Adverse Selection and Auction Design for Internet Display Advertising

Nick Arnosti, Marissa Beck and Paul Milgrom

Working Paper: September 2, 2014

"Half the money I spend on advertising is wasted; the trouble is, I don't know which half."
- John Wannamaker (pioneering brand advertiser)

Abstract

We model an online display advertising environment with brand advertisers and better-informed performance advertisers, and seek an auction mechanism that is strategy-proof, anonymous and insulates brand advertisers from adverse selection. We find that the only such mechanism that is also false-name proof assigns the item to the highest bidding performance advertiser only when the ratio of the highest bid to the second highest bid is sufficiently large For fat-tailed match-value distributions, this new mechanism captures most of the gains from good matching and improves match values substantially compared to the common practice of setting aside impressions in advance.

1 Introduction

Since the pioneering papers by Edelman, Ostrovsky, and Schwarz (2007) and Varian (2007), there has been a growing body of research focusing on auctions for sponsored search advertising on the Internet. These automated auctions, which are initiated when a consumer enters a search query, allow the winning advertisers to post ads near unpaid search results. When it works well, this auction process creates economic value by directing consumers to websites where a mutually profitable transaction can occur.

Markets for internet display advertising, which determine ad placements on all other kinds of pages, tend to operate quite differently. With no search phrase entered by the consumer, advertisers must try to identify good matches using other kinds of information, such as the content of the webpage (is it a gaming site or a newspaper's financial page?), as well as whatever can be inferred from cookies on the consumer's computer. Additionally, whereas bidders in sponsored search markets are nearly always "performance advertisers" placing ads designed to elicit an immediate performance (such as a click or a purchase), many display ads are placed by "brand advertisers," who are instead publicizing a product or an upcoming event. Finally, rather than paying different prices for each impression based on its individual characteristics, brand advertisers commonly often negotiate fixed-price contracts that require the publisher to deliver millions of impressions over a stated period of time to some target audience.

These differences raise several questions. Why are fixed-price contracts common for display advertising, when auctions are universally used for sponsored search? What are the benefits and costs of using these contracts, rather than auctions, for display advertising? Taking a market design perspective: Is there a different way to organize the market that might make all participants better off?

In this paper, we propose answers to each of these questions. We argue that information asymmetry is far more problematic in display advertising than in search advertising. We develop a simple model in which the current market organization shields brand advertisers from adverse selection,[1] but also causes many impressions to be allocated inefficiently. Motivated by this finding, we propose a novel auction design which aims to recover much of the efficiency provided by standard second-price auctions while simultaneously ensuring that brand advertisers are delivered a representative sample of impressions.

Why is adverse selection a particular risk for brand advertisers? Our hypothesis is that nearly all advertisers value certain consumer characteristics, such as high income and high responsiveness to online advertisements. These correlations in the impression's value across advertisers introduce a form of the "winner's curse," whereby the brand advertiser (who often has limited access to private cookies) wins only those impressions on which better-informed performance advertisers choose to bid low. Furthermore, as highlighted by the John Wannamaker quotation that opens this paper, a brand advertiser who wants consumers to come to a special sale or a movie opening next weekend may be unable to measure the effectiveness of individual ads, making it difficult for the advertiser to identify, monitor and manage the potential adverse selection problem on any particular website.[2]

Contracts that set aside some impressions for exclusive use in brand advertising can protect brand advertisers from adverse selection, although they also cause ad space to be allocated inefficiently. For example, a local Ford dealer who reserves the top-of-page banner impressions between 4pm and 8pm on a newspaper's automotive page faces no adverse selection. It can happen, however, that the Ford ad is shown to a user whose cookies identify him as a 19-year-old male in California who recently visited Amazon.com to search for skateboards. Amazon's value for this impression is likely to be much higher than Ford's, but the set-aside contract bars that more efficient allocation. Selecting a consumer for ad-targeting based on such browsing behavior would come at little or no cost to Ford, so long as another demographically similar impression was used to fill Ford's contract.

[1]Abraham, Athey, Babaioff, and Grubb (2013) also study adverse selection in Internet display advertising, but without posing this market design question.

[2]On the difficulty of measuring display ad performance, see Lewis and Reiley (2010).

To get a sense of the potential magnitude of this matching inefficiency, consider the following (hypothetical) example. Suppose that if the New York Times held an auction for every impression on its auto page over a certain period, it would collect \$5 million in revenue. Suppose, however, that 80% of these impressions are claimed in fixed-price contracts, so that only 20% of them will be auctioned to performance advertisers. If the impressions auctioned to performance advertisers are selected at random, they will likely generate about \$1 million in revenue.

The marketer's 80-20 rule-of-thumb states that 80% of the value of performance ads comes from the 20% of impressions that are the advertisers' best matches.[3] This suggests that if the New York Times could identify the "right" 20% of impressions to give to performance advertisers (for example, by running an auction for each impression and using those that generate the least interest to fill existing contracts), they could collect \$4 million in revenue from performance advertisers while still delivering the same number of impressions to brand advertisers – a four-fold increase in revenue!

We formalize the observation that set-asides lead to large losses through a simple model in Section 3. Theorem 1 states that in our model, among all mechanisms that deliver the required number of impressions to brand advertisers, the random set-aside has the *lowest* performance advertiser surplus and the *lowest* publisher revenue from performance advertisers. Any bidding mechanism that allows performance advertisers to win impressions that are more valuable to them seems to risk introducing adverse selection, which the set-aside contract avoids.[4]

Is there a way to capture much of the match value lost by set-asides while still avoiding the problem of adverse selection? The general answer must be "no," for if across all impressions the brand advertiser and the performance advertiser values were identical, then any value improvement for performance advertisers must be offset by a corresponding loss for brand advertisers. This example, however, is extreme and unrealistic. While an online store like Amazon might value a consumer due to her high income, it might have an even higher match value for a consumer with a lower income who has recently searched for a particular book or left several items in her Amazon shopping cart.

In Section 4, we aim to capture the idea that impressions can be valuable for these two quite different reasons by representing the value of each impression using two kinds of quality variables.

[3] See Wikipedia's entry on the "Pareto principle": `http://en.wikipedia.org/wiki/Pareto_principle`.

[4] McAfee, Papineni, and Vassilvitskii (2013) describe an algorithmic system recently designed and implemented by *Yahoo!* to use auctions in place of fixed price contracts. This system places random bids on behalf of brand advertisers, with the goal of acquiring the "most representative" sample of impressions – measured in terms of the highest performance bid – that the advertiser can afford, given its budget constraint. This does not, however, address the design tradeoff posed above. Conditional on the brand advertiser's randomized bid being very high (or very low), the bid nearly always wins (or nearly always loses), so this subset of bids leads to nearly random matching and does not improve match quality. Conditional on the bid taking an intermediate value, some matching value is added, because performance advertisers can then buy their most valuable impressions, but then the brand advertiser suffers from adverse selection. Thus, regardless of whether the *Yahoo!* randomization leads to a bid that is high, medium or low, the market outcome, conditional on the bid, is always gored by at least one of the two horns of this market design dilemma.

The first is the "consumer quality" variable C, which is proportional to the value of the impression to the brand advertiser; the second is the "match quality vector" M, which is the performance advertiser values divided by the consumer quality C. For efficient assignment of ad space, the impressions with the highest match quality should be assigned to performance advertisers. To the extent that the assignment is correlated with consumer quality, there is a danger of adverse selection.

We study efficiency and adverse selection among auctions that are strategy proof, anonymous, deterministic, and weakly efficient (meaning that only the highest performance bidder can win the auction). We will say that such an auction is *adverse-selection free* if, for every joint distribution of (C, M) such that C and M are statistically independent, the event that the impression is allocated to the brand advertiser is independent to C. It is *false-name proof* if no bidder can benefit from submitting multiple bids and the publisher cannot raise a winner's price by submitting very low bids.

Theorem 2 asserts that there is a unique auction mechanism that is deterministic, adverse-selection free and false-name proof and delivers the right fraction of the impressions to brand advertisers: it is the modified second-bid (MSB) auctions of Milgrom, Cunningham, and Beck (2012).[5] In an MSB auction, an impression is awarded to the brand advertiser (at its contract price) if and only if the ratio of the highest performance bid to the second highest is less than some factor $\alpha \geq 1$. If the ratio exceeds α, then the item is awarded to the highest performance advertiser at a price equal to α times the second-highest bid. The parameter α can be set so that the brand advertiser acquires any desired fraction of the available impressions.

Why does the MSB mechanism work? By inspection, it is strategy proof and anonymous among performance advertisers. When two bidders are already present, a winning bidder cannot reduce its price by submitting additional bids, and a seller cannot increase the price by submitting additional, very low bids. Those are the conditions that here define false-name-proof mechanisms. What is most interesting is that MSB auctions are also adverse-selection free. By design, the mechanism takes advantage of the following observation: the profiles of bids encountered when there is high consumer quality differ from those that occur when an advertiser has a high match quality. When consumer quality is high, all bids tend to rise proportionately, but high idiosyncratic match quality may often be signified by one bid being much higher than the others. When the consumer quality is independent of the match quality vector, then the fact that MSB determines the brand advertiser's allocation using the ratio of the highest bid to the second-highest bid implies that there is no statistical relationship between the allocation and the quality of the consumer.

