of Internet users, their reach performances over the entire Internet population are significantly reduced. Similarly, because the equal allocation and Danaher approaches allocate the entire budget to the eight travel websites, their reach to all Internet users is also naturally limited.

The right hand plots show the reach for the subset of users (6,431) who visited at least one of the eight aggregate travel websites in January 2011. Presumably, these consumers are most likely to be searching for travel deals. In this case, the unconstrained method performs worse than all other approaches. This is expected given all other methods are forced to allocate high budgets to the travel websites. Among them, Danaher’s method performs

---

<sup>17</sup>We omit the website category makeup description of this application due to its similarity to Table 2. It is available from authors upon request.

**Figure 8: Overall Reach with Mandatory Coverage in Aggregate Travel Sites (left) and on the subset of users who visited aggregate travel sites (right)**



slightly better than the rest. Again, this is expected given that Danaher’s approach allocates 100% of budget to these sites.

It is interesting to note that the constrained method only has a very small reduction in reach among the travel population in moving from 80% to 20% allocation, but a large increase in overall reach. Overall, the 20% constrained optimization appears to provide a nice compromise, since general users may still view the ad, but NCL can also be confident they have reached a high fraction of the people most likely to book a cruise.

## CONCLUSION

In the current advertising climate, firms need an online presence more than ever. Nevertheless, the ever-increasing number of websites presents not only endless opportunities but also tremendous challenges for firms' online display ad campaigns. While online advertising is limited only by the sheer number of websites, optimal Internet media selection among thousands of websites presents a prohibitively challenging task, particularly in the presence of programmatic advertising and real-time bidding. This has led to the inevitable rise of large-scale bidding using software such as DSPs.

One significant component of this framework, and the one we focus on, is ad-buying through the use of private auctions, which give both publishers and advertisers control over ad placement and inventory. Here, the websites on which ads may be placed is known in advance, generally to ensure reputability. Advertisers must still bid on ad impressions at these websites, but the pool of potential ad spaces is limited. We develop a method which leverages bid landscape procedures with both the marketing and statistics literature to provide a tool which sets DSP bidding guidelines while simultaneously providing ways to check the efficacy of the established bidding procedures and update as necessary.

Further, the method is designed to be run ahead of a campaign, meaning it can function as a budget-setting tool. To demonstrate the applicability and scalability of our algorithm in real-world settings, we consider two case studies using the comScore data. We also illustrate that the proposed method extends easily to accommodate common practical Internet advertising considerations, including consumer targeting, target frequency of ad exposures, and mandatory media coverage to matched content websites. Consequently, the proposed method provides a transparent bid- and budget-setting tool, with great flexibility for a range of online display advertising campaigns. As a result, we believe that this research has considerable value to both marketing academia and industry practitioners.

Our work also offers some promising avenues for further research. For example, while the proposed method emphasizes strategic bidding guidelines at the publisher (and/or consumer segment) level, future research may refine our approach by accounting for viewership

correlations at the impression level and developing optimal bidding strategies for individual impressions. In practice, this might become unwieldy for a large number of Internet users but this level of personalization would further improve campaign performance. Additionally, we currently consider the perspective of an individual firm that wishes to maximize reach for its particular campaign. This method could be further extended for use by an advertising broker who wishes to maximize reach over a set of clients. Thus, an interesting extension of our method would be to maximize over multiple campaigns from the perspective of an advertising agency. Lastly, future research may examine how to incorporate other important aspects of Internet display ads such as ad blocking (e.g., WSJ 2015, 2016), advertising spillovers (e.g., Lewis and Nguyen 2015; Sahni 2016), and temporal spacing between ad exposures (e.g., Sahni 2015) into our extant framework.

