Donald R. Lehmann Award, 2019; Paul E. Green Award Finalist, 2019

COURTNEY PAULSON, LAN LUO, and GARETH M. JAMES*

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 article is among the first efforts in marketing to develop a nonproprietary 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, this method explicitly accounts for viewership correlations across websites. Consequently, campaign managers can make optimal bidding decisions over the entire set of advertising opportunities. More Importantly, they can avoid overbidding for impressions from high-cost publishers, unless such sites reach an otherwise unreachable audience. The proposed method can also be used as a budget-setting tool, because it readily provides optimal bidding guidelines for a range of campaign budgets. Finally, this method 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, real-time bidding, constrained convex optimization

Online Supplement: http://dx.doi.org/10.1509/jmr.15.0307

Efficient Large-Scale Internet Media Selection Optimization for Online Display Advertising

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, of which Internet display advertising accounted for 40%. This percentage is expected to continue to grow, outpacing paid search ad spending (eMarketer 2012). This increasing trend in Internet display advertising is related to the wide range of benefits this advertising format offers, including building awareness and recognition, forming attitudes, and generating direct responses such as website visits and downstream purchases (Danaher, Lee, and Kerbache 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 (Rayport 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 (Rayport 2015; Vidakovic 2013).

Despite the increasing popularity of Internet display ads and the rise of programmatic advertising, very few marketing articles have examined optimal budget allocation of Internet displays ads within the context of programmatic advertising. To date, Danaher, Lee, and Kerbache's (2010) Sarmanov-based model was among the first and most successful attempts to do so in the nonprogrammatic setting. This method has proved to work well for budget allocation in settings on the order of ten websites, under the assumption that the cost to advertise at each website (i.e., cost

*Courtney Paulson is Assistant Professor, Robert H. Smith School of Business, University of Maryland (email: cpaulson@rhsmith.umd.edu). Lan Luo is Associate Professor, Marshall School of Business, University of Southern California (email: lluo@marshall.usc.edu). Gareth M. James is Professor and E. Morgan Stanley Chair in Business Administration, Marshall School of Business, University of Southern California (email: gareth@marshall.usc.edu). Peter Danaher served as associate editor for this article.

ISSN: 0022-2437 (print)
1547-7193 (electronic)

Journal of Marketing Research
Vol. LV (August 2018), 489–506
DOI: 10.1509/jmr.15.0307

per thousand impressions [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.

This article is among the first efforts in marketing to develop a nonproprietary 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 (GRPs; 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 real-time bidding (RTB) currently comprises approximately 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.[1]

Under RTB, when an ad space becomes available, the ad agency or demand-side platform (DSP) has only milliseconds to place a bid (Downey 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. Drawing 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 (Adelphic 2015; G2Crowd 2016; Wang, Yuan, and Zhang 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).

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 online viewership of *Bloomberg Businessweek* and *Reuters*. In such cases, firms can benefit greatly from strategically placing more bids for impressions at the more cost-effective website of the two, because both attract largely 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 optimizing budget allocations over a large number of websites is so difficult is that the problem of choosing the subset of websites on which to bid is generally nondeterministic polynomial-time hard (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. Although 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 articles include the least absolute shrinkage and selection operator (Lasso; Tibshirani 1996), the smoothly clipped absolute deviation (SCAD; Fan and Li 2001), the elastic net (Zou and Hastie 2005), the adaptive Lasso (Zou 2006), composite absolute penalties (CAPs; Zhao, Rocha, and Yu 2009), the Dantzig selector (Candes and Tao 2007), the relaxed Lasso (Meinshausen 2007), and variable inclusion and shrinkage algorithms (VISAs; 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 (AdWords API 2016; Cui et al. 2011; Iyer, Johari, and Sundararajan 2011; Wang, Yuan, and Zhang 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. Thus, we believe that the proposed optimization method will be of considerable value in Internet display ad campaigns.

The remainder of the article is structured as follows. First, we discuss the Internet display ad market and the real-world

[1]The U.S. display ad market is expected to be 17% nonprogrammatic, 31% programmatic direct, and 52% programmatic with RTB in 2017 (eMarketer 2013). In Web Appendix A, we outline 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, Lee, and Kerbache 2010) in this appendix.

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 Line (NCL) Wave Season advertising campaign) using 2011 comScore Media Metrix data to demonstrate the proposed method in a real-world setting. Finally, we conclude with a summary of our findings, contributions, and avenues for future work.

INSTITUTIONAL SETTING AND MANAGERIAL USAGE

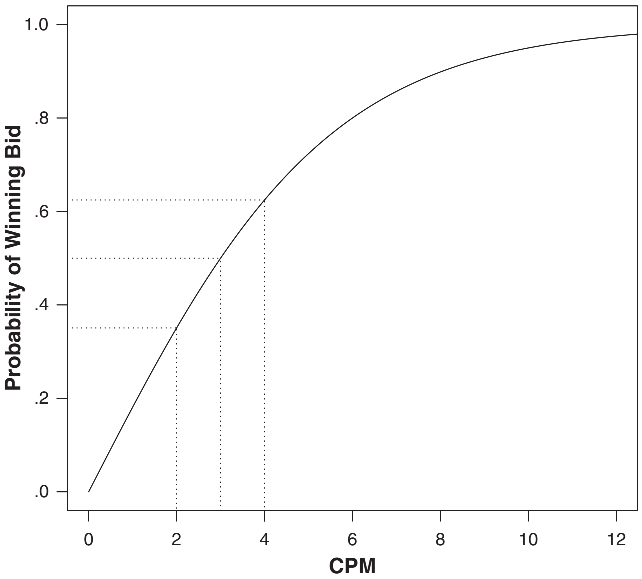
Programmatic Advertising with RTB for Internet Display Ads

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 (cost per click [CPC] buying) being the most popular for search ad campaigns (Bateman 2015; Mugridge 2016). Although a small percentage of advertisers are still pushing for pay-per-acquisition (cost 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 (Bateman 2015).

The primary focus of this article is on Internet display ads, so we emphasize pay-per-impression campaigns in which the KPIs are gauged by exposure-based metrics such as reach, frequency, and GRPs.[2] 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. In addition, we focus on the scenario of private marketplace auctions (the most popular form of programmatic auction) in which only select websites and advertisers are allowed to engage in the auction (Goldberg 2015).[3] 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 API 2016; Iyer, Johari, and Sundararajan 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.[4] Some advertising agencies even supply "bid simulators," giving advertisers an interactive tool to see how changing their bids affects their returns (AdWords API 2016). Furthermore, 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.

Figure 1
EXAMPLE COST CURVE



Drawing 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 showing that the DSP has a 50% probability of winning 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 the bid and the expected number of page views at this website. In the previous example, with a bid of \$3, the DSP can expect to win 5,000 bids if 10,000 impressions were available.

In practice, bid landscapes might fluctuate over time, possibly because of 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

Next, 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 previously, a DSP has only milliseconds to determine whether and how much to bid under RTB. Therefore, we suggest, like many extant DSP algorithms, that the campaign manager run the proposed method prior to the ad campaign to determine optimal bidding prices for each website (including placing zero bids on certain websites). As we have discussed, cost curves may fluctuate over time, but this can be easily handled by regularly rerunning the optimization, as is common for other DSP algorithms.