That still leaves an important question: how much of the match quality lost by a set-aside can be recaptured by use of an MSB mechanism? We study that question formally in Section

[5]If we drop the assumption that the auction is deterministic, then other satisfactory mechanisms include random set-asides (RSA) and certain hybrids of RSA and MSB.)

5, but an intuitive analysis is suggested by the preceding discussion: we should expect an MSB auction to work well when a large fraction of the expected maximum match quality is associated with the signature event that one bid is much higher than all of the others. To explore when this property may hold in a formal model as well as its implications for performance, we assume that the match quality variables, M_n, are IID draws from some distribution with a fat-tail property. Specifically, we assume that the distribution of $\ln(M_n)$ has a non-increasing hazard rate. With this specification, the permissible distributions for M_n include the power law distributions and mixtures of such distributions, as well as some others.

In Section 5, we show that the performance of MSB auctions for this class of distributions is very good. In particular, when compared to the set-aside mechanism, the MSB mechanism using parameter α multiplies both the revenue from performance advertisers and the performance advertiser profits by a factor of at least α, without reducing the value of impressions allocated to the brand advertiser.

How costly is the restriction to adverse-selection-free, false-name-proof mechanisms? We address this question in Section 6. Theorem 4 states that if the consumer quality is known by the publisher and match values for performance advertisers follow a power law distribution, then the use of an MSB auction yields revenues and bidder profits similar to those of the standard second-price auction. This means that even if the publisher could eliminate the winner's curse using information about the quality of each consumer, the gain from doing so (relative to running an MSB auction, which does not require this information) would be relatively small.

Section 7 concludes. All omitted proofs are presented in the Appendix.

2 Preliminaries

This section introduces our notation and definitions and characterizes the relevant strategy-proof mechanisms for the display advertising problem. We limit attention to direct mechanisms, but we suppress the adjective "direct" in our definitions and descriptions.

Throughout, we use capital letters to represent random variables, and lower case letters to represent their realizations. In our model, a random number N of performance bidders participate in an auction. We assume throughout that $P(N \geq 2) = 1$. Let $x = (x_i, x_{-i})$ denote the profile of performance bids, where x_i is i's bid and x_{-i} denotes the remaining $N-1$ bids.

We limit attention to auction mechanisms in which only the winner pays and we assume that all participants are risk neutral. The constraint on payments is standard in practice for reasons not modeled here: it has no consequences for the analysis of incentives, efficiency, or welfare below. We define a mechanism to be a pair (z, p) such that, for any realized number of performance bidders $n \geq 2$ and bid vector $x \in \mathbb{R}_+^n$,

1. $z(x)$ is an $(n+1)$-dimensional probability vector with $z_i(x)$ denoting the probability that

bidder i wins and $z_0(x) = 1 - \sum_{i=1}^{n} z_i(x)$ denoting the probability that the brand advertiser wins,

2. $p(x)$ is a vector in $\mathbb{R}_+^n$ with $p_i(x)$ denoting bidder i's expected payment,

3. for all $i \in \{1, \ldots, n\}$, $z_i(x) = 0 \Rightarrow p_i(x) = 0$,

4. for all $i \in \{1, \ldots, n\}$, $z_i(x)$ is left continuous in x_i, meaning that if x_i^k is an increasing sequence converging to x_i, then $z_i(x_i^k, x_{-i}) \to z_i(x_i, x_{-i})$.

Condition 1 implies that the impression cannot go unassigned, which can rule out the expected-revenue-maximizing auction. Condition 3 states that only winners make or receive payments. In this paper, we focus on dominant strategy direct revelation mechanisms, implying that each bidder faces a threshold price and wins the item if she is willing to pay this price. Condition 4 implies that a bidder must bid strictly above her threshold price in order to win the impression.

Definition 1. *The mechanism (z, p) is* **strategy proof** *if, for all $n \geq 2$, all $x \in \mathbb{R}_+^n$, all $i \in \{1, \ldots, n\}$, and all $\hat{x}_i$,*

$$x_i z_i(x) - p_i(x) \geq x_i z_i(\hat{x}_i, x_{-i}) - p_i(\hat{x}_i, x_{-i}).$$

Definition 2. *A mechanism (z, p) is* **anonymous** *(among performance advertisers) if, for any $n \geq 2$, any permutation σ on $\{1, \ldots, n\}$ and any $x \in \mathbb{R}_+^n$, the following hold:*

$$\sigma(z(x)) = z(\sigma(x)) \qquad \text{and} \qquad \sigma(p(x)) = p(\sigma(x)).$$

Definition 3. *A mechanism (z, p) is* **weakly efficient** *if, for all $n \geq 2$, all $x \in \mathbb{R}_+^n$, all $i \in \{1, \ldots, n\}$ and all $j \in \{1, \ldots, n\} \setminus \{i\}$,*

$$x_i < x_j \Rightarrow z_i(x) = 0.$$

So, a mechanism is weakly efficient if, among performance advertisers, only the highest bidder may win the impression.

Definition 4. *Let $W \sim U[0, 1]$ be a random variable that is independent from the number of performance bidders and from their valuations. An anonymous mechanism (z, p) is a* **threshold auction** *if there exists a function $h : [0, 1] \times \bigcup_{n=2}^{\infty} \mathbb{R}_+^{n-1} \to \mathbb{R}_+ \cup \{\infty\}$, symmetric in its arguments beyond the first, such that:*

I. bidder i wins if and only if $x_i > \max\{x_{-i}, h(W, x_{-i})\}$, and

II. if bidder i wins, he pays $\max\{x_{-i}, h(W, x_{-i})\}$.

In this definition, the random variable W is included to allow non-deterministic outcomes. In a threshold auction, a bidder wins if and only if he bids more than his threshold price. Note that any threshold auction is completely characterized by its "threshold price" function $\tilde{p} : [0,1] \times \bigcup_{n=2}^{\infty} \mathbb{R}_+^{n-1} \to \mathbb{R}_+$ defined by $\tilde{p}(w, x_{-i}) = \max\{x_{-i}, h(w, x_{-i})\}$. In what follows, whenever the threshold auction is deterministic, we omit the superfluous argument w from $\tilde{p}$. In this case, we have that for any n, any $i \in \{1, \ldots, n\}$, and any set of submitted bids $x \in \mathbb{R}_+^n$, $z_i(x) = \mathbf{1}_{x_i > \tilde{p}(x_{-i})}$ and $p_i(x) = z_i(x)\tilde{p}(x_{-i})$.

Lemma 1. *An anonymous mechanism (z, p) is strategy proof and weakly efficient if and only if it is a threshold auction.*

Two particularly simple examples of threshold auctions are second-price auctions with random set-asides and second-price auctions with a reserve. The former ignores all bids when deciding whether to allocate the impression to the brand advertiser and runs a simple second-price auction for the remnant. The latter sets a fixed reserve t and allocates an impression to the brand advertiser whenever no bid exceeds this reserve. We define these mechanisms formally below.

Definition 5. *A threshold auction is a* **random set-aside auction** *if there exists $\lambda \in [0,1]$ such that:*

$$\tilde{p}(w, x_{-i}) = \begin{cases} \max\{x_{-i}\} & w > \lambda \\ \infty & w \leq \lambda. \end{cases} \tag{1}$$

For any $\lambda \in [0,1]$, we denote by $RSA(\lambda)$ the threshold auction with threshold price given by (1)*.*

Definition 6. *A threshold auction is a* **second-price auction with reserve** *if there exists $t \geq 0$ such that:*

$$\tilde{p}(w, x_{-i}) = \max\{x_{-i}, t\}. \tag{2}$$

For any $t \geq 0$, we denote by $SPR(t)$ the threshold auction with threshold price given by (2)*.*

3 The Inefficiency of Random Set-Asides

In the introduction, we used the 80-20 rule to illustrate how set-aside mechanisms can lead to enormous losses of match value. In this section, we expand that critique by relaxing our distributional assumptions and expanding the comparison set to include all threshold auctions. For our calculations, we assume that match values are drawn IID from a distribution with cdf F. We compare the performance of alternative threshold auctions using the expected values of three welfare measures: (i) the total value generated by impressions allocated to the performance advertisers, (ii) the seller revenues from performance advertisers, and (iii) the net surplus to performance bidders. Of course, (i) = (ii) + (iii).

