

Risk-Aware Revenue Maximization in Display Advertising

William D. Heavlin
 Google, Inc.
 1600 Amphitheatre Parkway
 Mountain View, CA 94043
 bheavlin@google.com

Ana Radovanović
 Google Research
 76 Ninth Ave.
 New York, NY 10011
 anaradovanovic@google.com

ABSTRACT

Display advertising is the graphical advertising on the World Wide Web (WWW) that appears next to content on web pages, instant messaging (IM) applications, email, etc. Over the past decade, display ads have evolved from simple banner and pop-up ads to include various combinations of text, images, audio, video, and animations. As a market segment, display continues to show substantial growth potential, as evidenced by companies such as Microsoft, Yahoo, and Google actively vying for market share. As a sales process, display ads are typically sold in packages, the result of negotiations between sales and advertising agents.

A key component to any successful business model in display advertising is sound pricing. Main objectives for on-line publishers (e.g. Amazon, YouTube, CNN) are maximizing revenue while managing their available inventory appropriately, and pricing must reflect these considerations.

This paper addresses the problem of maximizing revenue by adjusting prices of display inventory. We cast this as an inventory allocation problem. Our formal objective (a) maximizes revenue using (b) iterative price adjustments in the direction of the gradient of an appropriately constructed Lagrangian relaxation. We show that our optimization approach drives the revenue towards local maximum under *mild* conditions on the properties of the (*unknown*) demand curve.

The major unknown for optimizing revenue in display environment is how the demand for display ads changes to prices, the classical demand curve. This we address directly, by way of a factorial pricing experiment. This enables us to estimate the gradient of the revenue function with respect to inventory prices. Overall, the result is a principled, risk-aware, and empirically efficient methodology.

This paper is based on research undertaken on behalf of one of Google's clients.

Categories and Subject Descriptors

H.4.2 [Information Systems Applications]: [decision support]; G.3 [Probability and Statistics]: [estimation, experiment design]; G.1.6 [Optimization]: [duality]

General Terms

Optimization, Econometrics, Experiment Design

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2012, April 16–20, 2012, Lyon, France.
 ACM 978-1-4503-1229-5/12/04.

Keywords

Display Advertising, Revenue Maximization, Pricing Experiment

1. INTRODUCTION

Display advertising is a fast-growing, multi-billion business, which provides a premium way of advertising online (see [7], [17]). This is much more than ads in Web browsers. People are watching video, reading newspapers, magazines, books and listening to digital music at an ever-increasing rate. They are turning to new devices like smartphones, tablets, e-readers and video game consoles. Display advertising has a significant advantage over advertising in magazines, newspapers and TVs: (i) it provides targeting options such as demographic and behavioral targeting to laser on a specific audience, (ii) one can track the performance of the advertising campaign daily to measure metrics such as impressions, clicks and conversions.

In order to make display advertising work better, large companies like Microsoft, Yahoo and Google have been investing in new technologies that should help grow display advertising for all publishers by orders of magnitude (see [7], [2]). One of such examples is a DoubleClick (Google) platform, called *DoubleClick for Publishers*, which is an ad serving platform that maximizes the value of ad space that publishers directly sold themselves. The overall goal is to give publishers a firm control and empower them with more data, reports and controls and, therefore, help them make better decisions about ad space forecasting, segmentation, targeting, allocation and *pricing*. Pricing of display ads is one of the most challenging tasks and can significantly impact publisher's revenue (see [12], [6], [1]).

Specific dynamics in the process of sales of display ads makes the pricing problem different from the related offline and online pricing schemes. Advertisers pre-purchase a reservation package of online inventory (impressions) on content sites (publishers). In this business context, an advertiser with certain advertising goals approaches a sales representative either directly or through an ad agency, after which they start a negotiation. The result of this process is a sold package of impressions, which represents the number of times a certain ad is displayed on a Web page when users access it within the desired window of time in the future. Impression (inventory) categories differ in their properties, such as size, type (text, video, etc.), position, as well as monitored performance measures (Click Through Rate (CTR), conversion rate, etc.), which usually impact their price. Data collected from the 2009 DoubleClick (Google)

User Group sessions revealed that publishers adhere to a wide spectrum of pricing practices, ranging from scientific methodologies to educated guesses. As a result, it is likely that many publishers aren't generating as much revenue as possible from their inventory. However, by applying scientific methodologies to publishers' pricing practice, they may be able to improve their monetization.

1.1 Our Contributions

This paper establishes a revenue maximization methodology for modifying prices of online ad inventory that is sold by sales people and is reserved for the guaranteed delivery. We model the problem as inventory (impression) allocation problem with capacity constraints for different types. Since in practice change in inventory allocations as a function of price is not known, we propose an iterative procedure for gradual increase in revenue that is based on tracking the gradient of a suitably chosen Lagrange relaxation of the revenue objective. The gradient of the impression allocation function at the current price point is obtained using the only practically-feasible way, pricing experiment. However, there are various practical constraints that one needs to satisfy in the process of estimating the revenue gradient and conducting consequent price adjustments.

As already described before, an intrinsic ingredient of the sales process is negotiation. Even though there is a menu of inventory categories and the corresponding prices, a negotiation takes place and every sales person's goal is to negotiate the 'most' profitable deal. Our goal in this paper is *not* to change the current sales process, but to change the listed prices which are used by sales representatives as a 'starting point' for their negotiation. Calculating the revenue-increasing inventory (listed) prices requires a rigorous and thorough collection of accurate pricing-related information, including product (ad) information, historical sales information, delivery data, traffic forecasts, impression availability and sell-through rates. For example, Google's (DoubleClick's) tool for data collection and storage is called DART Sales Manager. Records on various offers and advertisers' responses during the deal negotiations are *not* recorded, except for the information about the package of impressions that is finally sold. In this paper, we do not directly incorporate methods sales people use to negotiate the deal (e.g., price discounts, bonus (free) impressions) since these methods vary depending on a publisher. However, we present a fundamental approach in treating the negotiated price-controlled inventory allocation problem, which can be extended to specific negotiation practices. Due to the space limitation, we do not address these extensions in this paper.

Our first of a kind, practically feasible and statistically rigorous revenue maximization methodology incorporates:

- Iterative optimization algorithm for adjusting inventory prices to gradually increase revenue.
- A statistically rigorous, experimental method for estimating the gradient of impression allocations with respect to prices (i.e., allocation sensitivity). The proposed procedure incorporates several important properties:
 - It is very simple to execute and minimally interferes with the current sales process, which is necessary in convincing publishers to allow the experimentation.

