Buy-it-now or Take-a-chance:
A Mechanism for Real-Time Price Discrimination *

L. Elisa Celis   Gregory Lewis   Markus M. Mobius
Hamid Nazerzadeh

April 17, 2012

Abstract

Increasingly sophisticated consumer tracking technology allows advertisers to reach narrowly targeted consumer demographics. Such targeting enhances advertising efficiency by improving the match quality between advertisers and users, but can also result in thin markets for particular demographic groups. Using historical bidding data from a large premium advertising exchange, we show that there is often a substantial gap between the highest and second highest willingness to pay, consistent with this intuition. This motivates our new BIN-TAC mechanism, which is simple and effective in extracting revenue when such a gap exists. Bidders can “buy-it-now”, or alternatively “take-a-chance” in an auction, where the top \( d > 1 \) bidders are equally likely to win. The randomized take-a-chance allocation incentivizes high valuation bidders to buy-it-now. We show that for a large class of distributions, this mechanism outperforms the second-price auction, and achieves revenue performance close to Myerson’s optimal mechanism. We apply structural methods to our data to estimate counterfactual revenues, and find that our BIN-TAC mechanism improves revenue by 4.5% relative to a second-price auction with optimal reserve.

*Department of Computer Science, University of Washington (ecelis@cs.washington.edu), Department of Economics, Harvard University (glewis@fas.harvard.edu), Microsoft Research New England (mobius@microsoft.com), and USC Marshall School of Business (nazerzad@marshall.usc.edu) respectively. All authors are grateful to Microsoft Research New England for their hospitality, and Greg would also like to thank NET Institute (www.NETinst.org) for financial support. Josh Feng provided excellent research assistance. We thank Susan Athey, Alessandro Bonatti, Michael Grubb, Mallesh Pai and Lee Zen for fruitful discussions that improved this work. Finally we thank Lee Zen and members of Microsoft’s display advertising team— in particular Ian Ferreira, Keith Hawley, and Brandon Zirkle — for their help.
1 Introduction

Advertising technology is changing fast. Consumers can now be reached while browsing the internet, playing games on their phone or watching videos on YouTube. The large companies that control these new media — household names like Google, Facebook and Yahoo! — generate a substantial part of their revenue by selling advertisements. They also know a lot about their users. This allows them to match advertisers to potential buyers with ever greater efficiency. While these technologies generate a lot of surplus for advertisers, they also tend to create thin markets where perhaps only a single advertiser has a high willingness to pay. These environments pose special challenges for the predominant auction mechanisms that are used to sell online ads because they reduce competition among bidders, making it difficult for the platform to extract the surplus generated by targeting.

For example, a sportswear firm advertising on the New York Times website may be willing to pay much more for an advertisement placed next to a sports article than one next to a movie review. It might pay an additional premium for a local consumer who lives in New York City and an even higher premium if the consumer is known to browse websites selling sportswear. Each layer of targeting increases the sportswear firm’s valuation for the consumer but also dramatically narrows the set of participating bidders to fellow sportswear firms in New York City. Without competition, revenue performance may be poor (Bergemann and Bonatti 2010, Levin and Milgrom 2010).

Consider a simple model: When advertisers “match” with users, they have high valuation; otherwise they have low valuation. Assume that match probabilities are independent across bidders, and sufficiently low that the probability that any bidder matches is relatively small. Then a second-price auction will typically get low revenue, since the probability of two “matches” occurring in the same auction is small. On the other hand, setting a high fixed price is not effective since the probability of zero “matches” occurring is relatively large and many impressions would go unallocated. Hence, allowing targeting creates asymmetries in valuations that can increase efficiency, but decrease revenue. In fact, because of this phenomenon, some have suggested that it is better to create thicker markets by bundling different impressions together (Ghosh, Nazerzadeh and Sundararajan 2007, Even-Dar, Kearns and Wortman 2007, McAfee, Papineni and Vassilvitskii 2010). The question of how to optimally bundle is a subject of ongoing research (Bergemann, Bonatti and Said 2011).

Since targeting increases total surplus, platforms would like to allow targeting while still
extracting the surplus this creates. This paper outlines a new and simple mechanism for doing so. We call it \textit{buy-it-now or take-a-chance} (BIN-TAC), and it works as follows. Goods are auctioned with a buy-it-now price $p$, set relatively high. If a single bidder is willing to pay the price, they get the good for price $p$. If more than one bidder takes the buy-it-now option, a second price auction is held between those bidders with reserve $p$. Finally, if no-one participates in buy-it-now, an auction is held in which the top $d$ bidders are eligible to receive the good, and it is randomly awarded to one of them at the $(d + 1)$-st price.

In this manner, we combine the advantages of an auction and a fixed price mechanism. When matches occur, advertisers self-select into the fixed-price buy-it-now option, allowing for revenue extraction. This is incentive compatible because in the event that they “take-a-chance” on winning via auction, there is a significant probability that they will not win the impression, even if their bid is the highest. On the other hand, when no matches occur, the auction mechanism ensures the impression is still allocated, thereby earning revenue.

The BIN-TAC mechanism is just another form of second-degree price discrimination. It is simple, both in that it is easy to explain to advertisers and in that it requires relatively little input from the mechanism designer: a choice of buy-it-now price, take-a-chance parameter $d$ and optionally a reserve in the take-a-chance auction. As we show through both theory and monte carlo simulation, BIN-TAC generally outperforms the two leading alternatives: a second price auction with reserve, or the “bundling” solution in which the platform withholds targeting information. At least in principle one could do better still by using the revenue-optimal mechanism suggested in Myerson (1981), which is considerably more complicated. We also show that when the valuations are drawn iid from a mixture of two regular distributions — a weighted combination of high and low valuation distributions with disjoint supports — our mechanism closely approximates the allocations and payments of the optimal mechanism, achieving similar performance. This new price discrimination mechanism is the first contribution of the paper.

The second contribution is to analyze a subset of historical data from the Microsoft Advertising Exchange. There has been some theoretical work describing the way advertising exchanges work (Muthukrishnan 2010, McAfee 2011), but there has been little empirical work. We document that there is a large gap between the highest and second highest valuations in these auctions, consistent with targeting creating thin markets. We show that advertisers vary their bids based on the location of their users, taking advantage of user demographics provided by the platform to achieve better matches.
By estimating the distribution of advertiser valuations, we can simulate the effect of introducing the BIN-TAC mechanism. We also consider a bundling strategy in which all impressions on a given webpage browsed by a user located in a particular geographic region are sold as identical products. We find that the optimal BIN-TAC mechanism generates 4.5% more revenue than the optimal second-price auction. Both outperform the bundling strategy, although we cannot rule out better performance from an optimal bundling strategy.

Related Work: Our work is related to the literature on price discrimination and screening. Here we consider a mechanism that treats all bidders symmetrically, and proceeds sequentially. Other papers have suggested sequential screening approaches. Courty and Li (2000) consider a setting where the buyers themselves learn their type dynamically, in two stages. In this case, offering contracts after the first type revelation but before the second may be optimal; see Bergemann and Said (2010) for a survey on dynamic mechanisms. In the static setting, sequential screening and posted-price mechanisms can be used to design optimal (or near-optimal) mechanisms when the bidders have multi-dimensional private information (see for example Rochet and Chone (1998) and Chawla, Hartline, Malec and Sivan (2010)).

More generally, the question of whether sellers should provide information that allows buyers to target their bids arises in the analysis of optimal seller disclosure (see for example Bergemann and Pesendorfer (2001)). The idea of bundling goods together to take advantage of negative correlation in valuations — in this case the negative correlation in the valuations from “match” or “no match” — dates back to Adams and Yellen (1976); see also McAfee, McMillan and Whinston (1989). Our paper is similar in style to Chu, Leslie and Sorensen (2011), who combine theory, simulations and empirics to argue that bundle-size pricing is a good approximation to the more complicated (but theoretically superior) mixed bundling pricing scheme for a monopolist selling multiple goods.

Finally, our model considers only the private values setting. Abraham, Athey, Babioff and Grubb (2010) consider an adverse selection problem that arises in a pure common value setting when some bidders are privately informed. This is motivated by the display advertising and advertisement exchange markets when some advertisers are better able to utilize information obtained from cookies. They show that asymmetry of information can sometimes lead to low revenue in this market.

From an empirical perspective, our paper contributes to the growing literature on online advertising and optimal pricing. Much of the work here is experimental in nature — for
example, Lewis and Reiley (2011) ran a randomized experiment to test advertising effectiveness, while Ostrovsky and Schwarz (2009) used an experimental design to test the impact of reserve prices on revenues. There has also been recent work on privacy and targeting in online advertising (Goldfarb and Tucker 2011b, Goldfarb and Tucker 2011a).

Organization: The paper proceeds in three parts. First, we give an overview of the market for display advertising. In the second part we introduce a stylized environment, and prove existence and characterization results for the BIN-TAC mechanism. We also provide analytic results concerning the revenue maximizing parameter choices, and compare our mechanism to others using both theory and monte carlo simulation. Finally, in the third part we provide an empirical analysis of a premium exchange marketplace, including counterfactual simulations of our mechanism’s performance. All proofs are contained in the appendix.

2 The Display Advertising Market

This paper proposes a new second degree price discrimination strategy for advertising platform markets. In these markets, advertisers care about the characteristics of their audience (user demographics), but it is up to the platform to choose whether or not to disclose what they know about the users. We are interested in analyzing pricing mechanisms that work in “real-time”, in the sense that as users arrive on the one side of the market, they are matched with advertisers at a simultaneously determined price.

The online display advertising market is an example of such a market. On one side of the market are the “publishers”: these are websites who have desirable content and therefore attract Internet users to browse their sites. These publishers earn revenue by selling advertising slots on these sites, small pieces of webpage “real estate” in standardized sizes.