The conclusions, presented in Theorem 1, are stark. First, for *any* absolutely continuous distribution F, among the class of threshold auctions, the random set-aside *minimizes* the expected value generated by performance matches. Second, if the Myerson virtual value function associated with F is monotonic (i.e., F is regular), then the random set-aside also minimizes expected seller revenue. Third, if the hazard rate function associated with F is declining, then the random set-aside also minimizes the expected surplus of performance advertisers. By contrast, under the same assumptions, a second-price auction with reserve (hereafter SPR) *maximizes* these three performance measures.

For a precise statement, we introduce notation to correspond to the three measures of an auction's performance, (i), (ii) and (iii), described above. Let $X = (X_i)_{i=1}^N$ be the random vector of performance advertiser values and let (k) to denote the index of the k^{th} highest of these values, so that $X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(N)}$. For a fixed distribution over X, the relevant performance measures are $V(\tilde{p})$, $R(\tilde{p})$ and $S(\tilde{p})$. The first of these is the expected total value generated by impressions sold to performance advertisers when the publisher uses the threshold auction with pricing function $\tilde{p}$. Similarly, let $R(\tilde{p})$ and $S(\tilde{p})$ are the expected publisher revenue from performance bidders and the expected surplus enjoyed by performance bidders. In symbols,

$$V(\tilde{p}) = E\left[\sum_{i=1}^N X_i z_i(X)\right] = E\left[\sum_{i=1}^N X_i \mathbf{1}_{X_i > \tilde{p}(W, X_{-i})}\right] \tag{3}$$

$$R(\tilde{p}) = E\left[\sum_{i=1}^N p_i(X)\right] = E\left[\sum_{i=1}^N \tilde{p}(W, X_{-i}) \mathbf{1}_{X_i > \tilde{p}(W, X_{-i})}\right] \tag{4}$$

$$S(\tilde{p}) = E\left[\sum_{i=1}^N X_i z_i(X) - p_i(X)\right] = V(\tilde{p}) - R(\tilde{p}). \tag{5}$$

We use the following shorthand for conditional expectations (given the number of bidders): $E_n[\cdot] = E[\cdot|N = n]$ and $P_n(\cdot) = P(\cdot|N = n)$. We define $V_n(\tilde{p}) = E_n[V(\tilde{p})], R_n(\tilde{p}) = E_n[R(\tilde{p})]$, and $S_n(\tilde{p}) = E_n[S(\tilde{p})]$ to be the expected value, revenue, and bidder surplus given $N = n$, respectively.

The next theorem demonstrates that using these performance measures, among auctions that give the impression to the brand advertiser with the same probability λ, the RSA mechanism defined in (1) generates the least value among the class of threshold auctions, and the SPR mechanism from (2) generates the most. It also provides conditions under which corresponding statements can be made about revenue and bidder surplus.

Theorem 1. *Fix $N = n$ and suppose that the X_i are IID draws from a distribution with density f and corresponding cdf F, with support on an interval $[a, b) \subseteq [0, \infty)$ and with $E[X_i] < \infty$.*

Let $(\lambda, t, \tilde{p})$ satisfy

$$P_n\left(X_{(1)} \leq t\right) = P_n\left(X_{(1)} \leq \tilde{p}\left(W, X_{-(1)}\right)\right) = \lambda,$$

so that the probability that an impression is assigned to the brand advertiser is the same under all the mechanisms being compared. Then the following hold.

1. *The threshold auction given by $\tilde{p}$ generates at least as much value (from impressions allocated to performance advertisers) as the random set-aside auction with parameter λ and no more value (from these impressions) than is generated by the second-price auction with reserve t. In symbols,*
$$V_n(RSA(\lambda)) \leq V_n(\tilde{p}) \leq V_n(SPR(t)).$$

2. *If $x - \frac{1-F(x)}{f(x)}$ is a non-decreasing function, then the threshold auction given by $\tilde{p}$ generates at least as much revenue as the random set-aside auction with parameter λ and no more revenue than the second-price auction with reserve t. In symbols,*
$$R_n(RSA(\lambda)) \leq R_n(\tilde{p}) \leq R_n(SPR(t)).$$

3. *If $\frac{1-F(x)}{f(x)}$ is a non-decreasing function, then the threshold auction given by $\tilde{p}$ generates at least as much performance bidder surplus as the random set-aside auction with parameter λ and no more bidder surplus than the second-price auction with reserve t. In symbols,*
$$S_n(RSA(\lambda)) \leq S_n(\tilde{p}) \leq S_n(SPR(t)).$$

The intuition for this result starts with the observation that every threshold auction assigns the impression to the top performance bidder weakly more often when the highest performance value increases, but RSA makes that assignment completely at random. So, every threshold auction enjoys positive selection compared to RSA. Similarly, SPR enjoys positive selection with respect to any other threshold auction: the values of the winning bidders are higher. Those two observations account for the first statement of the theorem. The hypotheses of the second and third statements of the theorem require that the marginal contribution of assigning an impression to a performance bidder (to bidder profits in one case and seller revenues in the other) is increasing in the bidder's value. Then, positive selection works in the same way for those measures of performance as it does for total value. A proof of the theorem, found in the appendix, formalizes this intuition cleanly using majorization inequalities.

Given the bad theoretical performance of RSA, should publishers rush to adopt another threshold auction, such as SPR? Not necessarily. In general, we expect that alternative threshold auctions may boost the value of performance advertiser matches at the expense of brand advertisers – a fact which may make the adoption of such auctions undesirable in practice. We discuss this idea more formally in Section 4, where we introduce a model in which values of performance and brand advertisers are correlated and seek mechanisms that provide brand advertisers a "representative" sample of impressions while improving upon the value generated by RSA.

4 Derivation of a New Auction Design

In this section, we introduce a model of the bidding environment in which an ordinary second-price auction would cause adverse selection but a random set-aside would result in inefficient matching of impressions between brand and performance advertisers. We then define two conditions – the new condition of an *adverse-selection-free* mechanism and the more familiar one of a *false-name-proof* mechanism. Using this model, we show that there exists a unique deterministic threshold auction – the *modified second-bid* auction – that allocates a given fraction λ of impressions to the brand advertiser and is both adverse-selection free and false-name proof.

We suppress from our notation the publicly available information used by both brand and performance advertisers: the probabilities and expectations below should be understood to be conditional on that information. Then, fixing any particular impression, our notation is as follows: B is the value of the impression to the brand advertiser; N is the (possibly random) number of performance bidders; X_i is its value to performance advertiser i; $X = (X_1, ..., X_N)$; $C = E[B|X]/E[B]$ (so $E[C] = 1$); and $M = X/C$.

We will call C the "consumer quality" and think of it as a persistent consumer characteristic based on her income, responsiveness to online advertising, and so on. Similarly, we will call each $M_i = X_i/C$ a "match quality" and think of it as reflecting both persistent and transitory factors that affect i's value but not the value of the brand advertiser. Match quality may depend on the consumer's club memberships or recent browsing behavior, and on characteristics of advertiser i.

Informally, the brand advertiser is exposed to adverse selection whenever they are more likely to receive impressions with low values of C. No auction mechanism that uses bids to allocate impressions can eliminate adverse selection for every joint distribution of C and M, so we set a more modest goal: a mechanism is "adverse-selection free" if it avoids adverse selection whenever the consumer quality C and match quality vector M are statistically independent. This property is captured below in a definition. Although we describe some mechanisms as "adverse-selection free," the reader should remain aware that these mechanisms are only guaranteed to eliminate adverse selection if C and M are statistically independent. We discuss the performance of the mechanism when this condition is not satisfied at the end of the paper.

Definition 7. *A mechanism (z, p) is* **adverse-selection free** *if, for every joint distribution of (C, M) such that M is independent from C, $z_0(CM)$ is also independent of C.*

The following lemma characterizes deterministic adverse-selection free auctions.

Lemma 2. *A deterministic threshold auction with price function $\tilde{p}$ is adverse-selection free if and only if $\tilde{p}(x_{-i}) = \max\{x_{-i}, h(x_{-i})\}$ is homogeneous of degree one.*

Although threshold auctions are strategy proof when treated as direct mechanisms, in practice they can be vulnerable to other forms of manipulation. One is that a bidder may be able to take advantage of the anonymity of the web to submit multiple bids, and would be inclined to do so if such a manipulation were profitable. Meanwhile, the publisher could manipulate auction results by secretly acting as a bidder. It would be particularly troubling if the mechanism allowed the publisher, by submitting low shill bids that have little chance of winning, to raise the price paid by the winning bidder. We now define the class of mechanisms that are immune to such "false name" manipulations, and characterize these mechanisms in Lemma 3.