- It incorporates ways to reduce risks of different nature: (a) Effects from the variation in advertiser inventory requirements and budget constraints; (b) 'Noise' originating from changes in the current economy, seasonality, management decisions, etc.
- It is maximally efficient given the experimental constraints and incorporates a way to extend the experiment's duration (when more data is needed to make statistically reliable decisions).
- It scales with the number of inventory categories.

1.2 Literature Review

The *novel* methodology presented in this paper touches upon several streams of literature.

First, pricing reservation-based (or guaranteed) contracts in display advertising is a challenging task due to several intrinsic properties of its sales process: (i) contracts might be signed at different points of time before they actually start, (ii) advertisers' inventory and audience preferences can be also quite diverse which makes the demand categorization hard. In addition to negotiation which is currently the most dominant sales model for selling display inventory, estimating market value of publisher's inventory is a challenging task. This problem was addressed in [6] with the goal to design an automated pricing system used by sales agents as a starting point for negotiation and is a reflection of inventory's *market value*. In this paper, we track inventory allocations in time and measure the impact of their price change on publisher's revenue. Due to *many* noise factors that could make this task almost impossible, there is a need for risk-aware experimental methodology that would eliminate bias coming from (i) variations in advertiser inventory requirements, (ii) 'noise' originating from changes in the current economy, seasonality, management decisions, etc. The method proposed in this paper can be easily extended to an arbitrary number of different inventory types and handles the collected transaction data in the maximally efficient way (as will be described later).

In [16], there is a comparison in expected revenues between different pricing schemes used in display advertising with no consideration of uncertainties in advertiser's demand and viewers' supply. For more work addressing ad delivery, online search behavior and other decision making techniques for a Web publisher with an advertising operation, an interested user is referred to [19] and [12], and references therein.

There are other ways to sell online inventory, which are different from selling guaranteed display impression packages addressed in this paper. Real-time auctions are used extensively in sponsored search ([10]) and non-guaranteed display advertising ([18]). Here, the pricing method is combinatorial auctions ([9]), and the problem is to find a yield-maximizing way to allocate inventory. Guaranteed contracts are sold at different points in time, and are not amenable to auctions.

There are other optimization problems that arise in display advertising area and are not directly related to the revenue maximization in this paper. One of them addresses serving (or matching) display ads to incoming users. There is a series of work that casts this problem as a stochastic matching problem where decision control is obtained using primal-dual optimization method derived from the packing linear program (LP) (see [11], [5] and references therein). Apart from focusing on different optimization problems, all

of the mathematical models in these papers require knowledge of demand distributions and impose strong regularity conditions on the revenue function. This paper addresses a fundamental problem of iterative solving of the optimization problem under *no* distributional assumptions on model parameters and under very *mild* conditions on demand curve necessary to estimate its gradient (elasticity, sensitivity) at the current price point.

One of the main contributions of this work comes from the experimental methodology we use to estimate elasticity of the demand curve at the current price point. The literature on experiment design mostly relates to the *marketing area* and is quite extensive. Given the importance of understanding which inventory attributes are attractive to advertisers and, therefore, having choice menus is widely advisable for the Web-based information services (see [15] and references within). In a vast stream of literature on learning consumer choice behavior, there is a clear knowledge of the market and advertisers' preferences (see, for example, [3], [4], etc.), which focuses on model selection that could be later used to optimize revenue. However, these so called *conjoint studies* are essentially infeasible given the current way of selling display ads. The process of negotiating a deal is highly dependent on sales representative and the past relationship between a sales person and an advertiser. Also, currently used tools for sales data collection and storage do not record intermediate negotiation results and details, but only the final contract details, which include impression allocations, their price and campaign duration. In addition, there is a general tendency not to let advertisers leave without a contract after the negotiation. This eliminates an option of 'learning' whether advertiser actually rejected the offered package of impressions. Thus, the choice of the experimental approach is quite limited. Our *fractional factorial design* ([8]) has all of the desirable properties of an experiment in display advertising business: it provides a statistically sound and risk-aware approach that deals with small data volumes and various bias in the most efficient manner. The only drawback in conducting any pricing experiment is that there are no revenue guarantees while it lasts. In order to avoid large potential loss of revenue in the experimental phase, publishers allow only moderate price changes during this time, which further supports our step-wise approach in optimizing revenue.

2. PROBLEM FORMULATION AND ASSUMPTIONS

In this section we introduce our mathematical model that captures the sales process, including measurable model parameters and assumptions used in the analysis. It relies strongly on our prior research of Google's publisher clients. On a separate note, we refer an interested reader to [6], which describes the key features of the sales dynamics presented in this paper.

A stream of advertisers arrives to a given publisher according to some point process in time. Upon her/his arrival, an advertiser, a , $a \in \mathcal{A}$, brings budget B_a that she/he has available to spend on a package of impressions. In addition, advertisers usually have some initial preferences with respect to types of ads they want to purchase (video, text, multimedia, etc.), as well as certain performance expectations (measured through CTR, conversion rate, etc.), which are not

recorded in practice but influence the negotiation dynamics (discussed later). We assume that *transaction budgets are independent from their arrival points in time within a given time interval*. Analysis of actual transaction traffic implies that the transaction (arrival) rate of one advertiser is much smaller than the aggregate transaction (arrival) rate of all advertisers. Therefore, *assuming that purchasing decisions of different advertisers are mutually independent, we treat a collection of transactions that occurs in a given bounded interval of time (experimental phase) independent as well*.

Each advertiser has a single sales agent handling all of its campaigns. An assigned sales agent starts the negotiation by trying to incorporate all of advertiser's campaign goals by allocating a desirable package of impressions, denoted here by a K -tuple $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$ ($K < \infty$ is the number of inventory types). Apart from the campaign goals (including budget and past inventory choices), sales agent's allocation decision depends on the current inventory utilization levels and prices. Unfortunately, *no records* on the negotiation process are kept, except for the final package of agreed-to impressions. Impression allocations, $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$, are an (*unknown*) function of the current price vector, $\bar{p} = (p_1, \dots, p_K)$ where p_k , $1 \leq k \leq K$, is the price of inventory k . (In practice, the price is usually expressed in CPM units, the cost of 1000 impressions.) We use $U_k(\bar{p})$, $1 \leq k \leq K$, to denote a percentage of inventory k that is currently utilized (or already assigned) at the moment of transaction. The negotiated impression package, $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$, needs to add up to the advertiser's available budget spent, i.e., $B_a = p_1 M_{a1}(\bar{p}) + \dots + p_K M_{aK}(\bar{p})$.