## References

- 360i (2008), Point of View on Gaming, Technical report.
- AdAge (2012), ‘Can the McRib Save Christmas?’. URL: <http://adage.com/article/news/mcrib-save-christmas/237233/>. Accessed 25 November 2016.
- Adelphic (2015), ‘Are Media Buyers Using the Best Decision Algorithms?’. URL: <http://www.adelphic.com/2015/03/media-buyers-using-best-decision-algorithms/> Accessed 12 July 2016.
- AdWords (2016), ‘Bid Landscapes’, Online. URL: <http://venturebeat.com/2012/07/23/real-time-bidding-is-the-next-mobile-ad-breakthrough-heres-how-you-can-profit/>. Accessed 11 Jan 2016.
- Breheny, Patrick and Jian Huang (2011), ‘Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection’, *The Annals of Applied Statistics* **5**(1), 232–253.
- Candes, Emmanuel and Terence Tao (2007), ‘The Dantzig Selector: Statistical Estimation When  $p$  is Much Larger than  $n$ ’, *The Annals of Statistics* **35**(6), 2313–2351.
- Cui, Ying, Ruofei Zhang, Wei Li and Jianchang Mao (2011), Bid Landscape Forecasting in Online Ad Exchange Marketplace, in ‘Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, ACM, pp. 265–273.
- Danaher, Peter J. (2007), ‘Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction’, *Marketing Science* **26**(3), 422–437.
- Danaher, Peter J. (2008), *Advertising Models*, Springer US, Boston, MA, pp. 81–106.
- Danaher, Peter J., Janghyuk Lee and Laouchine Kerbache (2010), ‘Optimal Internet Media Selection’, *Marketing Science* **29**(2), 336–347.
- Danaher, Peter J. and Michael Smith (2011), ‘Modeling Multivariate Distributions Using Copulas: Applications in Marketing’, *Marketing Science* **30**(1), 1–18.
- DNB (2015), ‘Understanding the Programmatic Ecosystem’. URL: <http://www.dnb.com/perspectives/marketing-sales/understand-programmatic-ecosystem.html> Accessed 2 November 2016.
- Drewnowski, Adam and Nicole Darmon (2005), ‘Food Choices and Diet Costs: an Economic Analysis’, *The Journal of Nutrition* **135**(4), 900–904.
- Efron, Bradley, Trevor Hastie, Iain Johnston and Robert Tibshirani (2004), ‘Least Angle Regression (with discussion)’, *The Annals of Statistics* **32**(2), 407–451.
- eMarketer (2012), ‘Digital Ad Spending Tops 37 Billion’. URL: <http://www.emarketer.com/newsroom/index.php/digital-ad-spending-top-37-billion-2012-market-consolidates>. Accessed 4 Jun 2015.
- eMarketer (2013), ‘Is the Future of Programmatic Premium?’. URL: <http://www.emarketer.com/Article/Future-of-Programmatic-Premium/1010430> Accessed 2 November 2016.
- Fan, Jianqing and Runze Li (2001), ‘Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties’, *Journal of the American Statistical Association* **96**(456), 1348–1360.