Definition 8. *A strategy-proof mechanism* (z, p) *is* **bidder false-name proof** *if no bidder can benefit by submitting multiple bids, meaning that for all* $n \geq 2$*,* $x \in \mathbb{R}_+^n$*, all* $m \geq 1$*, and all* $y \in \mathbb{R}_+^m$*:*

$$x_i z_i(x) - p_i(x) \geq x_i \left(z_i(x, y) + \sum_{j=1}^{m} z_{n+j}(x, y) \right) - \left(p_i(x, y) + \sum_{j=1}^{m} p_{n+j}(x, y) \right).$$

A mechanism (z, p) *is* **publisher false-name proof** *if submitting bids below that of the lowest performance bidder cannot raise the price paid by a winning bidder; that is, if for all* $n \geq 2$*,* $x \in \mathbb{R}_+^n$*,* $m \geq 1$ *and* $y \in \mathbb{R}_+^m$ *satisfying* $\max y \leq \min x$*:*

$$z_i(x) > 0 \implies p_i(x, y) \leq p_i(x).$$

The mechanism (z, p) *is* **false-name proof** *if it is both bidder false-name proof and publisher false-name proof.*

Lemma 3. *A deterministic threshold auction with price function* $\tilde{p}$ *is false-name proof if and only if* $\tilde{p}(x_{-i}) = \tilde{p}(\max\{x_{-i}\})$ *for all* x_{-i} *and* $\tilde{p} : \mathbb{R}_+ \to \mathbb{R}_+$ *is nondecreasing.*

We now introduce the modified second-bid (MSB) auction, which allocates the impression to the top performance bidder whenever his bid is at least α times the bid of his top competitor. The main result of this section characterizes MSB auctions as the unique deterministic auctions possessing the two desirable properties discussed above.

Definition 9. *A threshold auction is a* **modified second-bid** *(MSB) auction if there exists* $\alpha \geq 1$ *such that:*

$$\tilde{p}(w, x_{-i}) = \alpha \max\{x_{-i}\}. \tag{6}$$

For any $\alpha \geq 1$*, we denote by* $MSB(\alpha)$ *the threshold auction with threshold price given by* (6)*.*

Theorem 2. *A threshold auction is deterministic, adverse-selection free, and false-name proof if and only if it is a modified second-bid auction.*

Proof. It is clear that any MSB auction is deterministic. Furthermore, it is adverse-selection free by Lemma 2 and false-name proof by Lemma 3. Conversely, Lemmas 2 and 3 imply that if $\tilde{p}$ represents a deterministic, adverse-selection-free, false-name-proof mechanism, then the price faced by bidder i must be a function of only the top competing bid, and this function must be homogeneous of degree one. The fact that $\tilde{p}(x_{-i}) \geq \max\{x_{-i}\}$ implies $\alpha \geq 1$. □

Theorem 2 characterizes MSB auctions as the unique threshold auctions which are deterministic, adverse-selection free, and false-name proof. We note here that other threshold auctions may have any two of these three properties. For example, SPR auctions are deterministic and false-name proof, but not adverse-selection free. RSA auctions are adverse-selection free and false-name proof, but not deterministic. Finally, consider the threshold auction in which $h(x_{-i}) = 2\min\{x_{-i}\}$ and the threshold price is $\tilde{p}(x_{-i}) = \max\{x_{-i}, h(x_{-i})\}$. This mechanism is deterministic and, by Lemma 2, adverse-selection free, but it is not false-name proof, for whenever $h(x_{-i}) > \max\{x_{-i}\}$, the winning bidder could reduce its threshold price by submitting additional bids that are just barely above zero.

5 Benchmark Against RSA

The axiomatic characterization of the MSB auction highlights its appealing characteristics, but how do its efficiency and revenues compare to alternatives? This section and the next one tackle that question by comparing the MSB to two different alternatives. In the next section, we compare MSB to SPR to explore the opportunity cost of using MSB when adverse selection can be addressed by other means. In this section, we compare MSB to RSA. RSA has the advantage that it completely avoids adverse selection even when M and C are not independent. In order for MSB to be an attractive alternative, it needs to have the potential to deliver large gains relative to RSA.

Theorem 3 delivers just such a conclusion, using the assumption that the match values $M_1, M_2, ...$ are IID with some common distribution F and density f, that $E[M_i]$ is finite, and that the distribution of match values is "fat-tailed."[6] By fat-tailed, we mean that $xf(x)/(1-F(x))$ is non-increasing. This condition is satisfied, for example, by power law distributions and mixtures thereof.

We study this class of distributions both because we believe that advertiser valuations exhibit heavy tails in practice, and because MSB auctions can be proven to perform *much* better than RSA in this case.[7] In particular, Theorem 3 asserts that for fat-tailed distributions, MSB auctions offer a Pareto improvement over RSA: brand advertisers are unaffected and both the publisher's revenue and the performance bidder surplus are increased by a (potentially large) multiplicative

[6] Note that the axiomatic results of Section 4 do not rely on any distributional assumptions.

[7] Recall that Theorem 1 provides much weaker conditions under which MSB auctions provably improve upon the revenue and bidder surplus provided by RSA. The stronger condition used in Theorem 3 allows us to lower-bound the magnitude of this improvement.

factor.

Theorem 3. *Suppose that C and M are independent, and that the M_i are IID draws from distribution F on $[1, \infty)$ for which $xf(x)/(1 - F(x))$ is non-increasing. Fix a distribution over N, and suppose that λ, α satisfy $E[P_N(M_{(1)}/M_{(2)} \leq \alpha)] = \lambda$. Then*

$$\min\left(\frac{V(MSB(\alpha))}{V(RSA(\lambda))}, \frac{R(MSB(\alpha))}{R(RSA(\lambda))}, \frac{S(MSB(\alpha))}{S(RSA(\lambda))}\right) \geq \alpha.$$

We defer a proof of Theorem 3 to the Appendix.

Remark 1. *When match values are drawn from a power law distribution, with cdf given by $F(x) = 1 - x^{-a}$, the quantity $x\, f(x)/(1 - F(x)) = a$ is constant and*

$$\frac{V(MSB(\alpha))}{V(RSA(\lambda))} = \frac{R(MSB(\alpha))}{R(RSA(\lambda))} = \frac{S(MSB(\alpha))}{S(RSA(\lambda))} = \alpha = (1 - \lambda)^{-1/a}.$$

Note that the factor α increases without bound as the fraction λ of impressions allocated to the brand advertiser increases, allowing the gains from adopting an MSB auction to be quite large. To get a sense for the magnitudes implied by Theorem 3, suppose that the match values follow a power law distribution and satisfy the 80-20 rule. Then, for $\lambda = 0.8$, we must have $\alpha = 4$, which implies a 300% increase in both the revenue extracted from performance advertisers and the performance advertiser surplus. Similarly, for $\lambda = 0.5$, we must have $\alpha = 1.81$, which implies an 81% increase in revenues and surplus over RSA.

6 Benchmark Against an Optimal Auction

Theorem 3 demonstrates that if the distribution of bidder values is fat-tailed, the gains from adopting MSB relative to RSA can be quite large. But if the problems of adverse selection and false-name bidding could be addressed by other means, could another mechanism could do much better than MSB? How much is lost in the quest to avoid adverse selection and false-name bidding?

The MSB mechanism uses the second highest bid to distinguish cases when the common value is high from those when the top performance advertiser is a particularly good match. A mechanism that considers the entire profile of bids (rather than only the second highest) or that has direct access to exogenous information about C might conduct this inference more effectively than MSB, and thereby allocate impressions more efficiently.

In this section, we show that while such improvements over MSB may be possible, if match values follow a power law distribution, they are limited in magnitude. Indeed, Theorem 4 implies that for the power law case, even if the publisher could directly observe the brand advertiser's value (so that $C = 1$), MSB generates *at least* 88.5% of the value, revenue, and bidder surplus as any weakly efficient auction mechanism. This is a worst-case bound, which applies regardless of the realized number of bidders n, the fraction of impressions λ, and the parameter a of the power law distribution. By contrast, we know from Theorem 3 that RSA cannot be guaranteed to provide *any* constant fraction of the value generated by MSB.

Theorem 4. *Fix n and λ. Suppose that (i) $F(m) = 1 - m^{-a}$ and $a > 1$; that is, the M_i are IID draws from a power law distribution with parameter a with finite mean, (ii) the brand advertiser's value B is known, so that C is identically one, and (iii) α, t satisfy*

$$P_n(M_{(1)}/M_{(2)} \leq \alpha) = P_n(M_{(1)} \leq t) = \lambda.$$

Then
$$\frac{V_n(MSB(\alpha))}{V_n(SPR(t))} = \frac{R_n(MSB(\alpha))}{R_n(SPR(t))} = \frac{S_n(MSB(\alpha))}{S_n(SPR(t))} \geq \Gamma(2 - 1/a) \geq 0.8865,$$
where the function Γ is defined by $\Gamma(s+1) = \int_0^\infty x^s e^{-x} dx$.

The theorem gives more than just a worst-case bound. Among the possibilities that it includes is the power law distribution corresponding to the 80-20 rule, with $a = 1.16$. For that parameter value, Theorem 4 implies that MSB delivers at least 93.7% of the value, revenue and bidder profits generated by SPR, for any values of λ, n.