To summarize, recorded (available) information from the past sales/transactions includes:

- The Advertiser's name, say a , and the assigned sales agent;
- Incoming budget, B_a ;
- Inventory percent-availability at the moment of sale, $(1 - U_1(\bar{p}), \dots, 1 - U_K(\bar{p}))$; and
- The final package of negotiated impressions, $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$.

We do not make any distributional assumptions on the arrival process of advertisers, budgets, contract duration, etc. Implicitly, however, we assume sufficient stability for estimates from one time period, in order to remain useful at later time periods.

Each publisher's objective is to increase revenue subject to its current available inventory. The total revenue equals to the sum of revenues spent on all inventories, i.e.,

$$R(\bar{p}) = \sum_{a \in \mathcal{A}} B_a = \sum_{a \in \mathcal{A}} \sum_{k=1}^K M_{ak}(\bar{p}) p_k. \quad (1)$$

Therefore, the optimization problem that a publisher faces is the following:

$$\begin{aligned} \max_{\bar{p}} \quad & \sum_{k=1}^K p_k \sum_a M_{ak}(\bar{p}) \\ \text{s.t.} \quad & \sum_a M_{ak}(\bar{p}) \leq C_k, \quad k = 1, \dots, K. \end{aligned} \quad (2)$$

Note that (2) is a constrained inventory allocation problem where the inventory allocation is controlled by the price

vector \bar{p} . Despite the relatively simple formulation, solving this optimization problem turns out to be challenging given that the allocation process $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$ as a function of \bar{p} (i.e. demand surface) is *not* known. Therefore, the only feasible optimization procedures are iterative or stepwise ones, which incorporate learning the gradient of $(M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p}))$ at the current price vector.

Our approach to solving the constrained optimization problem (2) uses Lagrangian relaxation of the objective function, i.e.,

$$\mathcal{L}(\bar{\lambda}, \bar{p}) = \sum_{k=1}^K p_k \sum_a M_{ak}(\bar{p}) + \sum_{k=1}^K \lambda_k (C_k - \sum_a M_{ak}(\bar{p})),$$

where $\bar{\lambda} = (\lambda_1, \dots, \lambda_K)$ represents a vector of *costs* (or dual variables) associated with available (non-sold) inventory. Assuming the specific relation between this cost and price vectors, we provide a stepwise optimization procedure that guarantees an increase in expected revenue after execution of each price update and under these rather mild (and natural) regularity and market *assumptions*:

- A1** $U_k(\bar{p})$, $1 \leq k \leq K$, is differentiable;
- A2** An increase in price of inventory k leads to non-decreases in demand (allocations) for inventory j , $j \neq k$ (substitution of demand assumption);
- A3** An increase in price of inventory k results (a) in a non-increase in demand for inventory k . Further, (b) the relative change in demand for inventory k to its relative price change is of smaller absolute value. In particular, halving the price of inventory k results is less than doubling of demand for that inventory (decreasing marginal utility assumption);
- A4** Sales agents act consistently across all advertisers in the presence of price perturbations at the current price vector (aggregation assumption).

The aggregation assumption from above is a common assumption in the econometric theory (see, for example, Chapter 2 of [20]) and, very often, it is not explicitly stated. Relaxing this assumption would increase the number of experimental conditions one needs to consider when estimating the gradient of function $\mathcal{L}(\bar{\lambda}, \bar{p})$ at the current price point \bar{p} . This is commonly not sustainable in the presence of small amounts of data.

2.1 Optimization Algorithm

In this subsection, we propose our optimization algorithm and show that each recommended price update increases publisher's expected revenue. Note that the utilization of the inventory k , $U_k(\bar{p})$, can be re-expressed as the sum of impression allocations across all advertisers, $C_k U_k(\bar{p}) = \sum_a M_{ak}(\bar{p})$. Our stepwise algorithm consists of repeating the following three steps:

ITERATIVE OPTIMIZATION ALGORITHM:

1. Estimate the gradient of $\mathcal{L}(\bar{\lambda}, \bar{p})$, $\nabla_{\bar{p}} \mathcal{L}(\bar{\lambda}, \bar{p})$ at the current price point where $\lambda_k \equiv \lambda_k(\bar{p}) = p_k U_k(\bar{p})$ (i.e., we assume that the cost of non-sold inventory equals to the effective inventory unit price);
2. Adjust the price vector in the direction of the estimated gradient;

3. Repeat 1. and 2. until $\nabla_{\bar{p}} \mathcal{L}(\bar{\lambda}, \bar{p}) = \mathbf{0}$.

Next, we state the main theorem of this paper.

THEOREM 1. *Assuming that the gradients are accurately estimated, the **ITERATIVE OPTIMIZATION ALGORITHM** increases revenue at each iteration and converges to the local maximum of the revenue function.*

Remarks:

- Estimating $\nabla_{\bar{p}} \mathcal{L}(\bar{\lambda}, \bar{p})$ requires measuring responses in the presence of perturbed prices which is captured by the experimental methodology in Section 3. Each experiment needs to last long enough to collect sufficient transaction data to be able to estimate gradient with a reasonable accuracy, which will be discussed later.
- Given that resolving each price update towards a revenue increase requires certain experimental time during which there are no guarantees on revenue change, publishers are commonly satisfied with only one iteration and an incremental revenue increase. However, no matter when a publisher decides to terminate the execution of the **ITERATIVE OPTIMIZATION ALGORITHM**, an increase in revenue is guaranteed.

Proof: Define $\bar{q} = (q_1, \dots, q_K)$ where $q_k = \log p_k$, $1 \leq k \leq K$, and

$$b_{jk} \triangleq \frac{\partial \log U_j}{\partial q_k}. \quad (3)$$

Then, for $\lambda_k = p_k U_k(\bar{p})$ and $U_k \equiv U_k(\bar{p})$ we obtain

$$\begin{aligned} \frac{\partial \mathcal{L}(\bar{\lambda}, \bar{p})}{\partial q_k} &= \frac{\partial}{\partial q_k} \sum_{j=1}^K C_j [U_j \exp q_j + (1 - U_j) \lambda_j] \\ &= C_k U_k p_k + \sum_{j=1}^K C_j (p_j - \lambda_j) U_j \frac{\partial \log U_j}{\partial q_k} \\ &= C_k U_k p_k + \sum_{j=1}^K C_j (p_j - \lambda_j) U_j b_{jk} \\ &= C_k U_k [p_k + (p_k - \lambda_k) b_{kk}] + \sum_{j \neq k} C_j U_j (p_j - \lambda_j) b_{jk} \\ &= C_k U_k p_k [1 + (1 - U_k) b_{kk}] + \sum_{j \neq k} C_j U_j (1 - U_j) p_j b_{jk}, \end{aligned} \quad (4)$$