The other side of the market consists of advertisers. They would like to display their advertisements to users browsing the publisher’s websites. They are buying user attention. Each instance of showing an advertisement to a user is called an “impression”. Advertiser demand for each impression is determined by which user they are reaching, and what the user’s current desires or intent are. For example, a Ferrari dealer might value high income users located close to the dealership. A mortgage company might value people that are reading an article on “how to refinance your mortgage” more than those who are reading an
article on “ways to survive your midlife crisis”, while the dealership might prefer the reverse. Some large publishers, primarily AOL, Microsoft and Yahoo!, sell directly to advertisers. Since the number of users browsing such publishers is extremely large (e.g. 1.5% of total worldwide Internet pageviews are on Yahoo!1), they can predict with high accuracy their user demographics. Consequently, they think of themselves of having a known inventory, consisting of a number of products in well-defined buckets: for example, male 15-24 year olds living in New York City viewing the Yahoo! homepage. They can thus contract to sell 1 million impressions delivered to a target demographic to a particular advertiser. Provided they have the inventory, they should be able to fulfill the contract. Transactions of this kind are generally negotiated between the publisher and the advertiser.

Alternatively, content is sold by auction through a centralized platform called an advertising exchange. Examples of leading advertising exchanges include the Microsoft Advertising Exchange (a subset of which we examine in this paper), Google’s DoubleClick, and Yahoo’s RightMedia.2 Advertising exchanges are a minor technological wonder, as they work in real-time. When a user loads a participating publisher’s webpage, a “request-for-content” is sent to the advertising exchange. This request will specify the type and size of advertisement to be displayed on the page, as well as information about the webpage itself (potentially including information about its content), and information about the user browsing the page.3

The advertising exchange will then either allocate the impression to an advertiser at a previously negotiated price, or hold a second-price auction between participating advertisers. If an auction is held, all or some of the information about the webpage and user is passed along to ad brokers who bid on behalf of the advertisers. These ad brokers can be thought of as proprietary algorithms that take as input an advertiser’s budget and preferences, and output decisions on whether to participate in an auction and how much to bid. The winning bidder’s ad is then served by the ad exchange, and shown on the publisher’s webpage.4

The bids placed in the auction are jointly determined by the preferences advertisers have, the ad broker interface and the disclosure policies of the ad exchanges or the publishers they interact with.

---

1Source: alexa.com
2“In Sept 2009, RightMedia averaged 9 billion transactions a day with 100s of thousands of buyers and sellers.” Muthukrishnan (2010)
3For example, it may include their IP address and cookies that indicate their past browsing behavior.
4To make things yet more complicated, in some ad exchanges, two different pricing models coexist. The first is pay-per-impression, which is what we analyze in the current paper; the second is pay-per-click, where the payment depends on whether or not the user clicks on the advertisement. Ad exchanges use expected click through rates to compare these different bids through a single expected revenue number.
represent. The ad brokers can only condition the bids they place on the information provided to them: if the user’s past browsing history is not made available to them, they can’t use it in determining their bid, even if their valuation would be influenced by this information. Similarly, the advertisers are constrained in expressing their preferences by the technology of the ad broker: if the algorithm doesn’t allow the advertiser to specify a different willingness to pay based on some particular user characteristic, then this won’t show up in their bids.

Ad exchanges have two main advantages over direct negotiation. First, they economize on transaction costs, by creating a centralized market for selling ad space. Second, they allow for very detailed products to be sold, such as the attention of a male 15-24 year old living in New York City viewing an article about hockey that has previously browsed articles about sports and theater. There is no technological reason why the products need to be sold in “buckets”, as publishers tend to do when guaranteeing sales in advance. This “real-time” sales technology is often touted as the future of this industry, as it potentially improves the match between the advertiser and their target audience.

3 Buy-it-Now or Take-a-Chance

We next describe our BIN-TAC mechanism. Like most auction formats, it is designed to allocate a single good among many potential buyers. A *buy-it-now price* $p$ is posted. Buyers simultaneously indicate whether they wish to *buy-it-now* (BIN). In the event that exactly one bidder elects to buy-it-now, that bidder wins the auction and pays $p$. If two or more bidders elect to BIN, a second-price sealed bid auction with reserve $p$ is held between those bidders. Bidders who chose to BIN are obliged to participate in this auction. Finally, if no-one elects to BIN, a sealed bid *take-a-chance* (TAC) auction is held between all bidders, with a reserve $r$. In that auction, one of the top $d$ bidders is chosen uniformly at random, and if that bidder’s bid exceeds the reserve, they win the auction and pay the maximum of the reserve and the $(d + 1)$-th bid. Ties among $d$-th highest bidders are broken randomly prior to the random allocation.

3.1 A Motivating Example

This section offers a simple numerical example that illustrates how BIN-TAC can outperform both a second price auction with targeting, and a second price auction with bundling.
Suppose a platform knows two binary user characteristics: whether the user is male or female, and whether they are older than 25 or not. This splits the user population into four demographic categories. There are four potential advertisers, each with a different target demographic. For simplicity, suppose all four advertisers will pay 10c an impression for their particular target demographic, and 2c otherwise. Assume that in the population all four demographic groups are equally likely. So without knowing the user characteristics an advertiser is willing to pay $10 \times 0.25 + 2 \times 0.75 = 4c$ for a generic impression. This is the revenue per impression the firm could earn from a second price auction with bundling. Expected consumer surplus is zero, implying total surplus of 4c per impression.

On the other hand, if they reveal the match information, exactly one of the advertisers will value it at 10c; the rest will value it at 2c. So then the per impression revenue from a second-price auction will be the second highest bid, 2c per impression. This makes it clear why the bundling solution may improve revenue. But on the other hand this mechanism is fully efficient, with total surplus of 10c per impression.

Now consider a BIN-TAC mechanism of the following form: the buy-it-now price is set at 8c, and if a take-a-chance auction is held, the item is awarded randomly to any bidder who bids at least 2c. As before, there is exactly one bidder who matches and has valuation 10c. That bidder is indifferent between buying-it-now and taking-a-chance: if they BIN, they get a guaranteed surplus of $10 - 8 = 2$, whereas if they take-a-chance, they get a 1/4 chance of a payoff of $10 - 2 = 8$. Breaking indifference in our favor, they will take the BIN option. Revenue per impression is therefore 8c — higher than both the second price auction options — and the allocation remains fully efficient.

This works exceptionally well, so one might wonder how robust this mechanism is to different assumptions about the information structure. Two will prove to be particularly important. First, in the example match probabilities were perfectly negatively correlated across bidders: if advertiser A matched, advertisers B, C and D did not. This artificially hurts the revenue of the second price auction with targeting. In the theory that follows we will not make this extreme informational assumption, but will instead assume that match probabilities are independent. We focus on independence for two reasons: first, it is an assumption that is often made in the screening and mechanism design literatures; second, because in the log data examined in this paper we observe little correlation in bids.\footnote{In practice, the information that platforms may choose to disclose is multidimensional, and some user characteristics may be “vertical” (e.g. income) and therefore induce positive correlation in match probabil-}
The second is that in the example it was certain that at least one advertiser would match. In fact, an important concern with providing “too much” targeting information is that no advertisers will want to target certain user demographics. So in the model we will relax this assumption, allowing for situations in which the total probability of a match is low.

### 3.2 The Model

A seller (publisher) has an impression to sell in real time, and they have information both about the webpage content and the user (summarized in a cookie). The seller is considering one of two policies: either disclosing the cookie content to the advertiser (the “targeting” policy), or withholding it (the “bundling” policy). When they allow targeting, bidders know whether the user is a “match” for them or not. When a “match” occurs, the bidder has high valuation. But this is rare, and so generally everyone in the auction has low valuation. This set of assumptions hardens in the idea that allowing targeting may make markets thin. Instead if the seller chooses to withhold the cookie, bidders are uncertain about the realization of the match and therefore have intermediate valuations.

The formal model is as follows. There are $n$ symmetric bidders who participate in an auction for a single good which is valued at zero by the seller. Bidders are risk neutral. They have value $V_H$ for the good when a match occurs, and value $V_L$ for the good if no match occurs, where $V_L \sim F_L$ and $V_H \sim F_H$. We assume that $F_L$ has support $[\omega_L, \bar{\omega}_L]$ and $F_H$ has support $[\omega_H, \bar{\omega}_H]$, and that these supports are disjoint (so $\bar{\omega}_L < \omega_H$). We assume both $F_L$ and $F_H$ have continuous densities $f_L$ and $f_H$.

The Bernoulli random variable $X$ indicates whether a match has occurred, and the event $X = 1$ occurs with probability $\alpha \in (0, 1)$. We are particularly interested in the case when $\alpha$ is small. The triple $(X, V_L, V_H)$ is drawn identically and independently across bidders, so that a user who is a match for one advertiser need not be a match for the others. In the case with targeting, each advertiser’s valuation $V = (1 - X)V_L + XV_H$ is private information, known only to the advertiser. Instead if the seller bundles all impressions, the advertiser knows $V_L$ and $V_H$ but does not know the realization of $X$, implying their expected valuation is $E[V] = (1 - \alpha)V_L + \alpha V_H$.

We also make some technical assumptions on the virtual valuations $\psi(v) = v - \frac{1-F(v)}{f(v)}$. When

---

*Notice: while others may be “horizontal” and have correlation implications that depend on the population of advertisers (e.g. age).*
ψ(v) is strictly increasing, the optimal mechanism is a second-price auction with a reserve price (Myerson 1981). We assume that ψ(v) is continuous and increasing over the regions $[\omega_L, \omega_L]$ and $[\omega_H, \omega_H]$. For simplicity of the presentation, we additionally assume that ψ(v) single-crosses zero and that this intersection occurs in the low valuation region $[\omega_L, \omega_L]$.

But the important thing to notice is that the virtual valuations are (infinitely) negative over the region $(\omega_L, \omega_H)$ since $F$ is unsupported on this region. This implies that the second price auction with reserve is not revenue-optimal, which is why we propose an alternative.