7 Conclusion/Discussion

Contracts for internet display advertising often require that certain impressions be set aside in advance, thereby preventing the highest quality consumers from being selected away in an auction. In this paper, we argue that while this practice protects brand advertisers from adverse selection, it comes at a great cost. This leads us to seek auction mechanisms that eliminate or mitigate adverse selection while being much less damaging to matching efficiency.

We introduce the Modified Second Bid auction as the unique deterministic mechanism that is adverse-selection-free and resistant to manipulation via the creation of false identities. As we show, the MSB auction can lead to a large, multiplicative increase in both revenues for the seller and profits for the performance bidders compared to merely selling to performance advertisers the 'remnant' impressions left unclaimed by brand advertisers. Furthermore, when value distributions are fat-tailed, the performance of MSB auctions approaches that of optimal adverse-selection-free mechanisms subject to relaxed constraints (they need not be false-name-proof and may rely on perfect, exogenous information about the consumer quality C).

Our theoretical findings are conditional on assumptions, and when those are varied, the conclusions will vary as well. Importantly, if the consumer and match quality are correlated, then the set of impressions delivered to the brand advertiser by MSB will not, in general, be independent of the consumer quality variable. However, so long as rare transitory factors are a major determinant of extreme match values, the dependence is likely to be small and adverse selection may be largely neutralized by an MSB auction.

Our quantitative analysis of the gains from MSB auctions incorporates the assumption that the match quality variables are independent and identically distributed. In practice, it seems reasonable to expect that advertisers selling similar products will sometimes target the same users, and thus their match values will be positively correlated. In such cases, the signature the MSB mechanism utilizes - a high bid by a single advertiser - may not be found in the bid data, some match value may be lost, and the average performance may not be as good as in the case of independent match values. Still, the improvements in our examples are so large that, even if some of the high value matches are missed in this way, substantial performance may still be possible.

If the MSB mechanism is as good as the theory and computations suggest, why wasn't it invented earlier? Perhaps the answer is that online display advertising is relatively new and has grown quickly in importance. Perhaps it is that most online display advertising initially came from brand advertisers, who originally bought online ads in the same way as they buy ads in broadcast media and whose resistance has slowed the adoption of auctions. Perhaps the problem of adverse selection is even more severe than we have imagined, and the solution that we propose offers too little protection to brand advertisers. Whatever the reasons, adverse selection in auctions and other markets has long been a central economic concern, and the design of nearly-efficient mechanisms that mitigate adverse selection is an interesting and important frontier in market design.

References

Ittai Abraham, Susan Athey, Moshe Babaioff, and Michael Grubb. Peaches, lemons, and cookies: designing auction markets with dispersed information. In *ACM Conference on Electronic Commerce*, pages 7–8, 2013.

Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97(1):242–259, 2007.

Randall Lewis and David Reiley. Does retail advertising work: Measuring the effects of advertising on sales via a controlled experiment on yahoo. Technical report, Working paper, 2010.

R.Preston McAfee, Kishore Papineni, and Sergei Vassilvitskii. Maximally representative allocations for guaranteed delivery advertising campaigns. *Review of Economic Design*, 17(2):83–94, 2013. ISSN 1434-4742.

P.R. Milgrom, S.J. Cunningham, and M.R. Beck. Impression allocation system and methods using an auction that considers losing bids, December 20 2012. US Patent App. 13/163,417.

Hal R. Varian. Position auctions. *International Journal of Industrial Organization*, 25(6):1163 – 1178, 2007. ISSN 0167-7187.

8 Appendix for Sections 2, 3, and 4

Proof of Lemma 1. Threshold auctions are anonymous by definition and it is straightforward to verify that any threshold auction is a strategy-proof, anonymous, and weakly efficient mechanism.

Conversely, suppose that (z, p) is a strategy-proof, anonymous, and weakly efficient mechanism. It is well known that for any strategy-proof mechanism and any x_{-i}, $z_i(x_i, x_{-i})$ is non-decreasing in x_i. By anonymity, we can suppress i to define a function h as follows:

$$h(w, x_{-i}) = \sup\{x_i : z_i(x) \leq w\}. \tag{7}$$

Applying the envelope theorem in the usual way, for any given monotone allocation rule, there is a unique payment rule that makes the mechanism strategy-proof. Thus, to prove that the threshold auction with pricing function $\tilde{p}(w, x_{-i}) = \max\{x_{-i}, h(w, x_{-i})\}$ implements (z, p), it is enough to show that $z_i(x) = P(x_i > \tilde{p}(W, x_{-i}))$.

Because z is weakly efficient, $z_i(x) = 0$ for $x_i \leq \max\{x_{-i}\}$. So, (7) implies that for all w and x_{-i}, $h(w, x_{-i}) \geq \max\{x_{-i}\}$. Therefore

$$\tilde{p}(w, x_{-i}) = \max\{x_{-i}, h(w, x_{-i})\} = h(w, x_{-i}). \tag{8}$$

Because z_i is left continuous in x_i, (7) implies that

$$x_i > h(w, x_{-i}) \Leftrightarrow z_i(x) > w, \tag{9}$$

Combining (8) and (9), we see that for all x,

$$P(x_i > \tilde{p}(W, x_{-i})) = P(x_i > h(W, x_{-i})) = P(z_i(x) > W) = z_i(x).$$

□

Throughout the remainder of the appendix, given a probability density function f that is positive on the interval $[a, b) \subseteq [0, \infty)$, we define

$$g(x) = x - \frac{1 - F(x)}{f(x)} \tag{10}$$

to be the virtual value function associated with f. Our proof of Theorem 1 makes use of the following Proposition.

Proposition 1. *Suppose that the M_i are IID draws from a distribution F with a density f that is positive on an interval $[a, b) \subseteq [0, \infty)$. Fix a threshold auction with price function $\tilde{p}(w, x_{-i}) =$*

$\max\{x_{-i}, h(w, x_{-i})\}$. *Then for any i,*

$$V_n(\tilde{p}) = nE_n\left[M_i\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] \tag{11}$$

$$R_n(\tilde{p}) = nE_n\left[g(M_i)\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] \tag{12}$$

$$S_n(\tilde{p}) = nE_n\left[(M_i - g(M_i))\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] \tag{13}$$

Proof of Proposition 1. To establish (11), note that

$$V_n(\tilde{p}) = E_n\left[\sum_{i=1}^{N} M_i z_i(M)\right] = E_n\left[\sum_{i=1}^{N} M_i\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] = nE_n\left[M_i\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right], \tag{14}$$

where the final equality follows from the symmetry of the mechanism and the assumption that the M_i are IID. Similarly, the expected revenue, given $N = n$, is

$$R_n(\tilde{p}) = E_n\left[\sum_{i=1}^{N} \tilde{p}(W, M_{-i})\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] = nE_n\left[\tilde{p}(W, M_{-i})\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right]. \tag{15}$$

For any $s \in [a, b)$,

$$\begin{aligned} E\left[g(M_i)\mathbf{1}_{M_i>s}\right] &= \int_s^b g(x)f(x)dx \\ &= \int_s^b \left[xf(x) - (1 - F(x))\right]dx \\ &= x(1 - F(x))\Big|_b^s \\ &= sP(M_i > s). \end{aligned} \tag{16}$$

Note that in the case that $b = \infty$, the final equality uses the fact that $\lim_{x\to\infty} x(1 - F(x)) = 0$, which follows because the distribution F has a finite mean. Since M_i is independent from $\tilde{p}(W, M_{-i})$, (16) shows that for any $s \in [a, b)$,

$$E_n\left[g(M_i)\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}|\tilde{p}(W, M_{-i}) = s\right] = sP(M_i > s).$$

Applying the law of iterated expectation, we establish that

$$E_n\left[\tilde{p}(W, M_{-i})\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right] = E_n\left[g(M_i)\mathbf{1}_{M_i>\tilde{p}(W,M_{-i})}\right],$$

completing the proof of (12). Because (13) follows immediately from (11) and (12), this completes the proof of Proposition 1. $\square$

The proof of Theorem 1 makes use of two facts about functions of random variables, which we state without proof.

Fact 1. *Let X be a real-valued random variable. If $u(\cdot)$ and $v(\cdot)$ are nondecreasing functions, then $E[u(X)v(X)] \geq E[u(X)]E[v(X)]$.*

Fact 2. *Let X be a real-valued random variable. If u is non-decreasing and v is single-crossing with $E[v(X)] = 0$, then $E[u(X)v(X)] \geq 0$.*
("Single-crossing" means that there is some real number z such that for all x, $(x-z)v(x) \geq 0$.)