In view of *Assumptions* from before, this latter expression is well defined and constant across all advertisers at the current price point \bar{p} . Note that the Taylor expansion implies

$$\Delta \log U_k(\bar{p}) = \frac{U_k(\bar{p} + \Delta \bar{p}) - U_k(\bar{p})}{U_k(\bar{p})} + o(\|\Delta \bar{p}\|) \text{ as } \|\Delta \bar{p}\| \rightarrow 0, \quad (6)$$

and that considering relative change or change in log utilization is a way of normalizing highly variable changes in inventory utilization that originate from advertisers' different budget ranges. Next, similarly as in (5), we obtain

$$\frac{\partial R(\bar{p})}{\partial q_k} = C_k U_k p_k [1 + b_{kk}] + \sum_{j \neq k} C_j U_j p_j b_{jk}. \quad (7)$$

Since Assumption **A2** implies $b_{jk} \geq 0$, and Assumption **A3** implies $-1 \leq b_{kk} < 0$, we conclude that $\frac{\partial \mathcal{L}(\bar{\lambda}, \bar{p})}{\partial q_k}$ and $\frac{\partial R(\bar{p})}{\partial q_k}$

are both positive, implying that both the Lagrangian and revenue function values increase at each iteration of the algorithm. Since both of the functions are bounded, the algorithm will naturally terminate at the price point that corresponds to local maximum of the revenue function. \diamond

2.2 Gradient Estimation

In this section, we describe the model we use to estimate price elasticities (sensitivities) b_{ji} . In order to estimate the sensitivity, we have to eliminate effects of 'noise' factors and business constraints that continuously change in time. For that reason, we experiment with different (perturbed) price vectors concurrently by assigning them to non-intersecting, but statistically similar experimental sales groups. More details on the experimental procedure is provided in Section 3.

As described above, for each transaction, DART for Publisher's database records: (i) negotiated inventory allocation quantities ($M_{a1}(\bar{p}), \dots, M_{aK}(\bar{p})$), and (ii) available inventories at the time of transaction (which we use to compute inventory utilization (U_1, \dots, U_K)). Then, after conducting the experiment, the transaction data corresponding to different experimental groups are analyzed together. This is one of the reasons why the design proposed in this paper is efficient. Similarly as in (4), using Taylor approximation, we obtain

$$\frac{\partial \log U_j}{\partial \log p_k} = \frac{1}{U_j} \frac{\partial U_j}{\partial \log p_k} + o(\Delta p_k) \text{ as } \Delta p_k \rightarrow 0.$$

Then, since

$$U_j = \frac{1}{C_j} \sum_a M_{aj},$$

using Assumption **A4** we obtain that in the presence of small perturbations Δp_k , $k = 1, \dots, K$,

$$\begin{aligned} b_{jk} &= \frac{\partial \log U_j}{\partial \log p_k} \\ &= \sum_a \frac{M_{aj}}{\sum_a M_{aj}} \frac{\partial \log M_{aj}}{\partial \log p_k} \\ &= \frac{\partial \log M_{aj}}{\partial \log p_k}. \end{aligned}$$

In order to compute b_{jk} we use linear model

$$dM = [1 \ dp \ dU^*]B + E, \quad (8)$$

where dM is the $n \times K$ matrix of *responses* that represent relative changes in impression allocations to the corresponding advertisers when compared to their values when the offered prices were \bar{p} (pre-experimental time), i.e.,

$$dM \triangleq \begin{bmatrix} \Delta \log M_{a_1 1} & \Delta \log M_{a_1 2} & \dots & \Delta \log M_{a_1 K} \\ \Delta \log M_{a_2 1} & \Delta \log M_{a_2 2} & \dots & \Delta \log M_{a_2 K} \\ \dots & \dots & \dots & \dots \\ \Delta \log M_{a_n 1} & \Delta \log M_{a_n 2} & \dots & \Delta \log M_{a_n K} \end{bmatrix}; \quad (9)$$

index a_i is used to denote the advertiser in i th transaction. Furthermore, the Δ -notation emphasizes the advertiser-specific changes in demand from the previous period. Such changes from baseline are a zero-order correction for differences in preference among advertisers. We use n to denote the number of transactions observed during the experimental period.

Next, dp is defined as

$$dp \triangleq \begin{bmatrix} \Delta \log p_1^{(1)} & \Delta \log p_2^{(1)} & \dots & \Delta \log p_K^{(1)} \\ \Delta \log p_1^{(2)} & \Delta \log p_2^{(2)} & \dots & \Delta \log p_K^{(2)} \\ \dots & \dots & \dots & \dots \\ \Delta \log p_1^{(n)} & \Delta \log p_2^{(n)} & \dots & \Delta \log p_K^{(n)} \end{bmatrix}. \quad (10)$$

Matrix dp is called a *design* matrix and will be discussed in detail in Section 3.

Then, dU^* represents generic features for each advertiser-transaction and, in view of our modeling assumptions described in Section 2, sales agents' allocation decisions depend on some function of inventory utilization, say U_{jn}^* , at the moment of transaction, i.e.

$$dU^* \triangleq \begin{bmatrix} \Delta U_{11}^* & \Delta U_{21}^* & \dots & \Delta U_{K1}^* \\ \Delta U_{12}^* & \Delta U_{22}^* & \dots & \Delta U_{K2}^* \\ \dots & \dots & \dots & \dots \\ \Delta U_{1n}^* & \Delta U_{2n}^* & \dots & \Delta U_{Kn}^* \end{bmatrix}. \quad (11)$$

As it was described before, utilization (or availability) of specific inventories influences the content of the negotiated package. We scale utilization in order to model non-linear functional dependence between allocation decisions and the utilization at the moment of transaction. Appropriate scaling is usually the result of empirical study conducted to learn sales agents' behavior. In Figure 1, we observe a relative change in impression allocations as a result of perturbations in log-scaled utilization, $-\log(1 - U) - (-\log(1 - \bar{U}))$, for a specific inventory category. The transformation $-\log(1 - x)$ describes the sales dynamics where the popularity and, therefore, high utilization of certain inventory induces larger allocations in negotiated deals. This dynamic corresponds to the regime of *near-capacity loading*. Also, one could consider the scaling function to be $-\log x$, which scores incremental change in low utilization more than at high utilization (the so called *herd* effect). Further, in Figure 2, we observe substitution effects, i.e., how changes in utilization of one impression category impacts the allocation in others.