3.3 Equilibrium Analysis

This is a sequential mechanism which we analyze by backward induction. The auctions that follow the initial BIN decision admit simple strategies. If multiple players choose to BIN, the allocation mechanism reduces to a second-price auction with reserve $p$. Thus, it is weakly dominant for players to bid their valuations.

Truth-telling is also weakly dominant in the TAC auction. The logic is standard: if a bidder with valuation $v$ bids $b' > v$, it can only change the allocation when the maximum of the $d$-th highest rival bid and the reserve price is in $[v, b']$. But whenever this occurs, the resulting price of the object is above the bidder’s valuation and if she wins she will regret her decision. Alternatively, if she bids $b' < v$, when she wins the price is not affected, and her probability of winning will decrease.

Taking these strategies as given, we turn to the buy-it-now decision. Intuitively, the BIN option should be more attractive to higher types: they have the most to lose from either random allocation (they may not get the good even if they are willing to pay the most) or from rivals taking the BIN option (they certainly do not get the good). This suggests that in a symmetric equilibrium, the BIN decision takes a threshold form: $\exists v$ such that types with $v \geq \overline{v}$ elect to BIN, and the rest do not. This is in fact the case.

Prior to stating a formal theorem, we introduce the following notation. Let the random variable $Y_j$ be the $j$-th highest valuation from $n - 1$ iid samples from $F$ and let $Y^*$ be the maximum of $Y_d$ and the TAC reserve $r$.

**Theorem 1 (Equilibrium Characterization)**

\[6\] Without this assumption we would have to analyze multiple cases, which is straightforward but tedious.

\[7\] Since participation is obligatory at this stage, the minimum allowable bid is $p$; but no bidder would take the BIN option unless they had a valuation of at least $p$. 

9
Assume $d > 1$ and $p \leq \frac{d-1}{d}a_H + \frac{1}{d}E[Y^*]$. Then there exists a unique symmetric pure strategy Bayes-Nash equilibrium of the game, characterized by a threshold $\bar{v}$ satisfying:

$$\bar{v} = p + \frac{1}{d}E[\bar{v} - Y^* | Y^1 < \bar{v}]$$

(1)

Types with $v \geq \bar{v}$ take the BIN option; and all types bid their valuation in any auction that may occur.

Equation (1) is intuitive: At what point is a bidder indifferent between the BIN and TAC options? If strategies are increasing, the only time the choice is relevant is when there are no higher valuation bidders (since otherwise those bidders would BIN and win the resulting auction). So if a bidder has the highest value and chooses to BIN, they get a surplus of $v - p$. Choosing to TAC gives $\frac{1}{d}E[v - Y^* | Y^1 < v]$, since they only win with probability $\frac{1}{d}$, although their payment of $Y^*$ is on average much lower. Equating these two to find the indifferent type $v$ yields Equation (1).

Now we consider the revenue-maximizing choices of the design parameters: the BIN price $p$, the TAC reserve $r$ and the TAC parameter $d$. We have not been able to get a nice characterization of the optimal $d$, as it is an integer programming problem which doesn’t admit standard optimization approaches. However, for a given $d$, the optimal BIN price and TAC reserve are given by some familiar looking equations. Again, we must introduce some notation. Let $p(\bar{v}, r) = \bar{v} - \frac{1}{d}E[\bar{v} - Y^* | Y^1 < \bar{v}]$ be the solution of Equation (1), expressing the BIN price as a function of the threshold and the TAC reserve. Let $R(\bar{v}, r)$ be the conditional expected revenue from a TAC auction when the highest valuation is exactly equal to $v$ and the reserve is $r$. Then we have the following theorem.

**Theorem 2 (Optimal Buy Price and Reserve)** For any $(p, d)$, the revenue-maximizing TAC reserve $r^*$ satisfies:

$$r^* = \frac{1 - F(r^*)}{f(r^*)}$$

(2)

The optimal BIN price is given by $p(\overline{v}^*, r^*)$ where $\overline{v}^*$ is the solution of the equation below:

$$f(\overline{v}) (((n-1)(1 - F(\overline{v}))(\overline{v} - p(\overline{v}, r^*)) + F(\overline{v})(p(\overline{v}, r^*) - R(\overline{v}, r^*)) = (1-F(\overline{v}))F(\overline{v})\frac{\partial p(\overline{v}, r^*)}{\partial \overline{v}}$$

The assumption that $p \leq \frac{d-1}{d}a_H + \frac{1}{d}E[Y^*]$ rules out uninteresting cases where the BIN price is so high that no-one ever chooses BIN.
If no such solution exists in \([\omega_H, \omega_H]\), then \(\overline{v}^*\) is equal to \(\omega_H\).

Equation (2) is somewhat surprising; the optimal TAC reserve is exactly the standard reserve in Myerson (1981), ensuring that no types with negative virtual valuation are ever awarded the object. This is despite the fact that our BIN-TAC mechanism is not the optimal mechanism. The key insight is that the TAC reserve is relevant for the BIN choice. Raising the TAC reserve lowers the surplus from participating in the TAC auction, and so the seller can also raise the BIN price while keeping the indifferent type \(\overline{v}\) constant. So the trade-off is exactly the usual one: raising the TAC reserve extracts revenue from types above \(r^*\) — even those above \(\overline{v}\) — at the cost of losing revenue from the marginal type. This is why we get the usual solution.

On the other hand, the implicit equation for the optimal BIN price is new. To get some intuition, notice that the BIN price in some sense sets a reserve at \(\overline{v}\). If two bidders meet the reserve, the seller gets the second highest bid; if only one, the BIN price; and if none, he gets the TAC revenue. So a marginal increase in the threshold has three effects. First, if the highest bidder has valuation exactly equal to the threshold, following an increase she will shift from BIN to TAC. This costs the seller \(p(\overline{v}, r^*) - R(\overline{v}, r^*)\). Second, if the second highest bidder has valuation equal to the threshold, an increase will knock her out of the BIN auction, and the seller’s revenue falls by \(\overline{v} - p(\overline{v}, r^*)\). Finally, if the highest bidder is above the reserve and the second highest is below, an increase gains the seller \(\partial p(\overline{v}, r^*) / \partial \overline{v}\). Working out the probabilities of these various events, and setting the expected value equal to zero, we get the result.

Sometimes there is no solution for \(\overline{v}^*\) in \([\omega_H, \omega_H]\). This occurs whenever the high valuations are substantially larger than the low valuations (i.e. \(\omega_H \gg \omega_L\)), so that it is not profitable to randomize the allocation for any of the high types. In this case the BIN price is set at \(p(\omega_H, r^*)\) so that the lowest high type at \(\omega_H\) is indifferent between TAC and BIN.

### 3.4 Performance Comparisons

We compare our mechanism to three benchmark mechanisms: the second price auction with targeting, the second price auction under bundling, and the Myerson (1981) mechanism with ironing. The latter is optimal within the class of mechanisms that allow targeting.\(^9\)

\(^9\)A seller may do better by withholding match information altogether (bundling), or by selling the rights to this information, possibly using sequential offers — see Bergemann and Pesendorfer (2001).
Note firstly that the SPA with targeting and reserve $r_0$ is just a special case of BIN-TAC for parameters $p = \omega_H$, $d = 1$ and TAC reserve $r = r_0$. In this case no-one takes the excessively high BIN price, and since $d = 1$ the TAC auction is just a second price auction. It follows immediately that BIN-TAC achieves at least weakly higher revenue performance.

How about bundling? In general it is hard to say if this performs better or worse than BIN-TAC: if $F_L$ exhibits sufficient dispersion, the revenue loss from randomizing the allocation among low value bidders may be greater than the revenue gained from extracting revenue at the top by providing information. However, we can show that this approach is worse than BIN-TAC in a special case of our environment. Assume that $F_L$ and $F_H$ are degenerate with all mass on $v_L$ and $v_H$ respectively, so that the only source of private information is the match variable $X$. Without targeting, all valuations are identical and equal to $(1 - \alpha)v_L + \alpha v_H$. This then is the revenue from the second price auction without targeting. With targeting, the second price auction returns the second order statistic of the realized valuations, which is equal to $v_H$ if two bidders match, and $v_L$ otherwise. Whether this is higher or lower in expectation is jointly determined by the number of bidders $n$, and the match probability $\alpha$: as either parameter increases, targeting becomes relatively more preferable.

The optimal BIN-TAC mechanism does better than this though. Consider the mechanism where the TAC parameter $d = n$ and a reserve $r = v_L$ is set, so that a high valuation bidder faces a lottery if she doesn’t take the BIN mechanism. Then a BIN price equal to $\frac{n-1}{n} v_H + \frac{1}{n} v_L$ will still induce bidders (weakly) to take the BIN option. Revenue is now $v_H$ if two bidders match, $\frac{n-1}{n} v_H + \frac{1}{n} v_L$ if a single bidder matches and $v_L$ otherwise. As we prove in the appendix, this is better than the bundling solution as long as $n \geq 3$. Summarizing:

**Theorem 3** Suppose $F_L$ and $F_H$ are degenerate. Then the revenue from the optimal BIN-TAC model is higher than the optimal second price auction without targeting, strictly for $n \geq 3$.

**Optimal Mechanism:** We now compare BIN-TAC to the optimal mechanism, which we derive below. As you will see, it is considerably more complex and harder for a platform to explain to potential advertisers. Let $r^*$ be defined as in Eq. (2). We show in the appendix that whenever $\alpha \omega_H \geq r^*(1 - F(r^*))$, the optimal mechanism is just a second-price auction with reserve equal to $\omega_H$. So suppose $\alpha \omega_H < r^*(1 - F(r^*))$. 