Proof of Theorem 1. To prove that $V_n(RSA(\lambda)) \leq V_n(\tilde{p})$, we apply Fact 1 to (11) from Proposition 1. For fixed W and M_{-i}, and define $u(M_i) = M_i$, $v(M_i) = \mathbf{1}_{M_i > \tilde{p}(W, M_{-i})}$, and $A = \{M_i > \max\{M_{-i}\}\}$. Then it follows from (11) that

$$\begin{aligned} V_n(\tilde{p}) &= nE_n[u(M_i)v(M_i)\mathbf{1}_A] \\ &= nP(A)E_n[u(M_i)v(M_i)|A] \\ &\geq E_n[u(M_i)|A]E[v(M_i)|A] \\ &= E_n\left[M_{(1)}\right] P_n\left(M_{(1)} > \tilde{p}\left(W, M_{-(1)}\right)\right) \\ &= V_n(RSA(\lambda)), \end{aligned}$$

where the inequality above follows from an application of Fact 1 and the fact that $nP(A) = 1$, and the final equality follows from the fact that $P_n\left(M_{(1)} > \tilde{p}\left(W, M_{-(1)}\right)\right) = 1 - \lambda$.

Similarly, if $g(x) = x - \frac{1-F(x)}{f(x)}$ is increasing, then we apply Fact 1 to (12), with $u(M_i) = g(M_i)$ and v and A as before. If $x - g(x)$ is increasing, identical logic with $u(M_i) = M_i - g(M_i)$ proves that $R_n(\tilde{p}) \geq R_n(RSA(\lambda))$.

Having completed the proof of the pessimality of random set-asides, we move on to the optimality of the second-price auction. We define $u(M_i) = M_i \mathbf{1}_{M_i > \max\{M_{-i}\}}$, $A = \{M_i > \tilde{p}(W, M_{-i})\}$, and $B = \{M_i > \max(M_{-i}, t)\}$. Then by (11), we have

$$V_n(\tilde{p}) = nE_n[u(M_i)\mathbf{1}_A], \qquad V_n(SPR(t)) = nE_n[u(M_i)\mathbf{1}_B].$$

Define $v(x) = P_n(B|M_i = x) - P_n(A|M_i = x)$. Note that $E_n[v(M_i)] = 0$ (because both auctions sell with probability $1 - \lambda$ when $N = n$). Furthermore, for $x \leq t$, $v(x) \leq 0$, and for $x > t$, $P_n(B|M_i = x) = P_n(\max\{M_{-i}\} \leq x) \geq P_n(A|M_i = x)$, so $v(x) \geq 0$ and thus v is single-crossing. By the law of iterated expectation,

$$V_n(SPR(t)) - V_n(\tilde{p}) = n(E_n[u(M_i)\mathbf{1}_B] - E_n[u(M_i)\mathbf{1}_A]) = nE_n[u(M_i)v(M_i)].$$

Applying Fact 2, we conclude that the above quantity is non-negative, i.e. $V_n(\tilde{p}) \leq V_n(SPR(t))$.

Having proved that the second-price auction is optimal for value creation, we move on to the claim about revenue. Define $u(M_i) = g(M_i)\mathbf{1}_{M_i > \max\{M_{-i}\}}$, with A, B, and v as before. If g is non-decreasing, then u is non-decreasing, and thus we may conclude from (12) and Fact 2 that

$$R_n(SPR(t)) - R_n(\tilde{p}) = nE_n[u(M_i)v(M_i)] \geq 0.$$

The claim about bidder surplus is proved identically, with $u(M_i) = (M_i - g(M_i))\mathbf{1}_{M_i > \max\{M_{-i}\}}$.

□

Proof of Lemma 2. If $\tilde{p}$ is homogeneous of degree one, then, for all $i \in \{1, \ldots, n\}$, $z_i(CM) = \mathbf{1}_{CM_i > \tilde{p}(CM_{-i})} = \mathbf{1}_{M_i > \tilde{p}(M_{-i})}$ is a function only of M. It follows that for every distribution such that M and C are independent, z_0 is independent of C, so the mechanism is adverse-selection free.

Conversely, suppose that $\tilde{p}$ is not homogeneous of degree one. Then there exists $c \in \mathbb{R}_+, n \geq 2$, and $x_{-i} \in \mathbb{R}_+^{n-1}$ such that (without loss of generality) $\tilde{p}(x_{-i}) < \tilde{p}(cx_{-i})/c$. Fix $x_i \in (\tilde{p}(x_{-i}), \tilde{p}(cx_{-i})/c)$. Suppose that $C \in \{1, c\}$ with $P(C = 1) \in (0, 1)$, that $P(M_{-i} = x_{-i}) = 1$, and that $P(M_i = x_i) = 1$. We show that $z_0(CM) = \mathbf{1}_{\{C=c\}}$, proving that the auction associated with $\tilde{p}$ is not adverse-selection free.

Because $\max\{x_{-i}\} \leq \tilde{p}(x_{-i}) < x_i$, we know that every impression is awarded to either bidder i or to the brand advertiser. When $C = 1$, $z_i(CM) = z_i(x) = \mathbf{1}_{\{x_i > \tilde{p}(x_{-i})\}} = 1$, so $z_0(CM) = 0$. When $C = c$, $z_i(CM) = z_i(cx) = \mathbf{1}_{\{cx_i > \tilde{p}(cx_{-i})\}} = 0$, so $z_0(CM) = 1$.

□

Proof of Lemma 3. We first show that the condition that $\tilde{p}(x_{-i}) = \tilde{p}(\max\{x_{-i}\})$ for all x_{-i} is necessary. Suppose that there is some x_{-i} such that this condition does not hold.

- If $\tilde{p}(x_{-i}) < \tilde{p}(\max\{x_{-i}\})$, then if there are two bidders, one with value exceeding $\tilde{p}(x_{-i})$ and the other with value $\max\{x_{-i}\}$, the first bidder can benefit by submitting the remaining bids in the profile x_{-i}.

- If $\tilde{p}(x_{-i}) > \tilde{p}(\max\{x_{-i}\})$, then if there are two bidders, one with value exceeding $\tilde{p}(x_{-i})$ and the other with value $\max\{x_{-i}\}$, the publisher can raise the price paid by the first bidder (without affecting the allocation) by submitting the remaining bids in the profile x_{-i} (which are all below both performance bids).

We now show that it is necessary that $\tilde{p} : \mathbb{R}_+ \to \mathbb{R}_+$ be non-decreasing. For, if there were $y < y'$ with $\tilde{p}(y) > \tilde{p}(y')$, then a bidder with value above $\tilde{p}(y')$ who faced a maximum competing bid of y would benefit from submitting a bid of y' in addition to their true value.

Conversely, suppose the conditions of the Lemma hold. It is clear that the publisher cannot change the allocation or payments by submitting bids below the lowest performance bid, since the threshold price faced by each bidder is a function only of the highest competing bid. Furthermore, because the auction is weakly efficient, any bidder who submits multiple bids knows that only the highest of these can win. Submitting additional bids can only raise the threshold price that their highest bid faces, since $\tilde{p}$ is a non-decreasing function of the maximum competing bid. It follows that no bidder can benefit from submitting multiple bids.

$\square$

9 Appendix for Section 5

Lemma 4. *$\frac{1-F(x)}{xf(x)}$ is non-decreasing in x if and only if $\frac{1-F(\alpha x)}{1-F(x)}$ is non-decreasing in x for all $\alpha \geq 1$.*

Proof of Lemma 4. The statement that $\frac{1-F(\alpha x)}{1-F(x)}$ is non-decreasing in x is equivalent to

$$\frac{d}{dx}\left(\log(1-F(\alpha x)) - \log(1-F(x))\right) \geq 0,$$

which is equivalent to

$$\frac{1-F(\alpha x)}{\alpha x f(\alpha x)} \geq \frac{1-F(x)}{xf(x)} \text{ for all } x. \tag{17}$$

Note that (17) holds for all $\alpha \geq 1$ if and only if $\frac{1-F(x)}{xf(x)}$ is non-decreasing in x. $\square$

Proposition 2. *Suppose that the M_i are IID from a continuous distribution with density f and cdf F on $[a, b)$. For any $n \geq 2$ and $1 \leq \alpha \leq b/a$,*

1.

$$P_n(M_{(1)}/M_{(2)} > \alpha) = E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right]$$

2.

$$R_n(MSB(\alpha)) = \alpha\ E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}M_{(2)}\right]$$

3.

$$S_n(MSB(\alpha)) = \alpha\ E_n\left[\frac{1-F(\alpha M_{(1)})}{f(M_{(1)})}\right]$$

Proof of Proposition 2. Note that

$$\begin{aligned}
P_n(M_{(1)}/M_{(2)} > \alpha) &= E_n\left[\sum_{i=1}^{N} \mathbf{1}_{M_i/\max\{M_{-i}\}>\alpha}\right] \\
&= nP_n(M_i/\max\{M_{-i}\} > \alpha) \\
&= n\int_a^b (1-F(\alpha x))(n-1)f(x)F(x)^{n-2}dx \\
&= \int_a^b \frac{1-F(\alpha x)}{1-F(x)} n(n-1)f(x)F(x)^{n-2}(1-F(x))dx \\
&= E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right].
\end{aligned}$$