Finally, entries in $n \times 1$ vector E represent noise. We assume that all these entries have zero mean, that they are mutually independent and independent from dp and dU^* . However, note that biases coming from variability among different advertisers and changing 'external' conditions, etc., will be diminished by a 'proper' experiment design as explained in Section 3. Then, assuming that $X \triangleq [1 \ dp \ dU^*]$, we use ordinary least squares (OLS) ([14]) to obtain the linear model coefficients:

$$B = (X^T X)^{-1} X^T dM. \quad (12)$$

OLS theory also provides the estimate of the variance-covariance

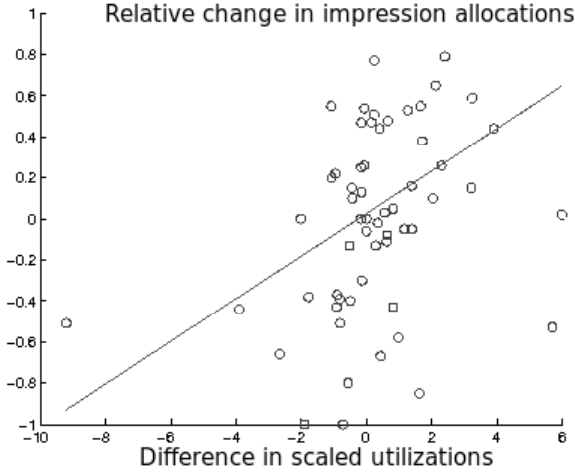


Figure 1: Capturing inventory allocation dynamics as a function of inventory utilization.

matrix of column j of B (j th inventory):

$$S_j = s_j^2 (X^T X)^{-1}, \quad (13)$$

where s_j^2 estimates the residual mean square error. Estimates of the covariances between two columns j_1 and j_2 of B , say B_{j_1} and B_{j_2} , depend on the correlation of residuals between the respective pair of columns of dM . Likewise, the covariance matrix of B_{j_1} and B_{j_2} has the same form:

$$\text{Cov}(B_{j_1}, B_{j_2}) = s_{j_1, j_2} (X^T X)^{-1}, \quad (14)$$

replacing the scalar value s_j^2 with the scalar covariance s_{j_1, j_2} .

2.3 Price Adjustments and Revenue Effects

Note that the Lagrangian gradient (5) and revenue gradient (7) are linear functions of elasticities $B = (b_{jk})$. Then, in view of (5), we can express its gradient as

$$g_k \equiv \frac{\partial \mathcal{L}(\bar{\lambda}, \bar{p})}{\partial q_k} = B_k \mathbf{c},$$

where $\mathbf{c}^T = (c_1, \dots, c_K)$ satisfies

$$c_k \triangleq \sum_{j=1}^K C_j U_j p_j (\delta_{jk} + 1 - U_j);$$

we use δ_{jk} to denote the function that is equal to 1 when $j = k$, and 0 otherwise. The variance of the gradient estimate g_k is obtained using

$$\text{Var}(g_k) = \mathbf{c}^T \text{Var}(B_k) \mathbf{c}, \quad (15)$$

where $\text{Var}(B_k)$ follows from (14).

The adjusted price vector \bar{p} that maximizes the positive change in the value of the Lagrangian function directly follows from (5) and is obtained by multiplying the current prices p_k , $1, \dots, K$, by

$$\mu_k \triangleq e^{\mu \frac{\sum_{j=1}^K C_j U_j p_j (\delta_{jk} b_{kk} + (1 - U_j) b_{jk})}{\sum_{j=1}^K C_j U_j p_j}}. \quad (16)$$

The unindexed factor μ which represents a desirable percent-increase of the Lagrangian value controls the overall *magnitude* of price changes. Usually μ is reserved as a business

decision, while the relative sizes, μ_i/μ_j , the *directions*, are determined by the gradient presented above.

Note that in the case when processing experimental results gives high coefficient of variation for the estimated gradient, i.e.,

$$\frac{\sqrt{\text{Var}(g_k)}}{g_k} > \eta,$$

for some choice of η (decision parameter), instead of adjusting prices and in order to increase the accuracy, publisher might decide to extend the experimentation. We include a way to conduct this extension in a statistically rigorous manner in Subsection 3.2.

3. PRICING EXPERIMENT

In order to provide gradient estimates discussed in Subsection 2.2, one needs to create disjoint experimental (cohort) groups of advertisers. Each group has assigned experimental prices $(\bar{p} + \Delta \bar{p})$ for K inventories.

The experimental design methodology incorporates procedures to create experimental groups and is highly dependent on the constraints of the problem. They should contain *statistically similar* mix of advertisers in terms of their baseline profiles. The idea is that such groups respond similarly to similar price changes. Most commonly, in reservation-based online advertising business, an advertiser's profile is determined by its annual budget with the publisher, and types of inventory previously purchased.

A statistically rigorous assignment of advertisers into, say G , pricing groups requires: (i) Features describing the advertisers, which need to be measurable in the pre-experimental (or baseline) period. Most commonly, these features include advertisers' 'typical' budgets and inventory preferences, i.e., average proportion of their budget spent on different inventory categories; (ii) Mapping of advertisers to corresponding sales people. In practice, the sales people of a given agency need to be assigned to exactly one pricing group, and they carry all their advertiser accounts with them into that group; (iii) An optimality criterion, which evaluates any proposed experimental design against any other; (iv) The combinatorial algorithm to do the assignment.

Details on the statistically rigorous procedures for creating statistically similar experimental groups an interested reader can find in [13] and [21]. In practice, a simple random assignment of sales agents to groups gives satisfactory results when the number of such agencies/advertisers is large. In contrast, the greatest benefits to more deliberate algorithms of [13] and [21] occur when the numbers of such agencies is not large.