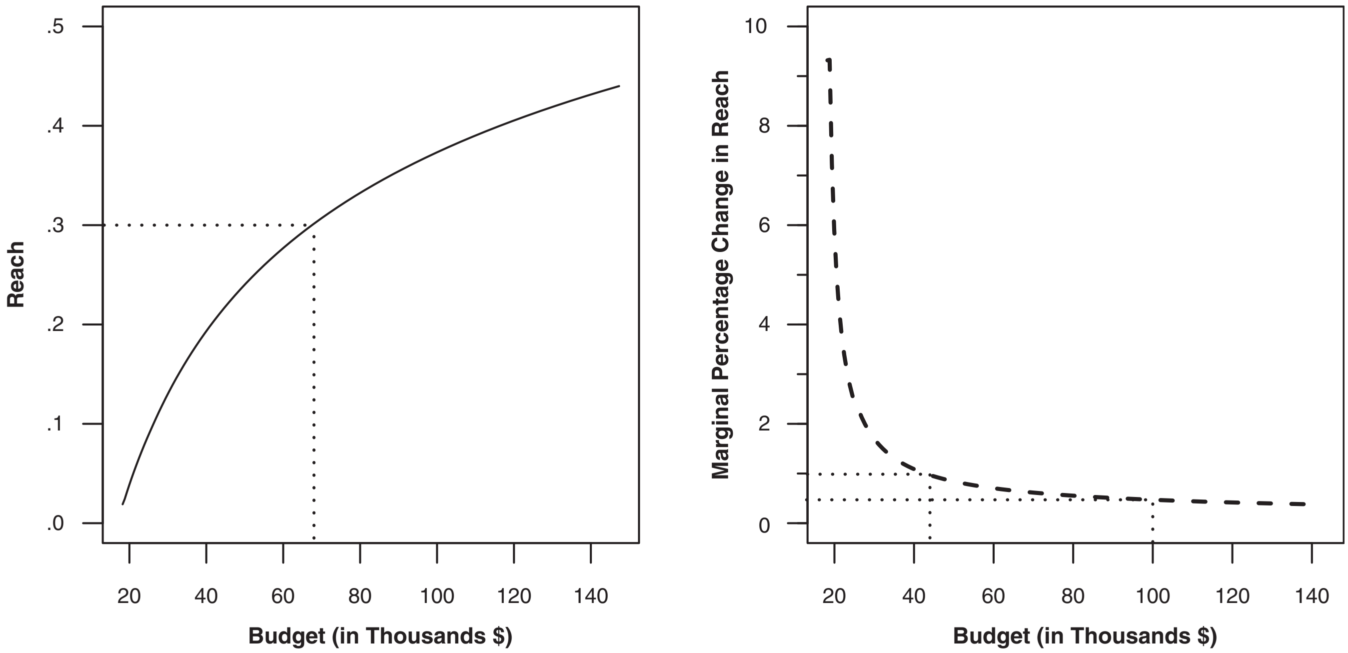
---

[2]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.

[3]A key advantage of private marketplace auctions is that advertisers do not have to worry about their ads popping up on less reputable sites, and publishers can also assert some control over the quality of advertisers that can place ads on their websites.

[4]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 to track how often bids are accepted at those websites at a variety of price points.

Figure 2
REACH AND THE CORRESPONDING MARGINAL PERCENTAGE CHANGE IN REACH



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 .11 seconds per budget calculation. The average per-budget time increases to .49 seconds with a maximum budget of $250,000 and 2.56 seconds for a maximum budget of $1 million.[5] Thus, for most realistic budget allocations, campaign managers can regularly recalibrate the optimization, if desired. 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, Sundararajan, and Saratchandran 1997). Campaign managers can update cost curves on the basis of recently accepted and rejected bids at each website, which in turn updates the bidding guidelines once the method is rerun. In addition, 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.

Although for illustrative purposes, we currently demonstrate our method using shifted logistic cost curves (Figure 1), we note that 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, because bid landscapes are frequently smoothed (Wang, Yuan, and Zhang 2016) and practical bidding data often result in concave bid landscapes (Feldman and Muthukrishnan 2008; Zhang and Wang 2015; Zhang, Yuan, and Wang 2014).

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 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 .5%—that is, as the budget increases from $100,000 to $101,000 (a 1% increase), the percentage change in reach is (37.5% – 37.3%)/37.3% = .5%.

Campaign managers can use such curves as a budget-setting tool. For example, the firm's managers may have originally designated $60,000 to reach 30% of their customers but discover that a budget of $70,000 is required (left panel of Figure 2) and adjust accordingly. In addition, 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 that 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.

[5]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 article. 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 nonzero websites, each of which requires more effort to determine the exact allocation.

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, 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. Subsequently, 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 $j = 1, \ldots, p$. Thus, $Y = \sum_{j=1}^{p} X_j$ 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 1,000 impressions). An example of such a cost curve appears in Figure 2. In addition, let τ_j represent the number of available impressions (in thousands) at the jth website during the campaign period. Then, the expected total number of ads served at the jth website will be $\tau_j s_j$ thousand at a total cost of $c_j \tau_j s_j$ dollars.[6]

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. Thus, our question is equivalent to the following optimization problem:

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

where $\mathbf{c} = (c_1, \ldots, c_p)$ 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. In addition, 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 next.

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 $\mathbf{Z} = (Z_1, \ldots, Z_p)$, 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 through the following:

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

where $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})$ are the page views at websites $j = 1, \ldots, p$, for the ith 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.[7]

The decomposition shown previously 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, because a customer who visits website j may 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 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, $k \ne j$. For example, again consider the websites of *Bloomberg Businessweek* and Reuters. 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 $s_1 = .5$ and $s_2 = .8$, then X_1 and X_2 become 10 and 15 independent random coin flips with respective probability of being served the ad at .5 and .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).

Thus, we model X_j, conditional on z_{ij} and c_j, as independent Poisson random variables—that is, $X_j | z_{ij}, c_j \sim \text{Pois}(\gamma_{ij})$, or equivalently,

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

The expected number of ad appearances γ_{ij} is given by the probability of an ad served on a random page view at the jth 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 web pages at that site, $\gamma_{ij} = .2 \times 10 = 2$. 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 γ_{ij}. Namely, the γ_{ij} terms are unconditionally correlated but are conditionally (on $\mathbf{z}_i$ and $\mathbf{c}$) independent.

[6]Note that although s_j depends on c_j, for notational simplicity, we refer to the $s_j(c_j)$ function as s_j.

[7]comScore data aggregate page views by domain names due to privacy considerations. For example, if a user reads two articles on Yahoo Sports, then reads three Yahoo News articles, comScore would enter the session as one visit to Yahoo.com with five web pages viewed. Because we use comScore data in our empirical investigation, we follow methodology proposed by Danaher (Danaher 2007; Danaher, Lee, and Kerbache 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 previous Yahoo example were logged as two page views at Yahoo Sports and three page views at Yahoo News), because of its computational efficiency, our method can also readily accommodate within-website correlation by treating each subdomain page view as a separate column in the $\mathbf{z}_i$ matrix.