12
Then, there exists $\tilde{v}, r^* \leq \tilde{v} \leq \omega_L$, such that

$$(2 - \alpha - F(\tilde{v}))F(\tilde{v}) + \alpha(\omega_H - \tilde{v})f(\tilde{v}) = 1 - \alpha$$

(3)

Define the ironed virtual valuations as follows:

$$\phi(v) = \begin{cases} 
0 & v \in [\omega_L, r^*_2) \\
\psi(v) & v \in [r^*_2, \tilde{v}] \\
\psi(\tilde{v}) & v \in (\tilde{v}, \omega_H) \\
\psi(v) & v \in [\omega_H, \omega_H],
\end{cases}$$

(4)

The allocation procedure works like this: award the good to the bidder with the highest ironed virtual valuation, breaking ties at random, provided the virtual valuation is positive. Notice that all types between $v^*$ and $\omega_H$ get the same ironed virtual valuations, and therefore if they tie, the winner is selected at random. Like BIN-TAC, this is inefficient, but allows additional revenue extraction from higher types.

The payments are determined as follows. Whenever the virtual valuation of the second highest bidder has a unique inverse (i.e. outside of the ironed region between $v^*$ and $\omega_H$), the winning bidder pays the maximum of the reserve and the valuation of the second highest bidder (as in a second price auction with reserve). Whenever both the highest and second highest bidder have virtual valuations in the ironed region, the required payment is $v^*$. Finally, when the winning bidder has valuation above $\omega_H$, but $k$ other bidders have valuations in the ironed region, the winner pays $\frac{1}{k+1}(k\omega_H + v^*)$. This last condition follows from incentive compatibility: the bidder on the top margin of the ironed region, type $\omega_H$, gets a payoff of $v - \frac{1}{k+1}(k\omega_H + v^*)$ when the highest bidder; but could alternately pretend to be in the ironed region, with payoff $\frac{1}{k+1}(v - v^*)$. These two payoffs must be equal.

**Theorem 4 (Optimal Mechanism)** Suppose $\psi(\omega_L) \leq \psi(\omega_H)$. If $\alpha\omega_H \geq r^*_2(1 - F(r^*))$, then the optimal mechanism is the second-price auction with reserve $\omega_H$. If $\alpha\omega_H < r^*_2(1 - F(r^*_2))$, then the ironed-mechanism described above is optimal.

Having obtained this characterization, we can compare BIN-TAC with the optimal mechanism. It is easy to prove that as either $n \to \infty$, $\alpha \to 1$ or $\omega_H/\omega_L \to \infty$, the BIN-TAC mechanism converges to the optimal mechanism. This, however, is not particularly interest-
Figure 1: **Comparison of Allocations and Payments.** Allocation probabilities (top panel) and expected payments (bottom panel) for the OPT, SPA and BIN-TAC mechanisms when the distributions $F_L$ and $F_H$ are uniform. The $x$-axis corresponds to the bid.

We show this in a particular case, where $F_L$ is uniform over $[0, 1]$ and $F_H$ is uniform over $[\Delta, \Delta + 1]$ for $\Delta \geq 3$. By Theorem 4, we have:

$$r^* = \frac{1}{2(1 - \alpha)} \quad \text{and} \quad v^* = \left(1 - \sqrt{\frac{\alpha(\Delta - 1)}{1 - \alpha}}\right).$$

Table 1 compares the expected revenue and welfare obtained by all the mechanisms we have considered for $n = 5$, $\Delta = 3$ and $\alpha = 0.05$. As you can see from the table, the performance of BIN-TAC is close to OPT (about 96% of OPT), much better than the optimal SPA (85%).
Figure 1 helps explain this. The top panel depicts the probability of allocation and the bottom panel the expected payment as bidder type varies along the x-axis. As you can see, BIN-TAC approximates the discontinuous increase in allocation probability at \( v^* \) with a smooth curve, whereas the SPA increases the probability of allocation to a much higher level. As a result, the SPA cannot extract revenue from the high types (who could easily pretend to be a lower type without losing much), while BIN-TAC has similar revenue performance to OPT at the top. The table also shows that bundling performs less well than both BIN-TAC and OPT, especially in terms of expected consumer surplus. This is because it often fails to match advertisers and users correctly.

Table 1: Revenue Comparison: Uniform Environment

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>OPT (d=2)</th>
<th>SPA (d=2)</th>
<th>BIN-TAC (d=2)</th>
<th>BIN-TAC (d=3)</th>
<th>Bundling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected Revenue</td>
<td>0.89</td>
<td>0.76</td>
<td>0.85</td>
<td>0.83</td>
<td>0.81</td>
</tr>
<tr>
<td>Expected Consumer Surplus</td>
<td>0.51</td>
<td>0.67</td>
<td>0.48</td>
<td>0.40</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Expected revenue and welfare under different mechanisms, for the uniform environment with \( \Delta = 3, \alpha = 0.05 \) and the number of bidders \( n = 5 \).

3.5 Monte Carlo Simulations

We would like to test our mechanism against the benchmarks in a variety of other settings. We drop the assumption that \( F_L \) and \( F_H \) have disjoint support. The optimal BIN-TAC mechanism is reasonably easy to calculate. Nothing in the proof of Theorem 2 required the disjoint supports for determining \( r^* \) and \( p^* \), and so these can be solved for numerically for each \( d \). Thus the optimization problem reduces to a one dimensional discrete optimization problem in the TAC parameter \( d \), which can be quickly solved. Finding the optimal mechanism is more challenging, but can be done using standard optimization techniques.

For our simulations, we restrict ourselves to location families where the distribution \( F_H(\cdot) = F_L(\cdot - \Delta) \) for some shift-parameter \( \Delta \), as in the uniform case above. This \( \Delta \) is the difference in mean valuation between the high and low groups. We consider two location families: one where \( V_L \) is normal, and another where \( V_L \) is log normally distributed. In both cases \( V_L \) has mean 1 and standard deviation 0.5. We allow \( \Delta, n \) and \( \alpha \) to vary across experiments, and compute \( r^*, p^* \) and \( d^* \) as discussed. The default parameters we consider are \( n = 10, \Delta = 5, \alpha = 0.05 \).
and $\alpha = .05$, and we vary one parameter at a time. Each experiment is repeated for 100000 impressions, and we calculate the average revenues.

The results are presented in Figures 4, 5 and 6. In all cases, on the y-axis we plot the revenue as a fraction of the revenue from the optimal mechanism. Recall that BIN-TAC generalizes the second price auction, so its performance is always at least as good, and often significantly better. In all cases, the BIN-TAC extracts at least 90% of the optimal revenue, compared to a worst-case performance of around 82% for the SPA with targeting. The SPA without targeting in some cases does even better than OPT (when there are very few bidders), but its performance sharply degrades as the probability or value of a match gets large.

We see this in Figures 4 and 5. The expected number of matches is $\alpha n$, and so as either increases, the performance of the mechanisms that allow targeting improves relative to the SPA without targeting. Over some range, BIN-TAC significantly outperforms the SPA, but as the number of bidders or the probability of match get sufficiently high, both converge to the OPT mechanism (which is itself an SPA with high reserve).

Figure 6 shows the dependence on the gap $\Delta$. As expected, the performance of BIN-TAC increases while that of a second price auction decreases as $\Delta$ gets larger, over some range. Since there is more revenue to be gained from high-valued bidders, BIN-TAC can only perform better with a large $\Delta$. For sufficiently high $\Delta$ though, both BIN-TAC and the SPA set high reserves, “throwing away” low-valued impression and extracting all their revenue from matches, with equal revenue performance.

Overall, the performance of BIN-TAC is very good, at least for the distributions and parameters chosen. The main caveats are that it doesn’t perform well with very few bidders (when no targeting is preferable), and has little to recommend it when matches are highly probable or very valuable (a second-price auction would do as well). Its niche, as we have argued, is in markets with relatively large numbers of bidders but low match probabilities, so that markets are “thin” in the sense of having relatively low matches in expectation.

4 Empirical Analysis

Our theoretical analysis has shown that there are cases in which BIN-TAC performs well. We now test our mechanism’s performance in a real-world setting. We have historical data from Microsoft Advertising Exchange, one of the leading ad exchanges. Our data comes from
a single large publisher’s auctions on this exchange and consists of a 0.1% random sample of a week’s worth of auction data from this publisher, sampled within the last two years. This publisher sells multiple “products”, where a product is a URL-ad size combination (e.g. a large banner ad on the sports landing page of the New York Times). This data includes information on both the publisher and advertiser side. On the publisher side, we see the url of the webpage the ad will be posted on, the size of the advertising space and the IP address of the user browsing the website. We form a unique identifier for the url-size pair, and call that a product. We determine which US state the user IP originates from, and call that a region. We use controls for product and region throughout the descriptive regressions. Unfortunately, we don’t have more detailed information on the product or the user, as the tags and cookies passed by the publisher to the ad exchange were not stored.

On the advertiser side, we see the bid they placed, the company name, the ad broker they employed, and a variable indicating the ad they intend to show. In the overwhelming majority of cases there is a single ad for each company, but some larger firms have multiple ad campaigns simultaneously. We treat these as being a single ad campaign in what follows. We observe who won the auction and the final price.

We drop auctions in which the eventual allocation was based on a previously negotiated price, rather than the bids. We also restrict attention to impressions that originate in the US, and where the publisher content is in English. Finally, we restrict only to reasonably frequently sold products, those with at least 100 sales in the dataset. This leaves us with a sample of 83515 impressions.

The dataset is summarized in Table 2. For confidentiality reasons, bids have been rescaled so that the average bid across all observations is equal to 1 unit. Bids are very skew, with the median bid being only 0.57 units. Perhaps as a consequence of this skewness, the winning bid — which is more heavily sampled from the right tail of the bid distribution — is much higher at 2.96 units. There are on average 6 bidders per auction, but there is considerable variation in participation, with a standard deviation of nearly 3. Bids are not strongly correlated: as the table shows, the correlation between a randomly selected pair of bids from each auction is only 0.01. This is not statistically significant at 5% (p-value 0.116, \( N = 15827 \)).