3.1 Fractional Factorial Design

In general, the complexity of any experimental design depends on several factors:

- It depends on the choice of the *underlying model* where the overall goal plays an important role. This determines the key responses and the kinds of experimental variables to be controlled.
- The design depends on the *number of experimental variables* (here, inventory prices) to be varied. In order to estimate the gradient with respect to changes in K prices, there is a minimum number of experimental groups required (see the discussion below). Then,

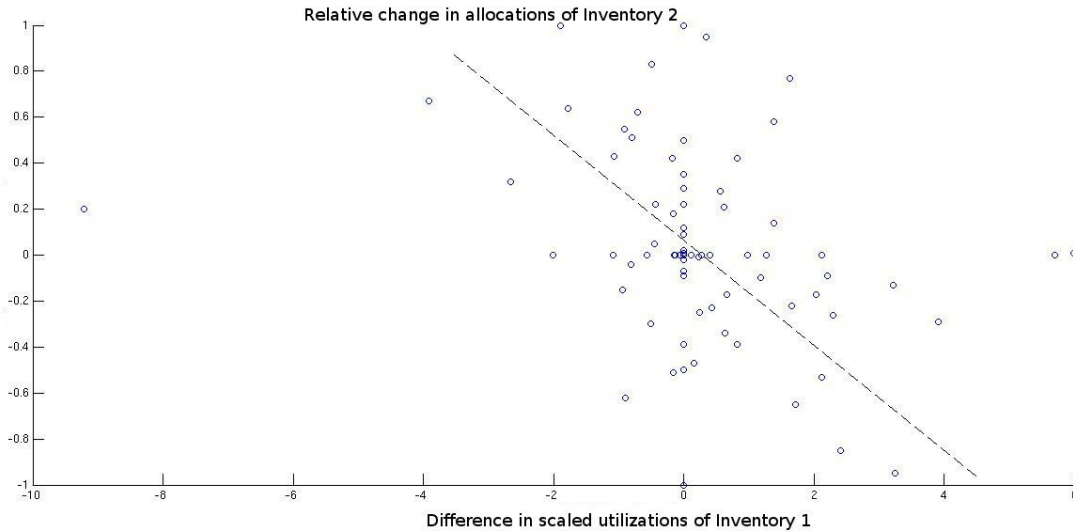


Figure 2: Capturing inventory substitution effects.

it is valuable to add the possibility of *extending* the experiment to a longer time (see Section 3.2 below).

- The current system needs to have the basic capacity to administer the design of the proposed complexity, i.e., it needs to be *easy to implement*.
- The design highly depends on the *underlying noise* of the experimental process.

As discussed before, our goal is to measure the first order effects, $b_{ij} = \partial \log M_j(\bar{p}) / \partial \log p_i$, of changes in log inventory allocation $\partial \log M_j(\bar{p})$ as a result of changes in K prices. In view of the analysis in Subsection 2.2, we define dp as the design matrix of price changes; it can also contain a column corresponding to the time effects – in case we extend the experiment to its second phase. In this paper we do not focus specifically on incorporating time drifts, even though it would be a straightforward extension of the current model. Each experimental group is exposed to a given vector of price changes and, after running the experiment, the responses from all experimental groups are aggregated for the estimation analysis. The method of exposing specific price vectors to specific experimental groups is the focus of the experiment design.

As an example, assume that we have three differently priced inventories. In order to explain what we mean by the efficient design, observe the following setup. Denote no price change, i.e. $\Delta \log p_i = 0$, by '0', and a price increase, i.e. $\Delta \log p_i = x > 0$, by '+'. Then, in Table 1, we consider four experimental groups, changing two prices at a time.

In the 3-dimensional space of prices, Table 1 design is symmetric, with design points at opposite corners of each face of the cube. With such a design, the effect of price i can be calculated as the average of groups 0 and j minus the average of the other two groups. Thus, all four groups are used in estimating each price elasticity. This latter observation, together with the smaller determinant and the design-space symmetry, indicate that this design is *maximally efficient*. The variance-covariance matrix for Table 1 gives zero correlation among the price elasticity estimates and is a manifestation of a deeper property called *orthogonality*.

In general, with K prices, the most theoretically scalable design has $K + 1$ groups since we are fitting the linear model with K first-order terms corresponding to prices, i.e., $b_{ji}, i = 1, \dots, K$ (factors next to $\Delta \log p_i, i = 1, \dots, K$, in the linear regression model), and one term corresponding to the baseline level or intercept. In the framework of the fractional factorial design, number of experimental groups, say G , is rounded to the smallest $G = 2^{K-p} \geq K + 1$. This is equivalent to achieving maximally efficient designs by taking $G \geq 4 \times \lceil (K + 1)/4 \rceil$ groups. Also, one can extend such designs to estimate terms beyond the first order experimental factors (here, price elasticities). For example, we could estimate higher order terms (a way to model non-linear effects) by having factors that correspond to $\Delta \log p_i \Delta \log p_j$ in the model (8). For more details on these types of design, the reader is referred to Chapter 6 of [8].

Price perturbations are business decisions, balancing customer impact with its potential revenue benefit. Assuming that such price changes are consistent with the goals of the study, larger price changes in the experiment offer the statistical advantage of requiring less extrapolation.

3.2 Two-Stage Experiment

The proposed experiment design provides an option for extending experimentation into a second phase. Using the second phase of experimentation explicitly doubles the size of the data set and thereby increases the precision of the first period results (gradient estimate). Prices in the second period change in a prescribed way, as it will be shown below, and estimating sensitivities involves differencing with respect to the first-period results from the same advertiser. Apart from increasing the amount of data points, the second stage reduces advertiser effects since differences in log impression allocations are computed with respect to the same advertiser. Furthermore, the second-stage in the experimentation allows for measuring time drifts and, therefore, provides a way to estimate seasonal-type effects. This seasonal-type effect is estimated orthogonally (i.e., with zero correlation) to the price elasticity effects.

The main disadvantage of the two-period design is that

Table 1: One example of price changes in experimental groups.

Exp. Group	constant	Inventory 1	Inventory 2	Inventory 3	role
0	1	+	+	+	change all prices
1	1	+	0	0	change price 1 from baseline
2	1	0	+	0	change price 2 from baseline
3	1	0	0	+	change price 3 from baseline

it can double the time required to complete the experiment and, very often, this is not a feasible option in the way a specific business is conducted.

The second phase design matrix is obtained from the first phase by reversing the modality of all price perturbations. For example, in the case of the example in Table 1 of the previous subsection, the second-stage matrix would be as in Table 2.

4. NUMERICAL EXAMPLE

In this section we provide an example on how to use the previously described experimental methodology to compute the revenue gradient. Since disclosing publisher’s transaction information is strictly forbidden, we simulate dynamics of the sales process using the understanding gained through the analysis of the real data. Note that the described procedure is used to perform one adjustment step in our algorithm proposed in Subsection 2.1. Each of these adjustment steps lasts long enough to collect enough transaction data to be able to adjust prices with a desirable accuracy (as explained in Subsection 2.3).

Consider a publisher who sells three types of premium inventory: (i) type 1 - low-cost (\$6 for 1000 impressions), (ii) type 2 - moderately priced (\$8 for 1000 impressions), and (iii) type 3 - high-value (\$15 for 1000 impressions). We assume that the overall capacities of these inventories are $C_1 = 1.37 * 10^9$, $C_2 = 1.013 * 10^9$ and $C_3 = 1.042 * 10^9$ for type 1, 2 and 3, respectively.