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

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

and, in particular, $P(Y = 0|\mathbf{Z} = \mathbf{z}_i, \mathbf{c}) = e^{-\gamma_i}$.[8] For ease of exposition, we present our method with γ_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 in which γ_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:

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

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 it is 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 multidimensional rhomboid (i.e., it has many "sharp points" on the axes where $c_j = 0$). Thus, the solution within the constraint region that minimizes the optimization function often falls on one of these points in a similar way to Lasso (for further intuition of this phenomenon, see Web Appendix B). As budget B decreases, the constraint becomes confined to a smaller region around the origin and, thus, all the c_{js} are shrunk toward zero, producing a sparse solution. Therefore, although our optimization function is not the 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 $s_1, ..., s_p$ are concave (for proof, see Web Appendix B). This in turn makes our method highly computationally efficient. Finally, as discussed previously, our optimization function directly incorporates correlations among viewerships. In particular, we show in Web 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 reexpress it as an unconstrained optimization through a Lagrangian formulation. We take that approach with Equation 5, reformulating it as

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

where $\lambda > 0$ is the Lagrangian multiplier.[9] Note λ 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 λ. For a given number of websites, as budget increases, λ decreases, and the algorithm allocates more budget to more websites. As budget decreases, λ 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 because τ_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. Thus, optimizing over w_j will also solve for c_j. Using this change of variable, Equation 6 can be simplified as follows:

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

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, Hastie, and Tibshirani 2010; Goeman 2010; Hesterberg et al. 2008; Rosset and Zhu 2007; Schmidt, Fung, and Rosales 2007). As a result, there exist highly efficient algorithms for solving such problems. In this article, we utilize one such algorithm known as coordinate descent to solve Equation 7 over a grid of values for λ, 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., Breheny and Huang 2011; Friedman, Hastie, and Tibshirani 2010; Wu and Lange 2008).

Essentially, the coordinate descent algorithm simplifies our optimization to a series of one-dimensional optimizations as described in Algorithm 1 (for more details of the algorithm, see Web Appendix B). 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 found quickly by initializing $\mathbf{w}_0$ in Algorithm 1 (see Table 1) with the allocation from the previous budget. Because 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

[8]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 in which the Poisson distribution is replaced by a negative binomial distribution.

[9]In statistics, this is commonly referred to as a penalized optimization equation. The $(\lambda/n)\sum_j c_j s_j \tau_j$ penalty would frequently be written as an ℓ_1 penalty rather than a summation term. However, for our setup, these two formulations are identical, because we have the condition $c_j \ge 0$ for all j.

Table 1
COORDINATE DESCENT ALGORITHM FOR BUDGET OPTIMIZATION

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

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 $\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_p$ represent the current estimates for **w**. Then, using standard calculations it is not hard to show that Step 3a can be approximately solved by optimizing

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

where $\theta_{ij} = e^{-\tilde{\gamma}_i}s_j'(\tilde{w}_j)z_{ij}$, $\eta_{ij} = e^{-\tilde{\gamma}_i}z_{ij}[z_{ij}s_j'(\tilde{w}_j)^2 - s_j^{s''}(\tilde{w}_j)]$, 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

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

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 (because 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.[10]

Extensions

Consumer targeting. In practice, campaign managers often wish to target particular consumer groups on the basis of 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. Next, we discuss how demographic and behavioral targeting can be readily incorporated into the proposed model.

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 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. Because 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 comprising 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 because 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.

With behavioral targeting, ads are aimed at certain website visitors in line with their past purchase/browsing behavior. For example, 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 behavior-tracking tools, a similar approach could be used to accommodate behavioral targeting. Within this setting, once the high-value consumers are identified (e.g., previous aggregate travel site users), they become a new demographic to target. Assuming it is more costly to retarget a user, these consumers will now have new cost curves compared with users who have not yet been tracked. Instead of each website having a cost curve that holds for all users, all retargeted 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.

Because of the 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 on request.

Alternative exposure-based key performance indicators. Consider the more general form of Equation 1:

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

where $g(\cdot)$ represents a given KPI of interest. So far, we have concentrated on reach—that is, $g(\mathbf{c}) = P(Y > 0|\mathbf{c})$. However, several alternative exposure-based KPIs are commonly used in Internet display ad campaigns. Next, we describe some of the most commonly applied KPIs and how our method can be adapted to these alternative metrics[11]:

- 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 $E_Z\{\gamma/(1 - e^{-\gamma})\}$ which we approximate by $(1/n)\sum_{i=1}^{n}[\gamma_i/(1 - e^{-\gamma_i})]$.

[10]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 whether 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.

[11]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 γ_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, ..., p is known, the optimization can then be carried out in a similar fashion to the approach already described.

- GRPs, frequency multiplied by reach: It is not difficult 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 $(1/n)\sum_{i=1}^{n}\gamma_i$.
- Effective frequency and frequency capping,[12] 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, Lee, and Kerbache 2010). For example, sales conversions and profits from online display ads might be the 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, $g(\mathbf{c}) = P(k_a \le Y \le k_b|\mathbf{c})$, which is given by $E_{\mathbf{Z}}\{\sum_{y=k_a}^{k_b}(e^{-\gamma}\gamma^y/y!)\}$. We can approximate this expectation using $(1/n)\sum_{i=1}^{n}\sum_{y=k_a}^{k_b}e^{-\gamma_i}\gamma_i^y/y!$. So, using the example of $1 \le Y_i \le 3$, our problem would involve maximizing $(1/n)\sum_{i=1}^{n}e^{-\gamma_i}[\gamma_i + (1/2)\gamma_i^2 + (1/6)\gamma_i^3]$.

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 θ_{ij} and η_{ij} in each case. When optimizing frequency, we apply Equations 8 and 9 with

$$\theta_{ij} = \frac{s_j' z_{ij}}{1 - e^{-\gamma_i}} - \frac{s_j s_j' z_{ij}^2 e^{-\gamma i}}{(1 - e^{-\gamma i})^2}, \text{ and}$$

$$\eta_{ij} = \frac{-s_j'' z_{ij}}{1 - e^{-\gamma_i}} + \frac{z_{ij}^2 e^{-\gamma_i}\left[s_j'^2(2 - z_{ij}) + s_j s_j''\right]}{(1 - e^{-\gamma_i})^2} - \frac{2z_{ij}^3 e^{-2\gamma_i} s_j s_j'^2}{(1 - e^{-\gamma_i})^3}.$$

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

$$\theta_{ij} = s_j' z_{ij} e^{-\gamma_i}(1 - \gamma_i^3/6), \text{and}$$
$$\eta_{ij} = z_{ij} e^{-\gamma_i}\left[\gamma_i^2 s_j'^2 z_{ij}\big/2 - \left(s_j'' - z_{ij} s_j'^2\right)(1 - \gamma_i^3/6)\right].$$

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," 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 $w_j \ge \min_j$, 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 nonzero amount. We would then replace Equation 9 with the following (an example of this extension appears in the NCL Wave Season online advertising campaign case study):

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

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., 5,000 websites).[13]

The cost curves, s_j, are an integral part of our methodology. Although, in practice, these curves can be assumed to be known to the campaign manager, in this article we demonstrate our method using a shifted logistic curve (Figure 2). A standard logistic curve follows the equation $s_j(c_j) = e^{c_j a_j + b_j}/(1 + e^{c_j a_j + b_j})$, where c_j is the CPM bid 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. Furthermore, because our cost curves should satisfy the constraint $s_j(0) = 0$, we shift the logistic curve as follows: $s_j(c_j) = 2\{[e^{c_j a_j}/(1 + e^{c_j a_j})] - 1/2\} = (e^{c_j a_j} - 1)/(e^{c_j a_j} + 1)$. 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 $s_j(5) = .5$—that is, $a_j = (1/5)\log 3$.

Conceptual Demonstration with Five Websites

We first consider a five-website setting so that features of our method can be represented graphically. Table 2 gives the average CPM, c_0 (i.e., the value at which there is a 50% chance of winning the bid); expected total number of page views, τ_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 τ_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 CPM bid on the right. Here, Website 1 (black solid) has the lowest c_0 (\$3), through to Website 5 (dashed purple) which has the highest c_0 (\$7). Consider a bid of CPM = \$5. For Website 1 (c_0 = \$3), $s_1(5) = .723$, so a \$5 bid results in approximately $s_1(5)\tau_1 = .723 \times 30 = 21.7$ million impressions bought during the campaign. By contrast, for Website 5 (c_0 = \$7), $s_5(5) = .373$, so a \$5 bid results in only

[12]This term typically refers to the setting in which 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.