The advertisers are themselves quite active in the market. On average they bid on 0.7% of all

\footnote{That bids are not positively correlated should not be taken to mean that underlying valuations are not positively correlated; it could just be that informational and technological constraints prevent advertisers from fully expressing their preferences.}
impressions, and win nearly 40% of those they bid on. These averages are somewhat misleading though. The median advertiser is far less active, bidding on only 0.02% of impressions, while the most active advertiser participates in nearly 90% of auctions. Our hypothesis is that some advertisers choose to participate in relatively few auctions, but tend to bid quite highly and therefore win with relatively high probability. Others bid lower amounts in many auctions, and win with lower probability. The first strategy is followed by companies who want to place their advertisements only on webpages with specific content or to target specific demographics, while the latter strategy is followed by companies whose main aim is brand visibility.

4.1 Descriptive Evidence

Before proceeding to the main estimation and simulations, we provide some evidence that advertisers bid differently on different users (i.e. there is matching on user demographics). We also show that the platform is doing poorly in extracting this match surplus as revenue.

The first thing we note is that leading advertisers do vary their bids on the same product over short periods of time, which suggests that their willingness to pay depends on the demographics of the user viewing the webpage. Figure 2 shows re-scaled bids in 50 auctions by large five advertisers for the most popular webpage slot sold by this publisher. The

Figure 2: Bids over Time. The figure shows the (rescaled) bids of five advertisers in our data, selected at random from the top 50 advertisers (ranked by purchases) on 50 randomly chosen successive impressions of the most popular product. Note that the set of impressions differs across bidders (there are no impressions on which all 5 participate).
advertisers were chosen at random from the top 50 advertisers in our dataset (ranked by purchases). The 50 auctions are chosen to be consecutive for each bidder. The bids exhibit considerable variation, even though all of these impressions were auctioned within a 3-hour period. While this could in principle be driven by decreases in the advertisers' available budget, since the bids go both up and down it seems more likely that this variation arises from differences in user demographics.

One direct test of advertiser-user matching is to look for the significance of advertiser-user fixed effects in explaining bids. Specifically, we estimate an unrestricted model where the dependent variable is bids and the controls are advertiser-user dummies, versus a restricted model with just advertiser and user fixed effects, but not their interaction. The restricted model is overwhelmingly rejected by the data: the relevant F-statistic is over 15, while the 99% critical value is just over 1. This points towards matching on demographics.

Determining whether this matching actually creates surplus is a little more difficult. The only user demographic we observe is the user region, and it is hard to know a priori what the advertisers’ preferences over regions are. To get a handle on this, we turn to another proprietary dataset that indicates how often an advertiser’s webpage was viewed by internet users in different regions of the country during the calendar month prior to the auction. Our intention is to proxy for the advertisers’ geographic preferences (insofar as these exist) using this pageview data. The idea is that firms who operate in only a few regions probably attract all their pageviews from those regions, and also only want to advertise in those regions. If this is right, advertisers who attract a large fraction of their pageviews from a particular region should participate more frequently and bid higher on users from those regions. Because the pageview data dates from a period before our exchange data we are not worried about reverse causality (i.e. advertisers who win more impressions from region X later get more views from region X). We normalize the pageviews from a particular state by the state population to get a per capita pageview measure, and then construct the fraction of normalized pageviews each region receives, calling this the “pageview ratio”.

In Table 3, we present results from regressions of auction participation (a dummy equal to one if the advertiser participated), and bid (conditional on participation) on the pageview ratio, as well as a number of fixed effects. Because the sheer size of our dataset makes it

---

11Since the same set of bidders don’t participate in every auction, the impression number on the x-axis corresponds to different impressions for different bidders.

12For example, if these auctions were in May, the pageview data would be taken from April.
difficult to run the fixed effect regressions, we run this on a subsample consisting of the top 10% of advertisers. The first column shows participation as a function of the pageview ratio, as well as product-region fixed effects, and time-of-day fixed effects (since participation and bids may vary with the user’s local time). We find a positive but insignificant effect. But when we include advertiser fixed effects to control for different participation frequencies across advertisers, we find a much bigger and now highly significant effect. All else equal, an advertiser is 3.3% more likely to bid on a user from a state that contributes 10% of the population-weighted pageviews for their site. This is a large increase, as the average baseline probability of participation is only around 1%.

Turning to the bids, we find similar estimates and significance levels from the specifications with and without advertiser fixed effects. We find that firms bid higher on users from more relevant regions, although this effect is relatively modest in economic terms. Given that our proxy for advertiser preferences is relatively crude, it is notable that we find these effects. This provides some evidence that the matching is surplus increasing, in that advertisers are able to target regions where their most active customers (as measured by pageviews) are.

The second stylized fact we note about this market is that there is often a substantial gap between the highest and second highest bid in the auction. To facilitate bid comparisons, we

---

Figure 3: **Bidding Gap and Virtual Valuations.** The left panel shows a kernel density estimate of the pdf of the (normalized) gap between the highest and second highest (rescaled) bids in auctions for the product with the highest sales volume in our dataset. The right panel shows the estimated virtual valuations as a function of bids.

---

---

13Since participation is highly skewed, these advertisers account for 90% of the bids. With only bidder fixed effects we could use a within transformation to reduce the computational burden; but unfortunately this is not possible with multiple non-interacting fixed effects.
look at the product with the highest sales volume in the data (over 38% of all impressions). The left panel of Figure 3, shows a kernel density estimate of this gap. The average bid in an auction is 0.88, while the mean gap is much larger at 1.89, indicating that there is a lot of money left on the table by a second-price mechanism (see Table 2 for other summary statistics). That gap itself is extremely skewed.

Assuming bids are equal to valuations — an assumption we will motivate in the next section — the right panel shows the virtual valuations $\psi(v)$ as a function of the bids. Although the virtual valuations are never infinitely negative, as in our stylized model, they are certainly non-monotone. This implies that BIN-TAC may be able to extract more revenue than a second price auction. We test this in the next section.

4.2 Estimation and Counterfactual Simulations

Our theoretical model is of a single auction with a particular valuation structure, rather than a whole market with a general valuation structure, and so in order to provide microfoundations for our simulation approach, we need to enrich the model.

We make the following assumptions for the estimation and counterfactual simulations. There is a fixed set of $N$ bidders who are always present in the market. As in the text, the model is symmetric independent private values. Each bidder draws their valuations for each impression identically, independently and privately according to some distribution $F_j$ supported on $[0, \infty)$ (where $j$ indexes products). Independence of valuations is a strong assumption, as it rules out common preferences for particular user demographics. For example, it rules out the possibility that all bidders prefer high income bidders, in which case we would observe positive correlation in bids. Some partial support for this assumption comes from the lack of bid correlation reported in Table 2. The symmetry assumption is also strong — and probably rejected by the data given the significance of the advertiser fixed effects in the reduced form regressions — but helps to keep the problem computationally tractable. To address the concern that the symmetry and independence assumptions are driving our results, we will do some robustness tests based on different informational assumptions in a later subsection.

\[14\] The assumption that bidders are continuously present in the market is in principle relatively innocuous since bidding is done by ad broker algorithms. Yet some bidding algorithms ignore certain auctions in order to respect advertiser budget constraints. We will not model this “inattention”, especially because it is hard to rationalize such behavior as optimal: bidding close to zero has almost no effect on the budget constraint since the maximum possible payment in a second-price auction is bounded above by the bid.
From the summary statistics we also know that participation varies across advertisers. We assume that participation costs are zero, and thus we can infer from non-participation that an advertiser has zero valuation for the impression (since with any positive valuation there is weakly positive surplus from bidding). This may seem like a strong assumption, but given that the 5th percentile of bids in our data is equal to 0.013 — with an almost zero probability of winning, and even lower surplus — it is hard to believe that participation costs are substantially different from zero. One reason for this may be that bidding is automated.

Given these assumptions, we are able to make the following inference from the second-price auction data. If bidder $i$ makes a bid of $b_{i,t}$ in auction $t$, their valuation is $b_{i,t}$, since it is weakly dominant for them to bid their valuations. Moreover, if bidder $i$ did not participate in auction $t$, their valuation for that particular impression must have been zero. Since there is a one-to-one mapping from the distribution of bids and participation to the valuations, $F_j$ is non-parametrically identified. We could therefore estimate the valuation density for each product using non-parametric methods. But, as we will show below, the counterfactual simulations will never require estimating more than some conditional moments of order statistics (e.g. the expected value of the $d$-th highest valuation when the highest valuation is less than $\bar{v}$). So instead we estimate these moments by the corresponding sample average. These estimates will be more robust to mis-specification of the information structure.$^{15}$

We are interested in comparing the “optimal” BIN-TAC mechanism to other leading mechanisms. For simplicity, we restrict attention throughout to the class of mechanisms that make the same parameter choices for all products (i.e. we rule out different reserves or randomization parameters by product or user-region). In each case we find these optimal parameters by maximizing the revenue functions defined in equations (5) and (6) below, using standard optimization methods. To get standard errors on our revenue and consumer surplus estimates, we bootstrap the estimation sample and re-run the simulation procedure, holding the parameter choices fixed.$^{16}$

**Mechanisms with Targeting:** The two policies we want to compare here are the second price auction with targeting and BIN-TAC. The two SPA mechanisms are easiest. For example, with a reserve of $r$, the expected revenue depends on the joint distribution of the

---

$^{15}$For example, if bidders are asymmetric, it will still be the case that the average highest bid is a consistent estimator of the expected highest valuation; which will not be true of the distribution of the highest order statistic derived from the non-parametrically estimated distribution of $F$ from the misspecified model.