Advertisers (clients) that are buying the inventory from the publisher are split into three main groups (buckets) based on their average budget invested in a transaction: (i) bucket 1 - (\$0, \$50000), (ii) bucket 2 - (\$50000, \$100000), and (iii) bucket 3 - (\$100000, \$500000). Within each of the three buckets, advertisers differ in inventory preferences, as described in Table 3 below. For example, an advertiser that belongs to bucket 1 has one of the three possible inventory preference vectors. Furthermore, we assume that all advertisers within a given bucket range have the same arrival rate and transaction durations as it is given in Table 3.

Next, we assume that different advertisers arrive to a publisher according to a Poisson process in time, purchasing and reserving the inventory for some random, exponentially distributed, duration (campaign length). We assume that arrivals of different advertisers are mutually independent, and that transaction durations are independent from each other and from the arrivals. Therefore, using the information provided in Table 3, one is able to compute average inventory utilization, which we seek to compute in real systems by looking at the pre-experimental history of transactions. Therefore, given the demand parameters and inventory capacities, average inventory i utilization, say U_i^0 , can

be obtained as

$$U_i^0 = \frac{1000}{p_i} \times \frac{\sum_{a \in \mathcal{A}} (\text{av. budget})_a \times (\text{arr. rate})_a \times (\text{av. dur.})_a \times (\text{i pref.})_a}{C_i},$$

implying that $U_1^0 = 0.165$, $U_2^0 = 0.578$ and $U_3^0 = 0.66$.

Given the arrival/departure dynamics described above, one is able to simulate an arrival process of advertisers, as well as the corresponding campaign lengths, incoming budgets and inventory preferences. In addition to these values, DART for Publisher’s database keeps records of negotiated inventory allocations (M_{a1}, \dots, M_{aK}) and available inventory (which we use to compute their utilization (U_1, \dots, U_K)) at the moments of transactions. These quantities depend on sales dynamics, which we model as follows.

When an advertiser arrives at the publisher, the assigned sales agent starts negotiating the deal. As shown in Table 3, we assume four advertisers of each type, which makes the process of splitting them into four experimental groups easy. There are many heuristics which one can choose to reach the so called D/D1-optimality criteria in more general cases, described in [13], allowing us to omit that discussion here.

Advertisers:

In this simulation, we assume that advertisers tends to keep their budgets fixed from transaction to transaction. However, oftentimes when a discount in price of their preferred inventory is mentioned, the advertiser responds by increasing his/her budget. We model this phenomenon by assuming that in the case of inventory discount (advertiser is exposed to the lower experimental inventory price when compared to the pre-experimental one) the advertiser’s budget increases with probability 0.5. When the increase happens, its actual value is chosen uniformly at random from the range between 0% and 20%.

Sales agents:

Sales agents tend to keep identical inventory allocations from transaction to transaction with the same advertiser. However, when the price of the preferred inventory is discounted, its allocation tends to be larger. We model this effect by increasing allocations by 20% and 50% for inventories 3 and 2 respectively, when compared to the corresponding pre-experimental average values. Apart from being influenced by price, sales agents’ allocation decisions depend on the inventory utilization at the moment of transaction. We assume that sales agents tend to allocate larger amounts of inventory if it is ‘popular’, believing that they make advertisers ‘happy’ this way. We model these nonlinear effects by assuming that changes in allocations are proportional to the change in the scaled inventory utilization, where utilization u is scaled as $-\log(1 - u)$. In simulating the previously described phenomena, we assume that a sales agent first starts allocating the most expensive inventory from the advertiser’s

Table 2: One example of price changes in experimental groups.

Exp. Group	constant	Inventory 1	Inventory 2	Inventory 3	role
0	1	0	0	0	change all prices
1	1	0	+	+	change price 1 from baseline
2	1	+	0	+	change price 2 from baseline
3	1	+	+	0	change price 3 from baseline

Table 3: Profiles of incoming advertisers.

Budget	Inv. Pref. (% of budget)	Daily Rate	Av. Duration (days)	Num. of Advert.
(\$ 0, \$50000)	(80%, 20%, 0%)	1/2	5	4
(\$ 0, \$50000)	(50%, 50%, 0%)	1/2	5	4
(\$ 0, \$50000)	(0%, 100%, 0%)	1/2	5	4
(\$ 50000, \$100000)	(0%, 50%, 50%)	1/8	14	4
(\$ 50000, \$100000)	(25%, 50%, 25%)	1/8	14	4
(\$ 50000, \$100000)	(0%, 25%, 75%)	1/8	14	4
(\$ 100000, \$500000)	(0%, 0%, 100%)	1/10	30	4
(\$ 100000, \$500000)	(0%, 50%, 50%)	1/10	30	4
(\$ 100000, \$500000)	(0%, 25%, 75%)	1/10	30	4
(\$ 100000, \$500000)	(25%, 25%, 50%)	1/10	30	4

Figure 3: Inventory allocation dynamics as a function of perturbations in its scaled utilization.

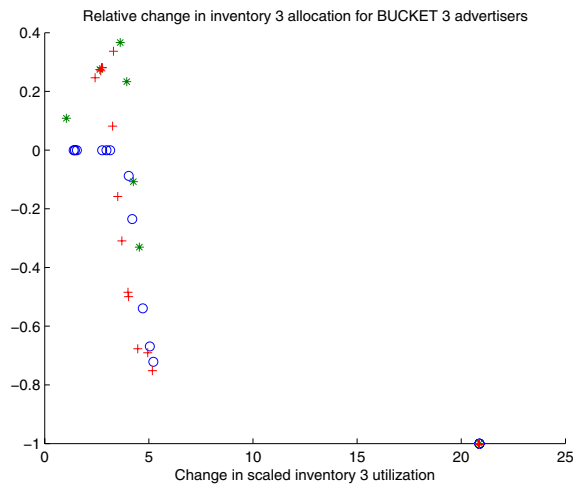
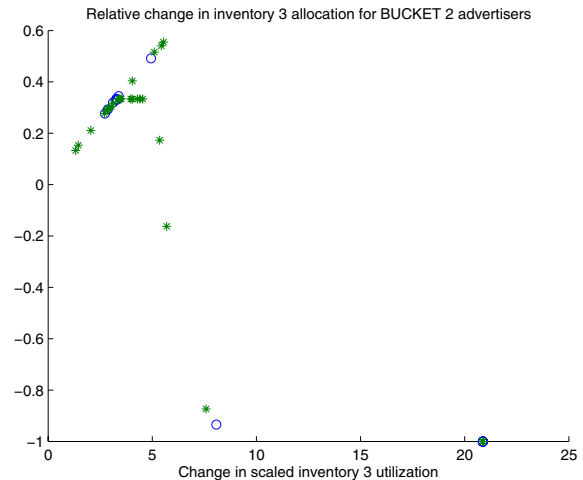


Figure 4: Inventory allocation dynamics as a function of perturbations in utilization of a substitute inventory.



preference list, while adjusting allocation decisions in case of price discounts and higher utilization. In case there are not enough inventory in stock, the rest is allocated from the other preferred inventory that is the closest in price to this one. In case there is not enough of the cheaper inventory, we allocate the lowest cost with the leftover budget. The cheapest inventory is usually much less utilized, and, in practice, never becomes unavailable.