[13]For simplicity, the simulation data in both examples are generated independently without correlations. Because the proposed method is designed to leverage correlations across sites, this setup provides a lower bound with respect to advantages from our approach. In addition, we chose to simulate the data set for the 5,000-website example because data cleaning in comScore for this many websites would be prohibitively time consuming.

Table 2
AVERAGE CPM AND EXPECTED NUMBER OF PAGE VIEWS FOR FIVE-WEBSITE EXAMPLE

	Avg. CPM (c_0)	*Page Views (in Millions)*	*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

$s_5(5)\tau_5 = .373 \times 40 = 14.9$ 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 (Δ), and $2,000,000 (⊠). 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 the algorithm allocated the budget across the websites. The location of each symbol represents the CPM bid 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 (Websites 1 and 3). 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: 5,000 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 zero to ten with higher weight on a value of zero. The average CPMs of these websites are randomly generated, chosen from .25 to 8.00 in increments of .25.

We compare the proposed method with 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 25, 50, and 100 of the most-visited websites; and (4) a benchmark greedy algorithm that sequentially allocates budget to one website at a time. To implement approaches 1–3, we select c_j so that $w_j = s_j\tau_jc_j$ 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., Hatano et al. 2015; Miyuchi et al. 2015;

Figure 3
EXAMPLE COST CURVES FOR THE FIVE SIMULATED WEBSITES AND NUMBER OF IMPRESSIONS BOUGHT AT EACH CPM

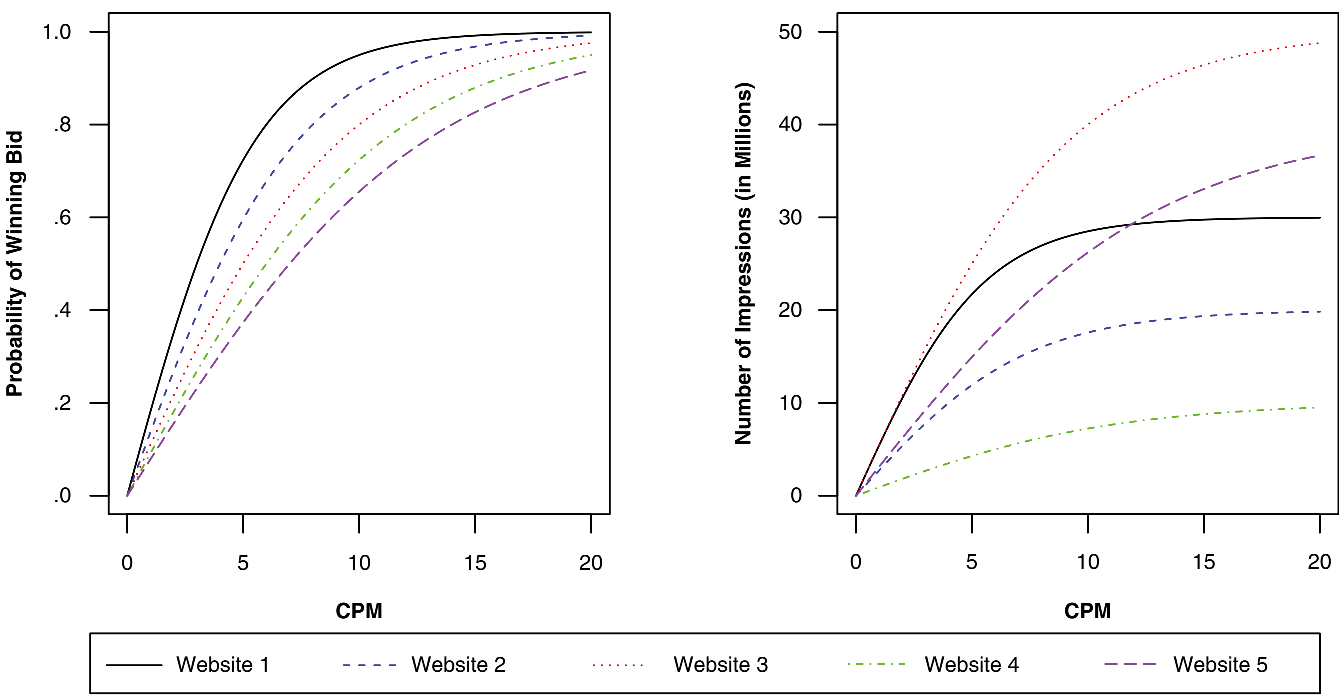
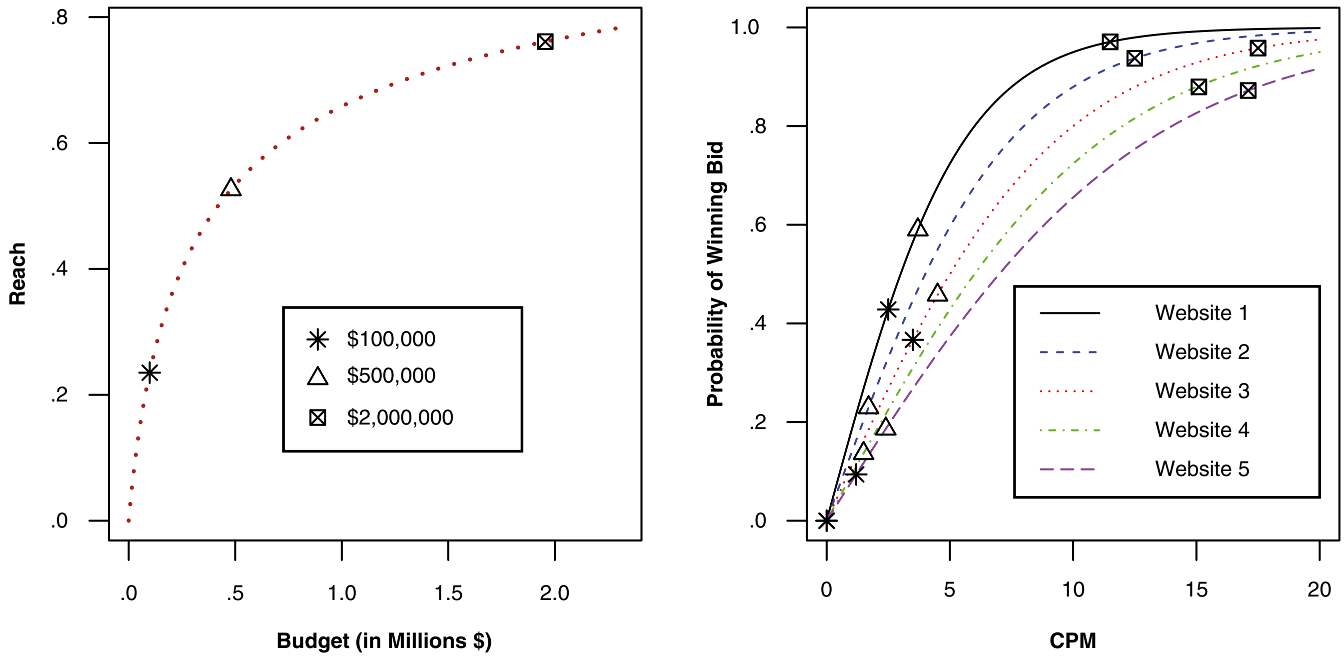


Figure 4

TOTAL REACH ACHIEVED USING THE FIVE WEBSITES GIVEN BUDGET AND THE ALLOCATION OF BUDGET BY BID CPM AT THREE POINTS



Parsons, Haque, and Liu 2004; Zhang, Vorobeychik, and Procaccia 2017; Zhang et al. 2000). The benchmark greedy approach is a simple, but less effective, way of optimizing Equation 5. As with all greedy methods, it looks only one step ahead and increases the allocation to the 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 surpass the first website to become the new 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 the budget spent on that site after the next website is added. By comparison, our approach fully optimizes Equation 5, because 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. For further discussions on conditions under which the advantages of our approach over this benchmark greedy approach are amplified or lessened, see Web Appendix H and the accompanying reach analysis in Table A1. 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.