$^{16}$We use 100 bootstrap samples (i.e. samples of $T$ impressions drawn randomly with replacement).
top two valuations: since bidders bid their valuations, the item sells if the highest valuation exceeds \( r \), and then the revenue is the maximum of the second highest bid and \( r \). Letting the \( k \)-th highest bid in an auction \( t \) be \( b_t^{(k)} \), our estimate is then given by the sample average across the \( T \) auctions:

\[
\text{Revenue}^{\text{SPA}}(r) = \frac{1}{T} \sum_{t=1}^{T} 1(b_t^{(1)} > r) \max\{b_t^{(2)}, r\}
\]  

(5)

BIN-TAC is harder, as an agent’s equilibrium decision to take the BIN option depends on their beliefs about the distribution of rival valuations. From the model, advertiser behavior is characterized by a threshold value \( \overline{v}_j = \overline{v}_j(p, d, r) \) for each product, above which they will take the BIN option, and below which they will TAC. From Theorem 1, this threshold solves the implicit equation \( \overline{v}_j - p = \frac{1}{d} \mathbb{E}[\overline{v} - Y^*|Y^1 < \overline{v}_j] \), where \( Y^* = \max\{Y^d, r\} \) and \( Y^1 \) and \( Y^d \) are the 1st and \( d \)-th order statistics of rival bids on product \( j \). To solve this equation for fixed \( (p, d, r) \), we need to estimate the expected TAC payment \( \mathbb{E}[Y^*|Y^1 < s] \) for varying \( s \).

Under symmetry, the joint distribution of valuations is exchangeable, and so the joint distribution of rival bids is exactly the same as the joint distribution of \( N - 1 \) randomly selected bids. So our estimate of the TAC payment conditional on winning on product \( j \) is given by:

\[
\text{TAC Payment}(s, r) = \frac{1}{T} \sum_{t=1}^{T} \frac{\sum_k 1(b_t^{(1)} < s) \max\{b_t^{(d)}, r\}}{\sum_k 1(b_t^{(1)} < s)}
\]

where \( k \) indexes the \( N \) choices of \( N - 1 \)-length bid vectors for each auction, including zeros for bidders that didn’t participate and restricting the sample only to product \( j \). We can then solve for the equilibrium \( \overline{v}(p, d, r) \) for each set of BIN-TAC parameters \( (p, d, r) \), and get a revenue estimate as follows:

\[
\text{Revenue}^{\text{BIN-TAC}}(p, d, r) = \frac{1}{T} \sum_{t=1}^{T} 1(b_t^{(2)} \geq \overline{v}(p, d, r)) b_t^{(2)} + \frac{1}{T} \sum_{t=1}^{T} 1(b_t^{(1)} \geq \overline{v}(p, d, r) > b_t^{(2)})p
\]

\[
+ \frac{1}{T} \sum_{t=1}^{T} 1(b_t^{(1)} < \overline{v}(p, d, r)) \sum_{j=1}^{d} 1(b_t^{(j)} \geq r) \max\{b_t^{(d+1)}, r\}
\]

(6)
Bundling Mechanisms: As we do not observe all the impression characteristics provided to advertisers in this market, we cannot consider the optimal bundling strategy. But we can consider bundling by product and user region, where the platform strips away all other user characteristics except for the region, so that advertisers are buying a random impression of a given size, on a given website, being viewed by a user from a particular US state. This is unlikely to be optimal, but provides a lower bound on the revenues from bundling.

For this analysis, we allow for bidder valuations to be asymmetric and vary by both product and region. Our estimate of their willingness to pay for this “generic impression” is just their average bid across all auctions of this product-region combination, taking their implicit bids when they didn’t participate as equal to zero. Given that participation costs are zero and all bidders have strictly positive mean valuations, in the counterfactual world all bidders will participate in all auctions. We assume that these impressions are sold by second-price auction without reserve (since the bundling creates thick markets, a reserve isn’t necessary).

Robustness to Informational Assumptions: The above theory and structural estimation follows the empirical auctions literature in treating bidder’s valuations as private information. A different modeling approach was suggested in an influential paper by Edelman, Ostrovsky and Schwartz (2007). They proposed a complete information model of sponsored search auctions. Their logic was that since these players compete with high frequency and can potentially learn each others’ valuations, a complete information model may be a better approximation to reality than an incomplete information model.

Following this intuition, we also consider counterfactual simulations under complete information. The only model this affects is the BIN-TAC model, as under weak refinements the SPA equilibria under incomplete and complete information coincide. However in the BIN-TAC model we unfortunately now have multiple equilibria.

To see this, consider a case where the bidder with the highest valuation is going to take the BIN option. Then the remaining bidders are indifferent between BIN and TAC, since in either case they will lose the auction and get payoff 0. We employ a trembling hand perfection refinement to eliminate this multiplicity. Specifically, for any probability $\epsilon > 0$ that the highest bidder will take the TAC option instead, the second-highest bidder faces a

---

17 See for example Laffont and Vuong (1996). See also Athey and Nekipelov (2010) for a model of sponsored search models in this tradition.

18 This arises also in the generalized second price auction — see Edelman et al. (2007) and Varian (2007).
non-trivial choice between BIN and TAC. Applying this logic restores a generically unique equilibrium prediction.\textsuperscript{19} We can therefore solve for the unique trembling hand perfect equilibrium of each auction, and estimate the expected revenues from the average sample revenues at any parameter vector.

Finally, for additional robustness, we also perform a worst-case analysis over all rationalizable beliefs about rival strategies. From the point of view of revenue, this worst-case occurs when agents are least inclined to take the BIN option: specifically, when they believe that all other agents will choose to TAC and then bid zero. This implies that incentives to take the BIN option must be provided directly by the design, through the randomization parameter \( d \) and the reserve price \( r \) in the TAC auction. Since these beliefs are identical across all auctions, we can compute the indifference threshold \( \bar{v}(p, d, r) \) implied by these beliefs, and then calculate revenue in exactly the same way as in the incomplete information case.

4.3 Results

The results are in Tables 4 and 5. We find that the optimal reserve when running a second price auction is high: nearly twice as high as the second highest bid, but still quite a lot lower than the highest bid. By contrast BIN-TAC, regardless of the information structure, uses relatively low reserves (all well below one), and instead threatens to randomize among 3-4 bidders in order to get agents to take the high buy price (which is close in magnitude to the optimal SPA reserve). Interestingly, in the worst-case scenario the platform has to threaten randomization among 4 agents to get bidders to take BIN, since bidder beliefs are such that the TAC auction looks relatively attractive.

The welfare performance of these mechanisms is detailed in Table 5. The SPA without reserve earns revenue of 0.98 per auction, and leaves substantial consumer surplus — on average 1.97 per auction.\textsuperscript{20} Adding the large optimal reserve improves revenue slightly (to 1.03 per auction), but hurts consumer surplus substantially (it falls to 1.44).

BIN-TAC does better than both of these mechanisms in terms of revenue. Interestingly, the consumer surplus is higher than under the SPA with targeting and reserve, implying

\textsuperscript{19}Proof available from the authors on request.

\textsuperscript{20}The per auction revenue of 0.98 is lower than the average second highest bid of 1.07 in Table 2 because of a small fraction (2.3\%) of auctions with only a single bidder, which will realize zero revenue in an SPA without reserve.
BIN-TAC dominates the SPA in terms of both revenue and consumer surplus. This happens because the optimal SPA reserve price is very high — to extract revenue from the long right tail — and so many impressions are not sold, resulting in inefficiency and lower total welfare. By contrast, BIN-TAC has the BIN price to extract this revenue, and so the reserve is much lower, and more impressions are sold. Even accounting for distortions owing to the TAC auction, this is a welfare improvement.

By contrast, the bundling strategy underperforms. Revenues are much lower than the simple SPA, and consumer surplus falls even more dramatically. This is because there is considerable variation in match surplus across impressions even after conditioning on product and region, and so bundling along only these two dimensions destroys a lot of surplus.

Finally, the BIN-TAC results in the bottom part of the table show that the revenue estimates are relatively robust to how we model the information structure. However in models where the bidders are more informed, or dubious about the BIN option, consumer surplus is lower. In those cases the BIN decision is taken less often, thereby increasing the distortion from TAC auctions.

5 Conclusion and Future Work

We have introduced the BIN-TAC mechanism, particularly suited for environments where there is considerable asymmetry in valuations. This mechanism has been shown to outperform the second-price auction mechanism in this setting, and appears to be preferable to bundling goods together through withholding information, at least when there is a reasonable size population of potential bidders. Moreover, we demonstrated through an example that the mechanism can closely approximate Myerson’s optimal mechanism with ironing, despite its relative simplicity.

Our analysis of the exchange marketplace revealed that it has many features that make it a good place to apply our mechanism: large differences between the highest and second highest bid, and evidence of matching on user characteristics that the platform has chosen to make available to advertisers. Although the market does not fit our stylized model, we found that the BIN-TAC mechanism would nonetheless improve revenues and consumer surplus relative to the existing mechanism, a second price auction with reserve.

Due to data limitations we were not able to compare our mechanism to an optimal bundling
strategy. Instead, we looked at what would happen if the platform only provided advertisers with product and user location information, rather than more detailed demographics. This bundling strategy performed poorly, but it is an interesting and open research question as to whether switching mechanisms to BIN-TAC is in fact better than retaining the SPA with a more clever bundling strategy.