Finally, we simulate dynamics using the description above and analyze the collected results that correspond to 90 days of running the experiment. As an example, in Figures 3 and 4, one is able to observe data points which correspond to the effects we described above: an increase in inventory allocation due to the price discount, higher allocations due to higher utilization, reduced allocation due to a lack of available inventory. Figure 3 captures allocations to the most

valuable, bucket 3, advertisers, while Figure 4 captures allocations to the advertisers belonging to bucket 2.

Using (12), (7) and $\Delta \log R_k(\bar{p}) \approx \frac{R_k(\bar{p} + \Delta \bar{p}) - R_k(\bar{p})}{R_k(\bar{p})}$ for small price perturbations (see (6)), we compute that the log-revenue elasticity vector is

$$\left(\frac{\partial \log R(\bar{p})}{\partial \log p_1}, \dots, \frac{\partial \log R(\bar{p})}{\partial \log p_3} \right) \approx (2.5, 1.5, 3.1),$$

implying that increasing prices of all three inventories increases revenue as well. Confidence (standard error) of the obtained estimates are 0.042, 0.2 and 0.17, respectively, which are obtained using the analogous expression as in (15), but for the revenue instead the Lagrangian function. Of course, in cases where it is required to obtain 'better' confidence, one can always extend the experiment using the methodology presented in Subsection 3.2. Finally, using (16), the

publisher should set the new price vector to be

$$(\$6e^{2.5\mu}, \$8e^{1.5\mu}, \$15e^{3.1\mu}),$$

where μ is a scalar that depends on publisher's business strategy.

5. CONCLUSIONS

In this paper we present a methodology for adjusting inventory prices with the objective of increasing (maximizing) revenue in display advertising. We cast this problem as an inventory allocation problem, where impression assignments that are performed by sales agents in the process of sales are controlled by price as well as the current inventory utilization. Since the demand curve (inventory allocation as a function of its price) is not known, the only feasible way to change prices in the direction of the revenue increase is by performing an experiment. Under *mild* assumptions on the inventory demand and *no* distributional assumptions on any of the parameters, we present a first of a kind, practically feasible and statistically rigorous revenue maximization methodology that incorporates (i) iterative algorithm for adjusting inventory prices, and (ii) a statistically rigorous method for estimating elasticity of the demand curve at each price point. Our fractional experiment design takes control over various biases that could impact impression allocations, such as: variation in advertiser inventory requirements and budget constraints, changes in the current economy, seasonalities and management decisions. Furthermore, our experimental method is simple to execute and maximally efficient in the presence of small data volumes, which are all *desirable* properties in convincing publishers to adopt the methodology.

6. REFERENCES

- [1] B. M. 2009. Advertising age - kill or cure: How reekitt's big buy rocked online ads. <http://breakmedia.break.com>, 2009.
- [2] M. Advertising. Display advertising. <http://advertising.microsoft.com/display-advertising>, 2011.
- [3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- [4] T. Amemiya. Qualitative response models: a survey. *Journal of Economic Literature*, XIX:1483–1536, 1981.
- [5] S. Balseiro, J. Feldman, V. Mirrokni, and S. Muthukrishnan. Yield optimization of display advertising with ad exchange. In *ACM Conference on Electronic Commerce*, 2011.
- [6] V. Bharadwaj, W. Ma, S. M., J. Shanmugasundaram, E. Vee, J. Xie, and J. Yang. Pricing guaranteed contracts in online display advertising. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, Toronto, Canada, October 2010.
- [7] T. O. G. Blogspot. Online publishers: growing the display advertising pie. <http://googleblog.blogspot.com/2010/08/online-publishers-growing-display.html>, 2010.
- [8] G. Box, S. Hunter, and H. W. *Statistics for experimenters: design, discovery, and innovation*. John Wiley and Sons, 2005.
- [9] P. Cramton, Y. Shoham, and R. Steinberg. *Combinatorial Auctions*. MIT Press, 2006.
- [10] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.
- [11] J. Feldman, M. Henzinger, N. Korula, V. Mirrokni, and C. Stein. Online stochastic packing applied to display advertising. In *ESA*, volume 1, pages 182–194, 2010.
- [12] K. Fridgeirsdottir and S. Najafi-Asadolahi. Cost-per-click pricing for display advertising. *Working paper*, <http://phd.london.edu/snajafi.phd2005/papers/snajafi-paper2.pdf>, 7(1), 2011.
- [13] W. Heavlin and G. Finnegan. Columnwise construction of response surface designs. *Statistica Sinica*, 8:185–206, 1998.
- [14] J. Johnston and J. DiNardo. *Econometric methods*. McGraw-Hill, 1997.
- [15] J. Liechty, V. Ramaswamy, and S. Cohen. Choice-menus for mass customization: An experimental approach for analyzing customer demand with an application to a web-based information service. *Journal of Marketing Research*, 38(2):183–196, 2001.
- [16] A. Mangani. Online advertising: Pay-per-view versus per-per-click. *Journal of Revenue and Pricing Management*, 2:295–302, 2003.
- [17] McAfee, P. Papineni, and S. Vassilvitskii. Maximally representative allocations for guaranteed delivery advertising campaigns. *Working paper, Yahoo Research*, 7(1), 2010.
- [18] S. Muthukrishnan. Ad exchanges: Research issues. In *Proc. of the 4th Inter. Workshop on Internet and Network Economics (WINE)*, 2009.
- [19] T. Novak and D. Hoffman. Advertising and pricing models for the web. In *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, D. Hurley, B. Kahin and H. Varian (editors), MIT Press, Cambridge, 2000.
- [20] O. Shy. *How to price: a guide to pricing techniques and yield management*. Cambridge University Press, 2008.
- [21] S. Silvey. *Optimal design*. Chapman and Hall, 1980.