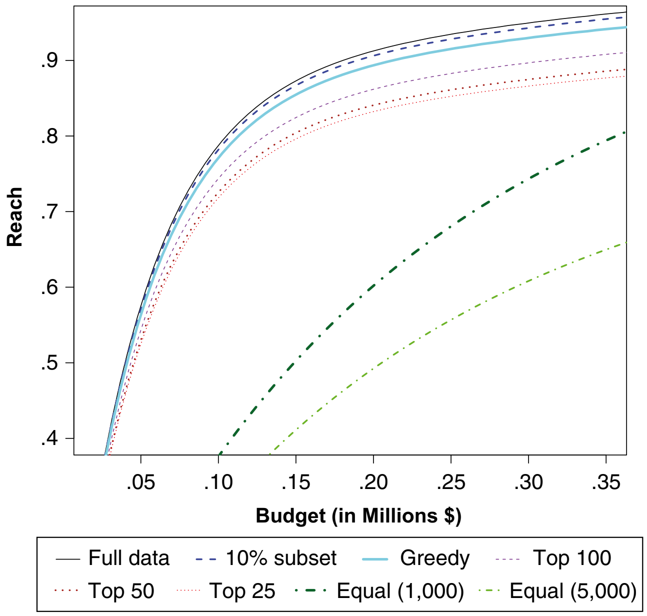
Figure 5 shows the resulting comparisons with the proposed method represented as a blue dashed line. To estimate reach, we first calibrate the proposed and benchmark greedy methods using a 10% subset of the 50,000 users as the training data set and then report reach on the 90% holdout data. We use a similar approach in all reach curve comparisons in this article. 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 that the relative rankings of the reach performances did not change in any of the 20 runs.

We compare the holdout reach calculated from the 10% subset with the results we would achieve over an optimization calibrated on the entire 50,000-user data set (Figure 5, black solid line). Even with only a 10% subset of the data, the proposed method yields reach estimates very similar to those calibrated on the entire data set. In addition, the proposed method outperforms all benchmark approaches. We find that the equal allocation approaches are clearly an inefficient use of budget. 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 NCL Wave Season online display advertising campaigns. These empirical illustrations are based on the 2011 comScore

Figure 5
SIMULATED DATA: 5,000 WEBSITES



Notes: We calculated reach using 90% holdout data.

Media Metrix data, which are from the Wharton Research Data Service (www.wrds.upenn.edu). comScore uses proprietary software to record daily web page use information from a panel of 100,000 Internet users (recorded anonymously by individual computers). Therefore, the comScore data can be used to construct a matrix of all websites visited and the number of page views each computer had for each website during a particular time period. Several prior studies in marketing have utilized the comScore data (e.g., Danaher 2007; Liaukonyte, Teixeira, and Wilbur 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 (Morrison 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 contain 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 × 500 matrix. We then separate our full data set into a 10% training data set (5,667 users), and a 90% holdout data set. We again 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. However, all reach curves in Figures 6–8 represent the average reach across the 100 runs. Relative rankings of the reach performances did not change in any of the 100 runs.

Table 3 provides the makeup for the 16 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 web pages viewed and thus the total number of ad impressions available (in millions).[14] 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 we discussed in the "Simulation Studies" section, we then calculate shifted logistic cost curves for each category of websites on the basis of a c_0 of these average CPMs.[15] Table 3 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 in which to advertise (all over \$6 on average). In addition, 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 16 website categories. Finally, Table 3 shows the number of websites in each category that fall in the top 10, 25, and 50 most-visited websites. Next, 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 that McDonald's is attempting to reach as many users as possible. Because McDonald's is a large company with significant brand awareness, the goal of its McRib campaign is to remind consumers who already know about McDonald's or the McRib that the seasonal sandwich is returning.

Tables 4 and 5 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 million, respectively) for this standard setup (original), as well as for the two extensions (targeted consumers and targeted frequency) considered subsequently.[16] Table 4 shows the results based on a relatively small budget, with an overall expected reach of just over 10%. Thus, budget is allocated to the most valuable websites and website categories (generally chosen by viewership and cost). In contrast, Table 5 presents results from a higher budget of \$1 million, in which reach is approximately 40%. At this budget, the allocation is more diversified, and budget is allocated across a greater number and variety of websites.

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

[14]To get these total impressions values, we follow the procedure of Danaher, Lee, and Kerbache (2010) by extrapolating the comScore data to the effective U.S. Internet population size in December 2011.

[15]As noted previously, in practice DSPs would apply actual, individualized cost curves for all sites in such an optimization, drawing on the results of their own empirical evaluations.

[16]Detailed budget allocation results for bid price at each website are available from authors on request.

Table 3
WEBSITE CATEGORIES IN McRIB EXAMPLE

Category	*Total Websites*	*Total Impressions (in Millions)*	*Avg. CPM*	*Top 10*	*Top 25*	*Top 50*
Community	23	3,322.5	2.10	0	1	4
Email	7	4,044.7	.94	1	3	5
Entertainment	92	7,511.6	4.75	0	0	0
Fileshare	28	3,226.2	1.08	0	2	7
Gaming	77	6,404.2	2.68	0	0	1
General News	12	1,468.6	6.14	0	0	0
Information	47	5,994.9	2.52	1	1	3
Newspaper	27	1,926.1	6.99	0	0	0
Online Shop	29	3,346.5	2.52	0	1	1
Photos	9	770.4	1.08	0	0	2
Portal	30	30,407.7	2.60	3	5	7
Retail	57	9,895.1	2.52	1	2	7
Service	18	4,574.4	2.52	1	2	2
Social Network	17	5,417.1	.56	3	8	11
Sports	17	1,637.2	6.29	0	0	0
Travel	10	529.7	2.52	0	0	0

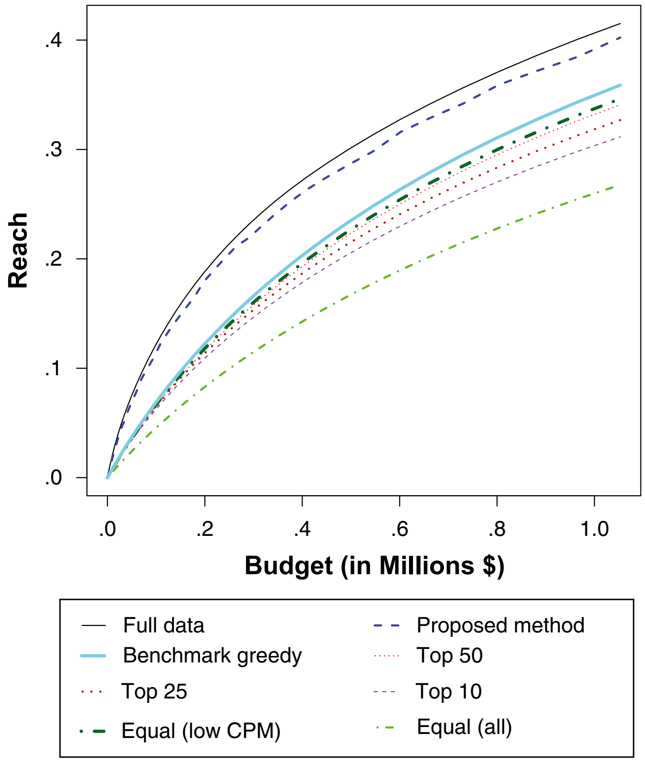
website can reach an otherwise unreachable audience. In this case, other websites, such as social networking sites, offer a relatively inexpensive way to reach consumers who are visiting other websites as well. Table A2 in Web Appendix I shows that social networking sites have relatively high correlations in viewership across other site categories, with the only exceptions being email and gaming sites. Consequently, the optimization ultimately bids heavily across the entire Social Network 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 million budget, our method places bids at all websites in the Email category. In addition to the relative lower cost of advertising on these websites, there is a very low within-category correlation in viewership among email sites (.01 absolute average correlation). This indicates that the same consumer often does not visit more than one email site, so bidding for more email 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, Email, Fileshare, Photos, and Social Network, which total 84 websites); 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 10% calibration data. The reach estimates based on the 10% calibration data are very close to those from the calibration based on the entire data set. In addition, 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