Whenever $M_i > \alpha\max\{M_{-i}\}$, the publisher receives revenue of $\alpha\max\{M_{-i}\}$. It follows that

$$\begin{aligned}
R_n(MSB(\alpha)) &= nE_n[\alpha\max\{M_{-i}\}\mathbf{1}_{M_i/\max\{M_{-i}\}>\alpha}] \\
&= \alpha n\int_a^b x(1-F(\alpha x))(n-1)f(x)F(x)^{n-2}dx \\
&= \alpha\int_a^b x\frac{1-F(\alpha x)}{1-F(x)} n(n-1)f(x)F(x)^{n-2}(1-F(x))dx \\
&= \alpha\ E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}M_{(2)}\right].
\end{aligned}$$

Finally, we have from (13) in Proposition 1 that

$$\begin{aligned}
S_n(MSB(\alpha)) &= nE_n\left[\frac{1-F(x_i)}{f(x_i)}\mathbf{1}_{x_i>\alpha\max\{x_{-i}\}}\right] \\
&= n\int_{\alpha a}^b \frac{1-F(x)}{f(x)} f(x)F(x/\alpha)^{n-1}dx \\
&= n\int_{\alpha a}^b \frac{1-F(x)}{f(x/\alpha)} f(x/\alpha)F(x/\alpha)^{n-1}dx \\
&= \alpha\int_a^{b/\alpha} \frac{1-F(\alpha u)}{f(u)} nf(u)F(u)^{n-1}du \\
&= \alpha E_n\left[\frac{1-F(\alpha M_{(1)})}{f(M_{(1)})}\right].
\end{aligned}$$

□

Proof of Theorem 3. For fixed α, we define

$$\lambda_n = 1 - P_n(M_{(1)}/M_{(2)} > \alpha). \tag{18}$$

Then we have from Proposition 2.2 that

$$
\begin{aligned}
R_n(MSB(\alpha)) &= \alpha\ E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}M_{(2)}\right] \\
&\geq \alpha\ E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right]E_n[M_{(2)}] \\
&= \alpha(1-\lambda_n)E_n[M_{(2)}]. \qquad (19)
\end{aligned}
$$

Note that the second line applies Fact 1, which is justified because Lemma 4 implies that $\frac{1-F(\alpha x)}{1-F(x)}$ is non-decreasing. The third line simply applies Proposition 2.1 and the definition of λ_n in (18). The final line follows from Proposition 1.

Similarly, Proposition 2.3 implies that

$$
\begin{aligned}
S_n(MSB(\alpha)) &= \alpha E_n\left[\frac{1-F(\alpha M_{(1)})}{1-F(M_{(1)})}\cdot\frac{1-F(M_{(1)})}{f(M_{(1)})}\right] \\
&\geq \alpha E_n\left[\frac{1-F(\alpha M_{(1)})}{1-F(M_{(1)})}\right]E_n\left[\frac{1-F(M_{(1)})}{f(M_{(1)})}\right] \\
&\geq \alpha E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right]E_n\left[\frac{1-F(M_{(1)})}{f(M_{(1)})}\right] \\
&= \alpha(1-\lambda_n)E_n[M_{(1)}-M_{(2)}]. \qquad (20)
\end{aligned}
$$

The second line is an application of Fact 1, which is justified by Lemma 4 and the fact that if $\frac{1-F(x)}{xf(x)}$ is non-decreasing, then so is $\frac{1-F(x)}{f(x)}$. The third line follows from the fact that $M_{(2)} \leq M_{(1)}$, and the final line from the definition of λ_n and from the fact that

$$
\begin{aligned}
E_n[M_{(1)}-M_{(2)}] &= \int_a^b P_n(M_{(1)} \geq x))dx - \int_a^b P_n(M_{(2)} \geq x)dx \\
&= \int_a^b P_n(M_{(2)} \leq x < M_{(1)})dx \\
&= \int_a^b \frac{1-F(x)}{f(x)}nf(x)F(x)^{n-1}dx \\
&= E_n\left[\frac{1-F(M_{(1)})}{f(M_{(1)})}\right]. \qquad (21)
\end{aligned}
$$

Combining (19) and (20), we see that

$$
V_n(MSB(\alpha)) = R_n(MSB(\alpha)) + S_n(MSB(\alpha)) \geq \alpha\ (1-\lambda)E_n[M_{(1)}]. \qquad (22)
$$

Because $V_n(RSA(\lambda)) = (1-\lambda)E_n[M_{(1)}]$ and $R_n(RSA(\lambda)) = (1-\lambda)E_n[M_{(2)}]$, equations (19), (20) and (22) jointly imply the result of Theorem 3 in the case where N is deterministic. To extend this result to arbitrary distributions satisfying $P(N \geq 2) = 1$, we claim that λ_n is non-increasing in n. To show this, we use Proposition 2.1 to conclude that

$$P_{n+1}(M_{(1)}/M_{(2)} > \alpha) = E_{n+1}\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right] \geq E_n\left[\frac{1-F(\alpha M_{(2)})}{1-F(M_{(2)})}\right] = P_n(M_{(1)}/M_{(2)} > \alpha), \tag{23}$$

where the inequality above follows because $\frac{1-F(\alpha x)}{1-F(x)}$ is non-decreasing (Lemma 4) and the distribution of $M_{(2)}$ conditioned on $N = n+1$ stochastically dominates that of $M_{(2)}$ conditioned on $N = n$. The fact that λ_n is non-increasing follows immediately from (23) and the definition of λ_n (18).

We now take expectations of (19) to obtain

$$\begin{aligned}
R(MSB(\alpha)) &= E[R_N(MSB(\alpha))] && \text{By definition}\\
&\geq \alpha E[(1-\lambda_N)E_N[M_{(2)}]] && \text{From (19)}\\
&\geq \alpha E[1-\lambda_N]E[E_N[M_{(2)}]] && \text{Fact 1}\\
&= \alpha(1-\lambda)E[M_{(2)}] && \lambda = E[\lambda_N]\\
&= \alpha\, R(RSA(\lambda)).
\end{aligned}$$

Analogously,

$$\begin{aligned}
V(MSB(\alpha)) &= E[V_N(MSB(\alpha))] && \text{By definition}\\
&\geq \alpha E[(1-\lambda_N)E_N[M_{(1)}]] && \text{From (22)}\\
&\geq \alpha E[1-\lambda_N]E[E_N[M_{(1)}]] && \text{Fact 1}\\
&= \alpha(1-\lambda)E[M_{(1)}] && \lambda = E[\lambda_N]\\
&= \alpha\, V(RSA(\lambda)).
\end{aligned}$$

For surplus, we have

$$\begin{aligned}
S(MSB(\alpha)) &= E[S_N(MSB(\alpha))] && \text{By definition}\\
&\geq \alpha E\left[(1-\lambda_N)E_N\left[\frac{1-F(M_{(1)})}{f(M_{(1)})}\right]\right] && \text{From (20)}\\
&\geq \alpha E[1-\lambda_N]E\left[E_N\left[\frac{1-F(M_{(1)})}{f(M_{(1)})}\right]\right] && \text{Fact 1}\\
&= \alpha(1-\lambda)E\left[M_{(1)} - M_{(2)}\right] && \text{From (21)}\\
&= \alpha\, S(RSA(\lambda)).
\end{aligned}$$

□

10 Appendix for Section 6

Lemma 5. *If the M_i are IID draws from a power law distribution with parameter a, and C and M are independent, then for* any *threshold pricing function $\tilde{p}$, the vector $(V_n(\tilde{p}), R_n(\tilde{p}), S_n(\tilde{p}))$ is proportional to $\left(1, \frac{a-1}{a}, \frac{1}{a}\right)$.*

Proof. Note that if $F(x) = 1 - x^{-a}$, $g(x) = x - \frac{1-F(x)}{f(x)} = \frac{a-1}{a}x$. The Lemma follows immediately from this fact and Proposition 1. □

In particular, Lemma 5 proves that for any α and t,

$$\frac{V_n(MSB(\alpha))}{V_n(SPR(t))} = \frac{R_n(MSB(\alpha))}{R_n(SPR(t))} = \frac{S_n(MSB(\alpha))}{S_n(SPR(t))},$$

so in the remainder of this section, we focus on the ratio $\frac{V_n(MSB(\alpha))}{V_n(SPR(t))}$.

Our proof of Theorem 4 also uses the following fact about the power law distribution, which we state without proof.