- Feldman, Jon and S. Muthukrishnan (2008), *Algorithmic Methods for Sponsored Search Advertising*, Springer, chapter 4, pp. 91–124.
- Friedman, Jerome, Trevor Hastie and Robert Tibshirani (2010), ‘Regularization Paths for Generalized Linear Models via Coordinate Descent’, *Journal of Statistical Software* **33**(1), 302–332.
- G2Crowd (2016), ‘Demand Side Platform (DSP)’, Online. URL: <https://www.g2crowd.com/categories/demand-side-platform-dsp> Accessed 17 May 2016.
- Goeman, Jelle (2010), ‘L1 Penalized Estimation in the Cox Proportional Hazards Model’, *Biometrical Journal* **52**(1), 70–84.
- Hatano, Daisuke, Takuro Fukunaga, Takanori Maehara and Ken ichi Kawarabayashi (2015), Lagrangian Decomposition Algorithm for Allocating Marketing Channels, in ‘Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence’, pp. 1144–1150.
- HBR (2015), ‘Is Programmatic Advertising the Future of Marketing?’. URL: <https://hbr.org/2015/06/is-programmatic-advertising-the-future-of-marketing> Accessed 2 November 2016.
- Hesterberg, Tim, Nam Hee Choi, Lukas Meier and Chris Fraley (2008), ‘Least Angle and L1 Penalized Regression: A Review’, *Statistics Surveys* **2**, 6193.
- Hoban, Paul R. and Randolph E. Bucklin (2015), ‘Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment’, *Journal of Marketing Research* **LII**, 375–393.
- iThinkMedia (2016), ‘CPC Bidding vs. CPM Bidding: Whats The Difference?’. URL: <https://www.ithinkmedia.co.uk/blog/cpm-cpc-bidding/>. Accessed 29 November 2016.
- Iyer, Krishnamurthy, Ramesh Johari and Mukund Sundararajan (2011), ‘Mean Field Equilibria of Dynamic Auctions with Learning’, *ACM SIGecom Exchanges* **10**(3), 10–14.
- Krugman, Herbert E. (1972), ‘Why Three Exposures May Be Enough’, *Journal of Advertising Research* **12**(6), 11–14.
- Lewis, David D. and William A. Gale (1994), *A Sequential Algorithm for Training Text Classifiers*, Springer-Verlag, pp. 3–12.
- Lewis, Randall and Dan Nguyen (2015), ‘Display Advertisings Competitive Spillovers to Consumer Search’, *Quantitative Marketing and Economics* **13**, 93–115.
- Liaukonyte, Jura, Thales Teixeira and Kenneth C. Wilbur (2015), ‘Television Advertising and Online Shopping’, *Marketing Science* **34**(3), 311–330.
- Lipsman, Andrew (2010), *The New York Times Ranks as Top Online Newspaper According to May 2010 U.S. comScore Media Metrix data*, Technical report, comScore, Inc.
- Luo, Zhi-Quan and Paul Tseng (1992), ‘On the Convergence of the Coordinate Descent Method for Convex Differentiable Minimization’, *Journal of Optimization Theory and Applications* **72**(1), 7–35.
- Manchanda, Puneet, Jean-Pierre Dub, Kim Yong Goh and Pradeep K. Chintagunta (2006), ‘The Effect of Banner Advertising on Internet Purchasing’, *Journal of Marketing Research* **43**, 98–108.
- MarketingLand (2013), ‘5 Common Myths About Real-Time Bidding’. URL: <http://marketingland.com/5-common-myths-about-real-time-bidding-58708> Accessed 2 November 2016.
- Meinshausen, Nicolai (2007), ‘Relaxed Lasso’, *Computational Statistics and Data Analysis* **52**, 374–393.
- Mintel (2014), *Kids as Influencers–U.S.*, Technical report, Mintel.
- Miyuchi, Atsushi, Yuni Iwamasa, Takuro Fukunaga and Naonori Kakimura (2015), Threshold Influence Model for Allocating Advertising Budgets, in ‘Proceedings of the 32nd International Conference on Machine Learning’, Vol. 37, pp. 1395–1404.

- Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C. Liechty (2004), ‘Modeling Online Browsing and Path Analysis Using Clickstream Data’, *Marketing Science* **23**(4), 579–595.
- Naples, Michael J. (1979), *Effective Frequency: The Relationship Between Frequency and Advertising Effectiveness*, Technical report, Association of National Advertisers, New York.
- Park, Young-Hoon and Peter S. Fader (2004), ‘Modeling Browsing Behavior at Multiple Websites’, *Marketing Science* **23**(3), 280–303.
- Parsons, Lance, Ehtesham Haque and Huan Liu (2004), ‘Subspace Clustering for High Dimensional Data: A Review’, *ACM SIGKDD Explorations Newsletter* **6**(1), 90–105.
- PromiseMedia (2015), ‘Online Advertising Models: CPC, CPM or CPA?’. URL: <http://www.promisemedia.com/online-advertising/best-revenue-deals-cpm-cpc-or-cpa>. Accessed 29 November 2016.
- Radchenko, Peter and Gareth James (2008), ‘Variable Inclusion and Shrinkage Algorithms’, *Journal of the American Statistical Association* **103**(483), 1304–1315.
- Rosset, Saharon and Ji Zhu (2007), ‘Piecewise Linear Regularized Solution Paths’, *The Annals of Statistics* **35**(3), 1012–1030.
- Sahni, Navdeep S. (2015), ‘Effect of Temporal Spacing between Advertising Exposures: Evidence from Online Field Experiments’, *Quantitative Marketing and Economics* **13**(3), 203–247.
- Sahni, Navdeep S. (2016), ‘Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising’, *Journal of Marketing Research* **LIII**, 459–478.
- Satchell, Arlene (2011), ‘Norwegian: Cruise Fares to Increase up to 10 Percent April 1’, *South Florida Sun-Sentinel*.
- Schmidt, Mark, Glenn Fung and Rmer Rosales (2007), ‘Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches’, *Machine Learning: ECML 2007* **4701**, 286–297.
- Tibshirani, Robert (1996), ‘Regression Shrinkage and Selection via the LASSO’, *Journal of the Royal Statistical Society, Series B* **58**, 267–288.
- VentureBeat (2012), ‘Real-Time Bidding is the Next Mobile Ad Breakthrough Heres How You Can Profit’, Online. URL: <http://venturebeat.com/2012/07/23/real-time-bidding-is-the-next-mobile-ad-breakthrough-heres-how-you-can-profit/>. Accessed 10 Jul 2016.
- Wang, Jun, Shuai Yuan and Weinan Zhang (2016), ECIR 2016 Tutorial: Real-Time Bidding Based Display Advertising: Mechanisms and Algorithms, ECIR.
- WSJ (2015), ‘Ad Blocking is a Hot Topic of Marketing, Media Executives’. URL: <https://www.wsj.com/articles/ad-blocking-is-a-hot-topic-for-marketing-media-executives-1443259981> Accessed: 12 August 2017.
- WSJ (2016), ‘The Rise of the Anti-Ad blockers’. URL: <https://www.wsj.com/articles/the-rise-of-the-anti-ad-blockers-1465805039>. Accessed 12 August 2017.
- Wu, Tong Tong and Kenneth Lange (2008), ‘Coordinate Descent Algorithms for Lasso Penalized Regression’, *The Annals of Applied Statistics* **2**(1), 224–244.
- Yingwei, Lu, N. Sundararajan and P. Saratchandran (1997), ‘A Sequential Learning Scheme for Function Approximation Using Minimal Radial Basis Function Neural Networks’, *Neural Computation* **9**, 461–478.
- Zhang, Haifeng, Yevgeniy Vorobeychik and Ariel D. Procaccia (2017), Multi-Channel Marketing with Budget Complementarities, in ‘AAMAS ’17 Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems’, pp. 1232–1240.
- Zhang, Weinan and Jun Wang (2015), ‘Statistical Arbitrage Mining for Display Advertising’, *ACM KDD ’15*.
- Zhang, Weinan, Shuai Yuan and Jun Wang (2014), ‘Optimal Real-Time Bidding for Display Advertising’, *ACM KDD ’14* pp. 1077–1086.

- Zhang, Zheng, Scott Schwartz, Lukas Wagner and Webb Miller (2000), ‘A Greedy Algorithm for Aligning DNA Sequences’, *Journal of Computational Biology* **7**(1/2), 203–214.
- Zhao, Peng, Guilherme Rocha and Bin Yu (2009), ‘The Composite Absolute Penalties Family for Grouped and Hierarchical Variance Selection’, *The Annals of Statistics* **37**(6A), 3468–3497.
- Zou, Hui (2006), ‘The Adaptive LASSO and its Oracle Properties’, *Journal of the American Statistical Association* **101**, 1418–1429.
- Zou, Hui and Trevor Hastie (2005), ‘Regularization and Variable Selection via the Elastic Net’, *Journal of the Royal Statistical Society, Series B* **67**, 301–320.