Table 4
WEBSITE CATEGORIES CHOSEN BY METHOD AT A BUDGET OF $100,000, McRIB

			Original		*Targeted Consumers*		*Targeted Frequency*	
Category	*Total Sites*	*Avg. CPM*	*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
Email	7	.94	5	.61	4	.59	7	.48
Entertainment	92	4.75	0	0	3	2.02	0	0
Fileshare	28	1.08	9	.54	9	.52	15	.37
Gaming	77	2.68	5	.77	20	1.35	25	.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	.89
Photos	9	1.08	2	.61	1	.9	4	.58
Portal	30	2.60	0	0	0	0	0	0
Retail	57	2.52	3	1.01	3	.98	11	.73
Service	18	2.52	0	0	0	0	0	0
Social Network	17	.56	16	.28	16	.31	17	.21
Sports	17	6.29	0	0	0	0	0	0
Travel	10	2.52	0	0	0	0	0	0

Figure 6
McRIB CAMPAIGN, MAXIMIZING REACH



Notes: We calculated reach using 90% holdout data.

details on reach comparisons between the proposed and the benchmark greedy methods at various percentiles across the 100 runs at selected budget points. As we 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 with other benchmark equal allocation methods.

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)—that is, 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.

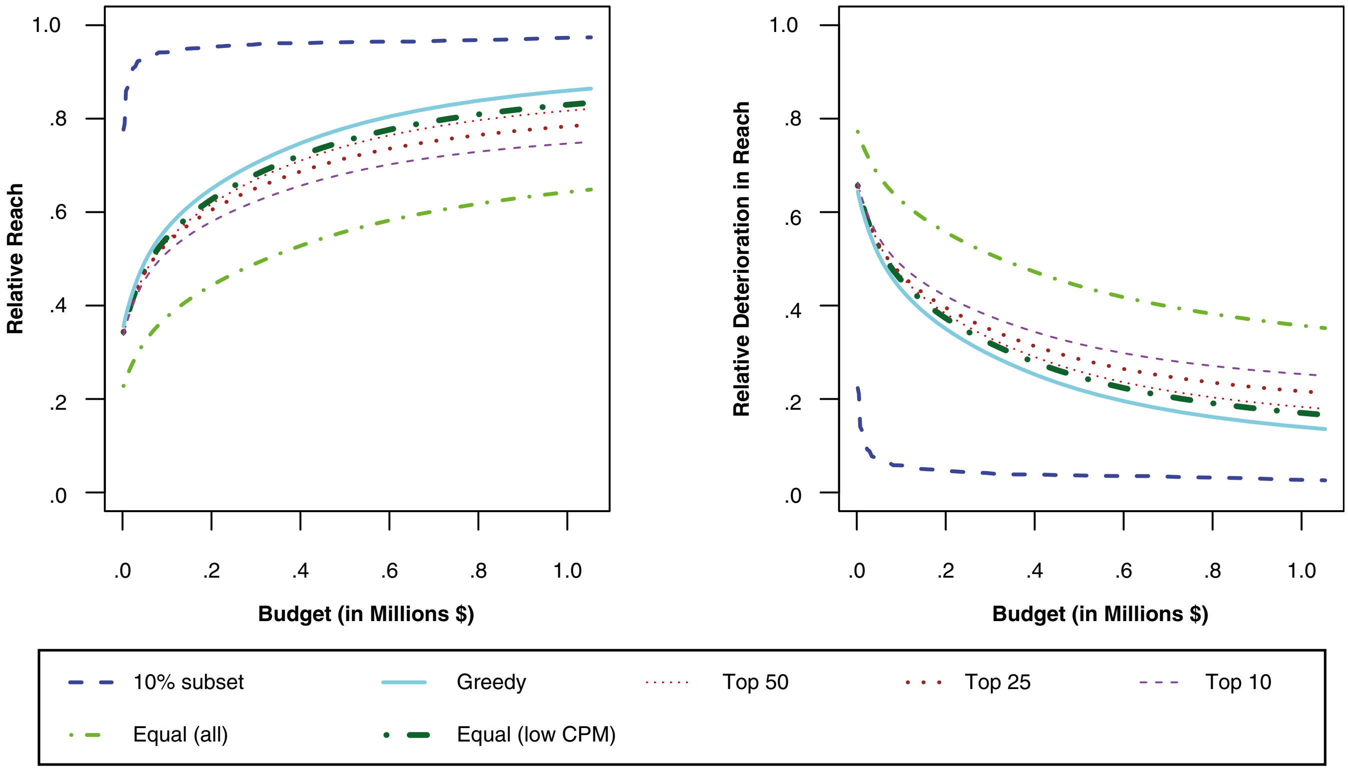
McRib campaign with targeted consumer demographics. In practice, companies often have specific target demographics in mind when running online display ads. In this subsection, we demonstrate how our method can be readily modified to accommodate such needs. We consider a setting involving two demographic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

We assume that an ad served to our target demographic costs twice as much as 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 wants 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

Table 5
WEBSITE CATEGORIES CHOSEN BY METHOD AT A BUDGET OF $1 MILLION, McRIB

			Original		*Targeted Consumers*		*Targeted Frequency*	
Category	*Total Sites*	*Avg. CPM*	*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
Email	7	.94	7	.94	7	.96	7	.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	.56	17	.63	17	.61	17	.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

Figure 7
McRIB CAMPAIGN: RELATIVE REACH AND DETERIORATION IN REACH ESTIMATES



curves, and second, on the remaining users, using the original cost curves. These metrics are then combined to create the values in Tables 4 and 5 (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 4 and 5 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). Thus, 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.