Fact 3 (Power Law Distribution)**.** *Suppose that $\{M_i\}_{i=1}^{n+1}$ are IID draws from a power law distribution with parameter a, i.e. $P(M_i \leq x) = 1 - x^{-a} = F(x)$. Let $f(x) = ax^{-a-1}$ be the density of each M_i, and let $M_{(j)}$ be the j^{th} order statistic of the M_i. Then $M_{(1)}/M_{(n+1)}, M_{(2)}/M_{(n+1)}, \ldots, M_{(n)}/M_{(n+1)}$ are independent from $M_{(n+1)}$ and are distributed as the order statistics of $M_1, \ldots, M_n$.*

For fixed n, λ, and f, define $\alpha(\lambda, n)$ and $t(\lambda, n)$ by

$$P_n(M_{(1)} \leq t(\lambda, n)) = P_n(M_{(1)}/M_{(2)} \leq \alpha(\lambda, n)) = \lambda. \tag{24}$$

It will be useful to have expressions for $t(\lambda, n)$ and $\alpha(\lambda, n)$ in the case where match values are drawn from a power law with parameter a, so that $F(x) = 1 - x^{-a}$. Repeated application of Fact 3 implies that for any n, $M_{(1)}/M_{(2)}$ has a power law distribution and is independent of $M_{(2)}$. From this fact and the definition of F, It is immediate that

$$t(\lambda, n) = (1 - \lambda^{1/n})^{-1/a}, \qquad \alpha(\lambda, n) = (1 - \lambda)^{-1/a}. \tag{25}$$

The following Lemma establishes that the performance of MSB (relative to SPR) decreases as the number of bidders grows.

Lemma 6. *Suppose that the M_i are IID draws from a power law distribution with parameter a, and that C and M are independent. Let $t(\lambda, n)$ and $\alpha(\lambda, n)$ be as given in* (25)*, so that $MSB(\alpha(\lambda, n))$ and $SPR(t(\lambda, n))$ sell the impression with equal probability when $N = n$. For any $\lambda \in (0, 1)$, $\frac{V_n(MSB(\alpha(\lambda,n))}{V_n(SPR(t(\lambda,n)))}$ is decreasing in n.*

Proof of Lemma 6. Note that Fact 3 implies that when $N = n+1$, the values $R_i = M_{(i)}/M_{(n+1)}$ for $i = 1, \ldots, n$ look identical to the order statistics of n IID draws from a power law distribution with parameter a, and are independent from $M_{(n+1)}$. Thus, for any α,

$$\begin{aligned} V_{n+1}(MSB(\alpha)) &= E_{n+1}\left[M_{(n+1)} R_1 \mathbf{1}_{\frac{R_1}{R_2}>\alpha}\right] \\ &= E_{n+1}\left[M_{(n+1)}\right] E_{n+1}\left[R_1 \mathbf{1}_{\frac{R_1}{R_2}>\alpha}\right] \\ &= E_{n+1}\left[M_{(n+1)}\right] V_n(MSB(\alpha)). \end{aligned}$$

The second line follows from the independence of $M_{(1)}/M_{(2)}$ from $M_{(n+1)}$ and the fact that $R_1/R_2 = M_{(1)}/M_{(2)}$, while the final line follows from Fact 3. Furthermore, (25) implies that when the M_i follow a power law distribution, $\alpha(\lambda, n)$ does not depend on n. Thus, to prove the lemma, it suffices to show that for any $\lambda \in (0,1)$ and $n \geq 2$,

$$V_{n+1}(SPR(t(\lambda, n+1))) \geq E_{n+1}[M_{(n+1)}] V_n(SPR(t(\lambda, n))),$$

We do this by considering an auction that generates value of exactly $E_{n+1}[M_{(n+1)}]V_n(SPR(t(\lambda, n)))$ when there are $N = n+1$ bidders. By Theorem 1, this alternative auction provides a lower-bound on the performance of $SPR(t(\lambda, n))$. When $N = n+1$, this auction uses the ratio $R_1 = M_{(1)}/M_{(n+1)}$ to determine how to allocate the impression: it goes to the top performance advertiser whenever R_1 exceeds $t(\lambda, n)$. Note that Fact 3 implies that $P_{n+1}(R_1 \leq t(\lambda, n)) = P_n(M_{(1)} \leq t(\lambda, n))$, so this auction allocates the impression to the brand advertiser with the desired probability λ. By Theorem 1 (which asserts the optimality of SPR), Proposition 1 (which provides an expression for the value generated by this alternative auction), and Fact 3 (which allows us to evaluate this expression), it follows that

$$\begin{aligned} V_{n+1}(SPR(t(\lambda, n+1))) &\geq E_{n+1}[M_{(n+1)} R_1 \mathbf{1}_{R_1 > t(\lambda, n)}] \\ &= E_{n+1}[M_{(n+1)}] E_{n+1}[R_1 \mathbf{1}_{R_1 > t(\lambda, n)}] \\ &= E_{n+1}[M_{(n+1)}] V_n(SPR(t(\lambda, n))), \end{aligned}$$

completing the proof. □

Our proof of Theorem 4 also makes use of the following proposition, which holds for *any* continuous probability density f.

Lemma 7. *Fix $n \geq 2$, and let $t(\lambda) = t(\lambda, n)$ (as defined by (24)). If the M_i are IID draws from a* *continuous distribution with density f and C and M are independent, we have that*

$$\frac{d}{d\lambda}V_n(SPR(t(\lambda)) = -t(\lambda).$$

Proof of Lemma 7. Differentiating the identity $F(t(\lambda))^n = \lambda$, we obtain

$$nF(t(\lambda))^{n-1}f(t(\lambda))\frac{d}{d\lambda}t(\lambda) = 1. \tag{26}$$

Therefore,

$$\begin{aligned}\frac{d}{d\lambda}V_n(SPR(t(\lambda)) &= \frac{d}{d\lambda}\int_{t(\lambda)}^{\infty} xnF(x)^{n-1}f(x)dx \\ &= -t(\lambda)nF(t(\lambda))^{n-1}f(t(\lambda))\frac{d}{d\lambda}t(\lambda) \\ &= -t(\lambda),\end{aligned}$$

where the final line follows from application of (26).

□

Proof of Theorem 4. Note that by Lemma 5, it suffices to prove that for all $\lambda \in (0,1)$ and $n \geq 2$,

$$\frac{V_n(MSB(\alpha(\lambda,n)))}{V_n(SPR(t(\lambda,n)))} \geq \Gamma(2-1/a). \tag{27}$$

By Remark 1,

$$V_n(MSB(\alpha(\lambda,n))) = \alpha(\lambda,n)V_n(RSA(\lambda)) = \alpha(\lambda,n)(1-\lambda)E_n[M_{(1)}]. \tag{28}$$

Meanwhile, Lemma 7 implies that

$$\begin{aligned}V_n(SPR(t(\lambda,n))) &= \int_1^{\lambda}\frac{d}{dx}V_n(SPR(t(x,n)))dx \\ &= \int_{\lambda}^{1} t(x,n)dx \\ &\leq \int_{\lambda}^{1}\left(\frac{1-x}{n}\right)^{-1/a}dx \\ &= \frac{a}{a-1}(1-\lambda)^{1-1/a}n^{1/a} \\ &= \frac{a}{a-1}\alpha(\lambda,n)(1-\lambda)n^{1/a}.\end{aligned} \tag{29}$$

The inequality above uses the fact that for any λ, n, $t(\lambda, n) \leq \left(\frac{1-\lambda}{n}\right)^{-1/a}$, which follows because

$$F\left(\left(\frac{1-\lambda}{n}\right)^{-1/a}\right)^n = \left(1 - \frac{1-\lambda}{n}\right)^n \geq \lambda = F(t(\lambda, n))^n.$$

The equality (29) follows from the expression for $\alpha(\lambda, n)$ in (25).

From (28) and (29), we see that

$$\frac{V_n(MSB(\alpha(\lambda, n)))}{V_n(SPR(t(\lambda, n)))} \geq \frac{a-1}{a} n^{-1/a} E_n[M_{(1)}]. \tag{30}$$

Furthermore,

$$\begin{aligned} E_n[M_{(1)}] &= \int_1^\infty x n f(x) F(x)^{n-1} dx \\ &= \int_0^n F^{-1}\left(1 - \frac{u}{n}\right)\left(1 - \frac{u}{n}\right)^{n-1} du \\ &= \int_0^n \left(\frac{u}{n}\right)^{-1/a}\left(1 - \frac{u}{n}\right)^{n-1} du, \end{aligned} \tag{31}$$

where we have made the substitution $u = n(1 - F(x))$ (the final line follows from $F(x) = 1 - x^{-1/a}$).

By (30) and Lemma 6, we have that

$$\frac{V_n(MSB(\alpha(\lambda, n)))}{V_n(SPR(t(\lambda, n)))} \geq \lim_{n\to\infty} (1 - 1/a) n^{-1/a} E_n[M_{(1)}]. \tag{32}$$

By (31) and the dominated convergence theorem,

$$\lim_{n\to\infty} n^{-1/a} E_n[M_{(1)}] = \lim_{n\to\infty} \int_0^n u^{-1/a}\left(1 - \frac{u}{n}\right)^{n-1} du = \int_0^\infty u^{-1/a} e^{-u} du = \Gamma(1 - 1/a). \tag{33}$$

Combining (32) and (33), and using identity $\Gamma(s+1) = s\Gamma(s)$ (verifiable by integration by parts), we establish (27), completing the proof. □