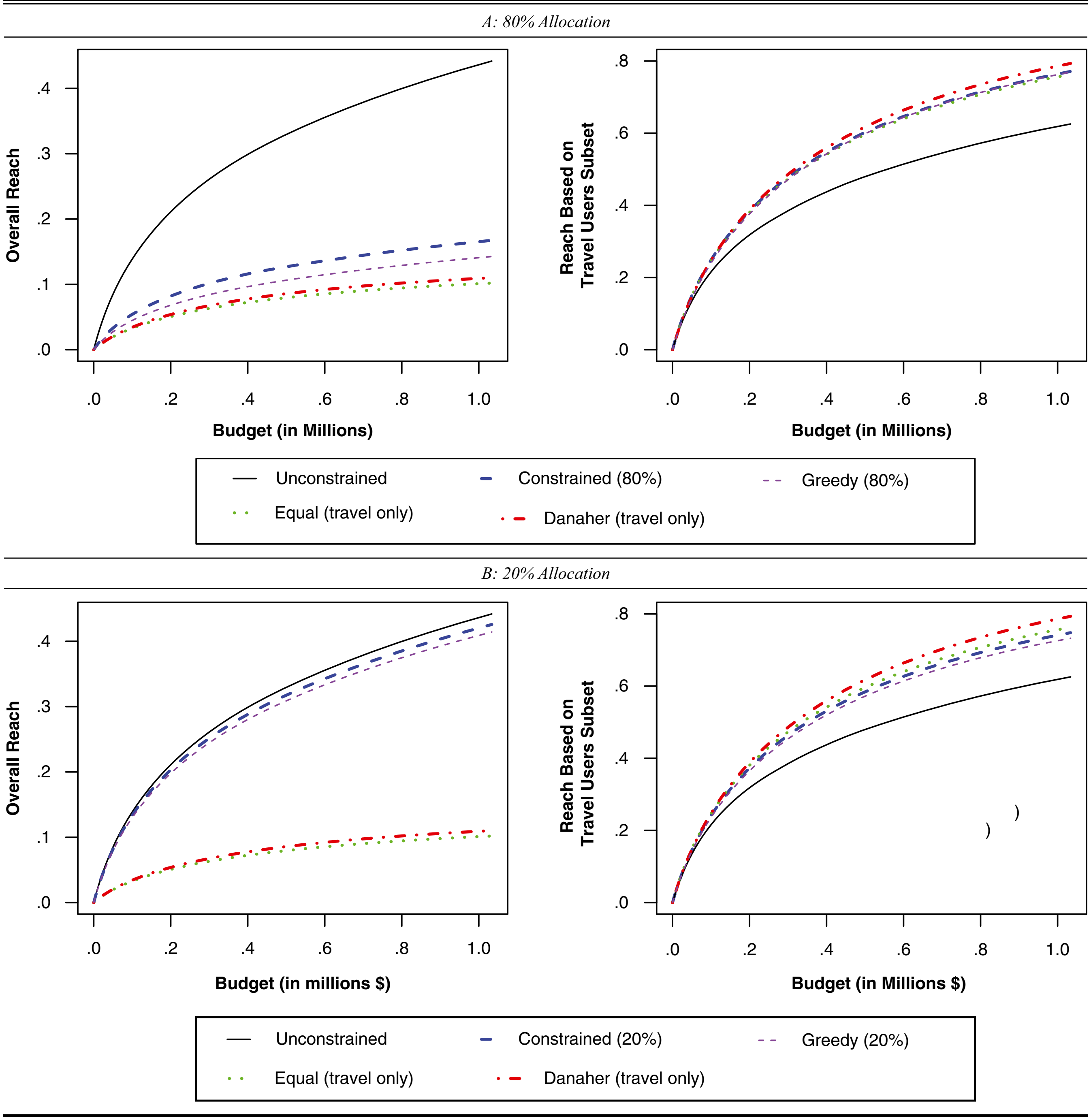
McRib campaign with target frequency of ad exposure. In this subsection, we demonstrate a case in which McDonald's wants to allocate its ad budget such that each person is exposed to the ad no more than three times during the course of the McRib campaign. For simplicity, we use the data set without demographic targeting, although both approaches could readily be used together. In this case, the "effective frequency" is the value of the function $e^{-\gamma}[\gamma + (1/2)\gamma^2 + (1/6)\gamma^3]$.

Again, Tables 4 and 5 show the optimization allocation across website categories for this extension (targeted frequency). In general, in this setting, our method bids 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: NCL 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 Line is among the cruise operators that participate

Figure 8
OVERALL REACH WITH MANDATORY COVERAGE IN AGGREGATE TRAVEL SITES AND ON THE SUBSET OF USERS WHO VISITED AGGREGATE TRAVEL SITES



heavily in wave season (Satchell 2011). Consumers who are interested in booking a cruise often use travel aggregation sites, such as Orbitz and Priceline, to compare offerings across multiple cruise lines. Thus, 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 that 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: In the first, 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). In the second, 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. We found that 48,628 users visited at least one of these 500 websites during January 2011, meaning our Z matrix is 48,628 × 500. We again divide these data into a 10% subset (4,863 users) of calibration data and use the remaining 90% as holdout data.[17]

Figure 8 demonstrates our reach curves under this extension, with the constrained optimization (mandatory media coverage of aggregate travel sites) represented by the dashed blue line, and the unconstrained optimization by the solid black line. 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%). In Figure 8, Panels A and B, we present the the 80% and 20% allocations, respectively.

The left-hand plots show the overall reach on the holdout data. As we expected, the unconstrained curve outperforms the constrained curve, because we cannot improve overall reach by constraining our optimization. As we observed previously, the benchmark greedy algorithm performed worse than the constrained method but better than the other benchmarks. As we 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 comprise only 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 that all other methods are forced to allocate high budgets to the travel websites. Among them, the Danaher approach performs slightly better than the rest. Again, this is expected, given that it allocates 100% of budget to these sites.

It is worth noting that the constrained method has only 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, because 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, selecting optimal Internet media among thousands of websites presents a prohibitively challenging task, particularly in the presence of programmatic advertising and RTB. 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 on which we focus—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 are 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 that leverages bid landscape procedures with both the marketing and statistics literature streams to provide a tool that sets DSP bidding guidelines while simultaneously providing ways to check the efficacy of the established bidding procedures and update as necessary.

Furthermore, the method is designed to be run ahead of a campaign, meaning that 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 could 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. In addition, we currently consider the perspective of an individual firm that wants to maximize reach for a specific campaign. This method could be further extended for use by an advertising broker who wants 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. Finally, future research could examine how to incorporate other important aspects of Internet display ads such as ad blocking (e.g., Marshall 2016; Vranica 2015), 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 (May 25), http://blog.360i.com/web-design/360i-report-gaming.

Adelphic (2015), "Are Media Buyers Using the Best Decision Algorithms?" (accessed July 12, 2016), http://www.adelphic.com/2015/03/media-buyers-using-best-decision-algorithms.

[17]We omit the website category makeup description of this application because of its similarity to Table 3. It is available from authors on request.

AdWords API (2016), "Bid Landscapes," (accessed January 11, 2016), https://developers.google.com/adwords/api/docs/guides/bid-landscapes.

Bateman, Scott (2015), "Online Advertising Models: CPC, CPM or CPA?" Promise Media (accessed November 29, 2016), http://www.promisemedia.com/online-advertising/best-revenue-deals-cpm-cpc-or-cpa.

Breheny, Patrick, and Jian Huang (2011), "Coordinate Descent Algorithms for Nonconvex Penalized Regression, with Applications to Biological Feature Selection," *Annals of Applied Statistics*, 5 (1), 232–53.

Candes, Emmanuel, and Terence Tao (2007), "The Dantzig Selector: Statistical Estimation When p Is Much Larger Than n," *Annals of Statistics*, 35 (6), 2313–51.

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*. New York: Association for Computing Machinery, 265–73.

Danaher, Peter J. (2007), "Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction," *Marketing Science*, 26 (3), 422–37.

Danaher, Peter J. (2008), *Advertising Models*. Boston: Springer, 81–106.

Danaher, Peter J., Janghyuk Lee, and Laouchine Kerbache (2010), "Optimal Internet Media Selection," *Marketing Science*, 29 (2), 336–47.

Danaher, Peter J., and Michael Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30 (1), 1–18.

Downey, Mike (2012), "Real-Time Bidding Is the Next Mobile Ad Breakthrough—Here's How You Can Profit," *VentureBeat* (July 23), http://venturebeat.com/2012/07/23/real-time-bidding-is-the-next-mobile-ad-breakthrough-heres-how-you-can-profit.

Drewnowski, Adam, and Nicole Darmon (2005), "Food Choices and Diet Costs: an Economic Analysis," *Journal of Nutrition*, 135 (4), 900–04.

Efron, Bradley, Trevor Hastie, Iain Johnston, and Robert Tibshirani (2004), "Least Angle Regression (with Discussion)," *Annals of Statistics*, 32 (2), 407–51.

eMarketer (2012), "Digital Ad Spending Tops 37 Billion," (September 20), http://www.emarketer.com/newsroom/index.php/digital-ad-spending-top-37-billion-2012-market-consolidates.

eMarketer (2013), "Is the Future of Programmatic Premium?" (December 4), http://www.emarketer.com/Article/Future-of-Programmatic-Premium/1010430.

Fan, Jianqing, and Runze Li (2001), "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of the American Statistical Association*, 96 (456), 1348–60.

Feldman, Jon, and S. Muthukrishnan (2008), "Algorithmic Methods for Sponsored Search Advertising," in *Performance Modeling and Engineering*, Zhen Liu and Cathy H. Xia, eds. New York: Springer, 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), 302–32.

G2 Crowd (2016), "Demand Side Platform (DSP)," (accessed May 17, 2016), https://www.g2crowd.com/categories/demand-side-platform-dsp.

Goeman, Jelle (2010), "L1 Penalized Estimation in the Cox Proportional Hazards Model," *Biometrical Journal. Biometrische Zeitschrift*, 52 (1), 70–84.

Goldberg, Michael (2015), "Understanding the Programmatic Ecosystem," Dun & Bradstreet (August 27), http://www.dnb.com/perspectives/marketing-sales/understand-programmatic-ecosystem.html.

Hatano, Daisuke, Takuro Fukunaga, Takanori Maehara, and Kenichi Kawarabayashi (2015), "Lagrangian Decomposition Algorithm for Allocating Marketing Channels," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence, 1144–50.

Hesterberg, Tim, Nam Hee Choi, Lukas Meier, and Chris Fraley (2008), "Least Angle and L1 Penalized Regression: A Review," *Statistics Surveys*, 2, 61–93.

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*, 53 (3), 375–93.

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," in *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, B.W. Croft and C.J. van Rijsbergen, eds. London: Springer-Verlag, 3–12.

Lewis, Randall, and Dan Nguyen (2015), "Display Advertising's Competitive Spillovers to Consumer Search," *Quantitative Marketing and Economics*, 13 (2), 93–115.

Liaukonyte, Jura, Thales Teixeira, and Kenneth C. Wilbur (2015), "Television Advertising and Online Shopping," *Marketing Science*, 34 (3), 311–30.

Lipsman, Andrew (2010), "The New York Times Ranks as Top Online Newspaper According to May 2010 U.S. comScore Media Metrix Data," press release (June 16), 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.

Marshall, Jack (2016), "The Rise of the Anti-Ad Blockers," *The Wall Street Journal* (June 13), https://www.wsj.com/articles/the-rise-of-the-anti-ad-blockers-1465805039.

Meinshausen, Nicolai (2007), "Relaxed Lasso," *Computational Statistics & Data Analysis*, 52 (1), 374–93.

Mintel (2014), "Kids as Influencers—US," technical report (April), https://store.mintel.com/kids-as-influencers-us-april-2014.

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. New York: Association for Computing Machinery, 1395–404.

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), 579–95.

Morrison, Maureen (2012), "Can the McRib Save Christmas?" *Ad Age* (September 17), http://adage.com/article/news/mcrib-save-christmas/237233/.

Mugridge, Andrew (2016), "CPC Bidding vs. CPM Bidding: What's the Difference?" iThinkMedia (August 2), https://www.ithinkmedia.co.uk/blog/paid-media/cpm-cpc-bidding/.

Naples, Michael J. (1979), *Effective Frequency: The Relationship Between Frequency and Advertising Effectiveness*. New York: Association of National Advertisers.

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," *SIGKDD Explorations*, 6 (1), 90–105.

Radchenko, Peter, and Gareth James (2008), "Variable Inclusion and Shrinkage Algorithms," *Journal of the American Statistical Association*, 103 (483), 1304–15.

Rayport, Jeffrey R. (2015), "Is Programmatic Advertising the Future of Marketing?" *Harvard Business Review* (June 22), https://hbr.org/2015/06/is-programmatic-advertising-the-future-of-marketing.

Rosset, Saharon, and Ji Zhu (2007), "Piecewise Linear Regularized Solution Paths," *Annals of Statistics*, 35 (3), 1012–30.

Sahni, Navdeep S. (2015), "Effect of Temporal Spacing between Advertising Exposures: Evidence from Online Field Experiments," *Quantitative Marketing and Economics*, 13 (3), 203–47.

Sahni, Navdeep S. (2016), "Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising," *Journal of Marketing Research*, 53 (4), 459–78.

Satchell, Arlene (2011), "Norwegian: Cruise Fares to Increase up to 10 Percent April 1," *South Florida Sun-Sentinel* (March 1), http://articles.sun-sentinel.com/2011-03-01/business/fl-cruise-price-hike-20110301_1_norwegian-epic-price-hike-cruise-fares.

Schmidt, Mark, Glenn Fung, and Rmer Rosales (2007), "Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches," in *Machine Learning: ECML*Joost N. Kok, Jacek Koronacki, Raomon Lopez de Mantaras, Stan Matwin, Dunja Mladenič, and Andrzej Skowron, eds. Warsaw: Springer, 286–97.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B. Methodological*, 58 (1), 267–88.

Vidakovic, Ratko (2013), "5 Common Myths About Real-Time Bidding," MarketingLand (September 16), http://marketingland.com/5-common-myths-about-real-time-bidding-58708.

Vranica, Suzanne (2015), "Ad Blocking is a Hot Topic of Marketing, Media Executives," *The Wall Street Journal* (September 27), https://www.wsj.com/articles/ad-blocking-is-a-hot-topic-for-marketing-media-executives-1443259981.

Wang, Jun, Shuai Yuan, and Weinan Zhang (2016), "ECIR 2016 Tutorial: Real-Time Bidding Based Display Advertising: Mechanisms and Algorithms," presented at the 2016 European Conference on Information Retrieval, http://wnzhang.net/slides/ecir16-rtb.pdf.

Wu, Tong Tong, and Kenneth Lange (2008), "Coordinate Descent Algorithms for Lasso Penalized Regression," *Annals of Applied Statistics*, 2 (1), 224–44.

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 (2), 461–78.

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*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 1232–40.

Zhang, Weinan, and Jun Wang (2015), "Statistical Arbitrage Mining for Display Advertising," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 1465–74.

Zhang, Weinan, Shuai Yuan, and Jun Wang (2014), "Optimal Real-Time Bidding for Display Advertising," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 1077–86.

Zhang, Zheng, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A Greedy Algorithm for Aligning DNA Sequences," *Journal of Computational Biology*, 7 (1/2), 203–14.

Zhao, Peng, Guilherme Rocha, and Bin Yu (2009), "The Composite Absolute Penalties Family for Grouped and Hierarchical Variance Selection," *Annals of Statistics*, 37 (6A), 3468–97.

Zou, Hui (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101 (476), 1418–29.

Zou, Hui, and Trevor Hastie (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society. Series B. Methodological*, 67 (Pt 2), 301–20.