References


Appendix

Proof of Theorem 1

Let $a$ be a binary choice variable equal to 1 if the agent takes BIN and zero if TAC. Fix a player $i$, and fix arbitrary measurable BIN strategies $a_j(v)$ for the other players. Let $q$ be the probability that no other agent takes the BIN option, equal to $\prod_{j \neq i} \left( \int 1(a_j(v) = 0) dF(v) \right)$. Let $\pi(a, v)$ be the expected payoff to action $a$ for type $v$ given that the agent bids their valuation in any auction that follows. Then we have that $\frac{\partial}{\partial v} \pi(1,v) \geq q$, as a marginal increase in type increases the payoff by the probability of winning, which is lower bounded by $q$ when taking the BIN option. Similarly we have that $\frac{\partial}{\partial v} \pi(0,v) \leq \frac{q}{d}$, as the probability of winning when taking-a-chance is bounded above by $q/d$. Then $\pi(a, v)$ satisfies the strict single crossing property in $(a, v)$; it follows by Theorem 4 of Milgrom and Shannon (1994), the best response function must be strictly increasing in $v$, which in this case implies a threshold rule. It follows that any symmetric equilibrium must be in symmetric threshold
strategies. So fix an equilibrium of the form in the theorem, and let the payoffs to taking taking BIN be $\pi_B(v)$ and to TAC be $\pi_T(v)$. They are given by:

$$
\pi_B(v) = E \left[ 1(v > Y^1 > \bar{v})(v - Y^1) \right] + E \left[ 1(Y^1 < \bar{v})(v - p) \right] \\
\pi_T(v) = E \left[ 1(Y^1 < \bar{v})1(Y^* < v) \frac{1}{d}(v - Y^*) \right]
$$

The threshold type $\bar{v}$ must be indifferent, so

$$
\pi_B(\bar{v}) = E \left[ 1(Y^1 < \bar{v})(\bar{v} - p) \right] \quad (7) \\
= E \left[ 1(Y^1 < \bar{v})\frac{1}{d}(\bar{v} - Y^*) \right] = \pi_T(\bar{v}).
$$

Next, we show a $\bar{v}$ satisfying Eq. (1) exists and is unique. The right hand side of Eq. (1) is a function of $\bar{v}$ with first derivative $\frac{1}{d}(1 - \frac{2}{\bar{v}d}E[Y^*|Y^1 < \bar{v}]) < 1$. Since at $\bar{v} = 0$ it has value $p > 0$ and globally has slope less than 1, it must cross the 45° line exactly once. Thus there is exactly one solution to the implicit Eq. (1).

**Proof of Theorem 2**

By assumption, $\psi(v)$ single-crosses zero exactly once from below on $[\underline{w}, \bar{w}]$, so the implicit equation for $r^*$ has exactly one solution. We next show that this first order condition is necessary. Fix $d$ and $\bar{v} > p \geq r$ and define $p(r)$ implicitly as the BIN price that holds $\bar{v}$ constant as $r$ changes. Then there are two effects of increasing the reserve $r$ slightly: first, you can raise the BIN price without changing $\bar{v}$; second, if all bidders TAC, increasing the reserve raises the expected payment of some types, while decreasing the probability of sale. The marginal increase in revenue due to the first effect is:

$$
nF(\bar{v})^{n-1}(1 - F(\bar{v}))\frac{1}{d} \Pr(Y^d \leq r)
$$
With probability $F(\bar{v})^n$ there are no BIN bidders. Writing $F_\pi$ for $F(v|v < \bar{v})$:

$$F(\bar{v})^n \frac{1}{d} \sum_{k=1}^{d} \left[ \sum_{j=k}^{d} \binom{n}{j} (1 - F_\pi(r))^j F_\pi(r)^{n-j} \right. \left. + \int_{r}^{\bar{v}} \frac{n!}{d!(n-1-d)!} f_\pi(s) F_\pi(s)^{n-d-1} (1 - F_\pi(s))^d ds \right]$$

Taking a first order condition in $r$, canceling telescoping terms and simplifying:

$$F(\bar{v})^n \frac{1}{d} \sum_{k=1}^{d} \binom{n}{k} k (1 - F_\pi(r))^{k-1} F_\pi(r)^{n-k} (1 - F_\pi(r) - rf_\pi(r))$$

Summing both marginal effects and expanding $P(Y^d \leq r)$:

$$n(1 - F(\bar{v})) \left( \sum_{k=0}^{d-1} \binom{n-1}{k} (1 - F_\pi(r))^k F_\pi(r)^{n-1-k} \right) +$$

$$F(\bar{v}) \sum_{k=1}^{d} \binom{n}{k} k (1 - F_\pi(r))^{k-1} F_\pi(r)^{n-k} (1 - F_\pi(r) - rf_\pi(r))$$

Changing summation limits, factorizing, eliminating constants and setting the FOC = 0:

$$(1 - F(\bar{v})) + (1 - F_\pi(r) - rf_\pi(r)) F(\bar{v}) = 0$$

Now since $F_\pi = F(v|v < \bar{v}) = F(v)/F(\bar{v})$, we can simplify and solve to get $r^* = \frac{1 - F(r^*)}{f(r^*)}$.

Next, the optimal BIN price $p > r^*$ must be such that $\bar{v} \geq \omega_H$ (only high types take the BIN option). Let $p(\bar{v}, r)$ and $R(\bar{v}, r)$ be defined as in the text. There are three effects of a marginal increase in $\bar{v}$. First, the second highest bidder may have valuation $\bar{v}$ and choose not to take BIN, which decreases revenue by $\bar{v} - p(\bar{v}, r)$. The probability of $V^2 = \bar{v}$ is given by $n(n-1)f(\bar{v})(1 - F(\bar{v}))F(\bar{v})^{n-2}$. The second is that that highest bidder may have valuation $\bar{v}$ and choose not to take BIN, reducing revenue by $p(\bar{v}, r) - R(\bar{v}, r)$. This happens with probability $nf(\bar{v})F(\bar{v})^{n-1}$. Finally, the highest bidder may have valuation above $\bar{v}$ and the second highest below it, in which case this raises revenue by $\frac{\partial p(\bar{v}, r)}{\partial \bar{v}}$. This happens with probability $n(1 - F(\bar{v}))F(\bar{v})^{n-1}$. Setting the sum of these effects equal to zero, evaluating
the expression at \( r^* \) and eliminating common factors we get:

\[
f(v) \left( (n-1)(1-F(v))(v-p(v, r^*)) + F(v)(p(v, r^*) - R(v, r^*)) \right) = (1-F(v))F(v) \frac{\partial p(v, r^*)}{\partial v}
\]

**Proof of Theorem 3**

The optimal BIN-TAC mechanism in this case threatens to randomize among all bidders to induce a high BIN price, so \( d = n \), and has a reserve of \( v_L \). It follows that the optimal BIN price is \( v_H - \frac{(v_H-v_L)}{n} = \frac{n-1}{n} v_H + \frac{1}{n} v_L \). Then we can expand the revenue of BIN-TAC:

\[
R^{BIN-TAC} = (1-(1-\alpha)^n - n\alpha(1-\alpha)^{n-1}) v_H + n\alpha(1-\alpha)^{n-1} p^{BIN} + (1-\alpha)^n v_L
\]

\[
= (1-(1-\alpha)^n - n\alpha(1-\alpha)^{n-1}) v_H + n\alpha(1-\alpha)^{n-1} \left( \frac{n-1}{n} v_H + \frac{1}{n} v_L \right) + (1-\alpha)^n v_L
\]

\[
= (1-(1-\alpha)^n - \alpha(1-\alpha)^{n-1}) v_H + (\alpha(1-\alpha)^{n-1} + (1-\alpha)^n) v_L
\]

and similarly for the SPA without info:

\[
R^{SPA}_{no Info} = (\alpha v_H + (1-\alpha) v_L)^{(2:n)}
\]

\[
= (\alpha v_H + (1-\alpha) v_L) \text{ (degeneracy)}
\]

Since these are both probability distributions with two point support, to rank revenues it suffices to show that the mass on \( v_L \) is lower under BIN-TAC. So we must show that

\[
(\alpha(1-\alpha)^{n-1} + (1-\alpha)^n) < (1-\alpha).
\]

After a bit of simple algebra, this is equivalent to showing

\[
(1-\alpha) (1-\alpha(1-\alpha)^{n-2} - (1-\alpha)^{n-1}) \geq 0,
\]

which holds by binomial expansion of 1 with equality for \( n = 2 \) and strictly for \( n > 2 \). This proves the claim.

**Proof of Theorem 4**

Since the payment structure is well-known given the ironed virtual valuations, the challenge is to compute the ironed virtual values. We follow the approach proposed by Myerson (1981). This approach requires the distribution of values, \( F \), to be strictly increasing. Hence, we
consider the following distribution of the values.

$$f_\varepsilon(x) = \begin{cases} \beta f_L(x) & x \in [\omega_L, \omega_L] \\ \varepsilon & x \in (\omega_L, \omega_H) \\ f_H(x)\alpha & x \in [\omega_H, \omega_H] \end{cases}$$

$$F_\varepsilon(x) = \begin{cases} \beta F_L(x) & x \in [\omega_L, \omega_L] \\ \beta + \varepsilon(x - \omega_L) & x \in (\omega_L, \omega_H) \\ (1 - \alpha) + \alpha F_H(x - \omega_H) & x \in [\omega_H, \omega_H] \end{cases}$$

where $\beta + \varepsilon(\omega_H - \omega_L) + \alpha = 1$. As $\varepsilon$ tends to 0 we get the original model back. We need to “iron” the virtual values. For $q \in [0, 1]$, let $F_\varepsilon^{-1}(q)$ be the inverse of $F_\varepsilon(\cdot)$. Define:

$$h(q) = F_\varepsilon^{-1}(q) - \frac{1 - q}{f_\varepsilon(F_\varepsilon^{-1}(q))}$$

$$H(q) = \int_0^q h(y)dy$$

$$G(q) = \min_{\lambda, r_1, r_2 \in [0, 1], \lambda r_1 + (1 - \lambda) r_2 = q} \{\lambda H(r_1) + (1 - \lambda) H(r_2)\}$$

This implies that $G(\cdot)$ is the highest convex function on $[0, 1]$ such that $G(q) \leq H(q)$ for every $q$. Define $\phi(v) = G'(F(v))$ as the virtual value of type $v$. By Theorem 6.1 (Myerson 1981), the optimal mechanism randomly allocates the item to one of the bidders with the highest positive virtual value. We first show that the ironed virtual values are the same as the original virtual valuations, except for a set of quantiles between $q^*$ and $(1 - \alpha)$:

**Lemma 1** Let $q^* = (1 - \alpha)\bar{v}$ and $\bar{v}$ be the solution of

$$-F^2(\bar{v}) + (2 - \alpha)F(\bar{v}) + \alpha(\omega_H - \bar{v})f(\bar{v}) = 1 - \alpha.$$  

Under the assumption of Theorem 4, as $\varepsilon \to 0$,

$$G'(q) = \begin{cases} h(q) & q \in [0, q^*] \\ h(q^*) & q \in (q^*, 1 - \alpha) \\ h(q) & q \in [1 - \alpha, 1] \end{cases}$$
Proof  First note that \( H(q) \) is convex in \([0, \beta]\) because of the assumption that \( x - \frac{1-F(x)}{f(x)} \) is increasing in \([\omega_L, \omega_L]\). It is also decreasing in \([0, q_0]\) and increasing in \([q_0, \beta]\), where \( q_0 = F(r^*) \) is the minimum of \( H(\cdot) \) in this range. Also, observe that \( H(q) \) is decreasing in \([\beta, 1-\alpha]\) because \( h(q) < 0 \) in this interval. In addition, by the assumption that \( \psi(v) \) is increasing over the regions \([\omega_L, \omega_L]\) and \([\omega_H, \omega_H]\), \( H(q) \) is increasing and convex in \([1-\alpha, 1]\). Therefore, \( G(\cdot) \) includes the tangent line from the point \((1-\alpha, H(1-\alpha))\) to \( H(q) \) in \([0, \beta]\). Let \( q^* \) be the tangent point. We have

\[
G(q) = \begin{cases} 
H(q) & q \in [0, q^*] \\
(q-q^*)H(q^*)+(1-\alpha-q)H(1-\alpha) & q \in (q^*, 1-\alpha) \\
H(q) & q \in [1-\alpha, 1]
\end{cases}
\]

which immediately leads to the claim. \( \square \)

In the rest we compute \( q^* \). For \( q \in [0, \beta] \),

\[
H(q) = \int_0^q \left( F^{-1}_\varepsilon(y) - \frac{1-y}{f(x)(F^{-1}_\varepsilon(y))} \right) dy = \int_{\omega_L}^{F^{-1}_\varepsilon(q)} \left( x - \frac{1-F(x)}{f(x)} \right) f(x) dx \\
= \int_{\omega_L}^{F^{-1}_\varepsilon(q)} ((xf(x) + F(x)) - 1) dx \\
= (q - 1)^{-1} F^{-1}_\varepsilon(q) + \omega_L
\]

In particular,

\[
H(\beta) = (\beta - 1)\omega_L + \omega_L
\]

For \( q \in (\beta, 1-\alpha) \), because \( h(q) = \frac{2q-(1+\beta)}{\varepsilon} + \omega_L \), we get

\[
H(q) = H(\beta) + \left[ \frac{x^2 - (1 + \beta - \varepsilon \omega_L)x}{\varepsilon} \right]_{\beta}^q \\
= (\beta - 1)\omega_L + \omega_L + \frac{q^2 - \beta^2 - (q - \beta)(1 + \beta - \varepsilon \omega_L)}{\varepsilon} \\
= (q - 1)^{-1} \omega_L + \omega_L + (q - \beta)^{-1} \omega_L + (1-\alpha-\beta)^{-1} \omega_L \\
= (q - 1)^{-1} \omega_L + \omega_L + (q - \beta)^{-1} \omega_L
\]

\[
H(1-\alpha) = -\alpha \omega_L + \omega_L + \omega_L + (1-\alpha-\beta)^{-1} \omega_L = -\alpha \omega_H + \omega_L
\]
To iron the distribution, we compute the tangent from $H(1 - \alpha)$ to $H(q)$, for $q \in [0, 1 - \alpha]$. Note that if $q^*$ is the tangent point then

$$ h(q^*) = \frac{H(1 - \alpha) - H(q^*)}{1 - \alpha - q^*} \quad (10) $$

Observe that by Eq. (8) we have

$$ \frac{H(1 - \alpha) - H(q^*)}{1 - \alpha - q^*} = \frac{(-\alpha \omega_H + \omega_L) - ((q^* - 1)F^{-1}_\varepsilon(q^*) + \omega_L)}{1 - \alpha - q^*} $$

Let $\tilde{v} = F^{-1}_\varepsilon(q^*)$, i.e., $q^* = F_\varepsilon(\tilde{v}) = \beta F_L(\tilde{v})$. Therefore,

$$ \frac{H(1 - \alpha) - H(q^*)}{1 - \alpha - q^*} = \frac{-\alpha \omega_H - (q^* - 1)F^{-1}_\varepsilon(q^*)}{1 - \alpha - F_\varepsilon(\tilde{v})} $$

$$ h(q^*) = \tilde{v} - \frac{1 - F_\varepsilon(\tilde{v})}{f_\varepsilon(\tilde{v})} $$

As $\varepsilon \to 0$, the $F_\varepsilon(\cdot) \to F(\cdot)$. Plugging into Eq. (10) we get

$$ (\tilde{v}f(\tilde{v}) - 1 + F(\tilde{v}))(1 - \alpha - F(\tilde{v})) $$

$$ = f(\tilde{v})(-\alpha \omega_H - (F(\tilde{v}) - 1)\tilde{v}) $$

Hence, re-arranging the terms,

$$ -F^2(\tilde{v}) + (2 - \alpha)F(\tilde{v}) + \alpha(\omega_H - \tilde{v})f(\tilde{v}) = 1 - \alpha $$

Observe that only if $H(1 - \alpha) > H(q_0)$, then $h(q^*)$ is positive.

$$ -\alpha \omega_H + \omega_L \geq (q_0 - 1)F^{-1}(q_0) + \omega_L = (F(r^*) - 1)r^* + \omega_L $$

This is equivalent to $\alpha \omega_H \leq (1 - F(r^*))r^*$. If this fails, the optimal reserve $r^*$ is above the ironed region, and so a second price auction is optimal.
Finally, observe that

\[ h(q^*) \leq h(\beta) = \bar{\omega}_L \leq \omega_H - \frac{1 - F_\varepsilon(\omega_H)}{f_\varepsilon(\omega_H)} \]

which shows that \( G(\cdot) \) is convex and completes the proof.

Figure 4: **Relative Performance 1: Number of bidders.** Simulated expected revenues for different mechanisms as the number of bidders \( n \) varies, in an environment where \( F_L \) has mean 1 and standard deviation 0.5, the match probability is 0.05 and the match increment is 5.
Figure 5: **Relative Performance 2: Match Probability.** Simulated expected revenues for different mechanisms as the probability of a match $\alpha$ varies, where $F_L$ has mean 1 and standard deviation 0.5, the number of bidders is 10 and the match increment is 5.

Figure 6: **Relative Performance 3: Match Increment.** Simulated expected revenues for different mechanisms as the size of the gap between the low and high distributions $\Delta$ varies, where $F_L$ has mean 1 and standard deviation 0.5, the match probability is 0.05 and the number of bidders is 10.
Table 2: Summary Statistics: Microsoft Advertising Exchange Display Ad Auctions

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bid-Level Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average bid</td>
<td>1.000</td>
<td>0.565</td>
<td>2.507</td>
<td>0.0000157</td>
<td>130.7</td>
</tr>
<tr>
<td>Number of bids</td>
<td>508036</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Auction-Level Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winning bid</td>
<td>2.957</td>
<td>1.614</td>
<td>5.543</td>
<td>0.00144</td>
<td>130.7</td>
</tr>
<tr>
<td>Second highest bid</td>
<td>1.066</td>
<td>0.784</td>
<td>1.285</td>
<td>0.00132</td>
<td>39.22</td>
</tr>
<tr>
<td>Number of bidders</td>
<td>6.083</td>
<td>6</td>
<td>2.970</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Bid correlation</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of auctions</td>
<td>83515</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Advertiser-Level Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of auctions participated in (p1)</td>
<td>0.697</td>
<td>0.0251</td>
<td>4.641</td>
<td>0.00120</td>
<td>88.28</td>
</tr>
<tr>
<td>% of auctions won if participated (p2)</td>
<td>38.90</td>
<td>29.59</td>
<td>35.50</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Correlation of (p1,p2)</td>
<td>-0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary statistics for the full dataset, which is a 0.1 percent sample of a week’s worth of auction data sampled within the last two years. An observation is a bid in the top panel; an auction in the middle panel; and an advertiser in the last panel. Bids have been normalized so that their average is 1, for confidentiality reasons. The bid correlation is measured by selecting a pair of bids at random in every auction with at least two bidders, and computing the correlation coefficient.

Table 3: Matching on Region

<table>
<thead>
<tr>
<th></th>
<th>Participation</th>
<th>Bids</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advertiser Website Pageview Ratio</strong></td>
<td>0.029</td>
<td>0.329***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.015)</td>
</tr>
<tr>
<td><strong>Time-of-Day Fixed Effects</strong></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Product-Region Fixed Effects</strong></td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Advertiser Fixed Effects</strong></td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>5581749</td>
<td>5581749</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.02</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Results from OLS Regressions. In the first two columns, the dependent variable is a dummy for participation. The sample used in the regressions consists of all auction-bidder pairs, limited to the 10% of bidders who participate most often. In the last two columns, the dependent variable is the bid. The sample used in the regressions only includes bids from the 10% of bidders who bid most often. The independent variable is the population-weighted fraction of pageviews of the advertiser’s website that come from the region the user is in. Time-of-day fixed effects refer to a dummy for each quarter of the day, starting at midnight. Product-region fixed effects are dummies for the page-group advertised on, and the state the user is located in. Standard errors are robust. Significance levels are denoted by asterisks (* p < 0.1, ** p < 0.05, *** p < 0.01).
Table 4: Optimal Parameter Choices

<table>
<thead>
<tr>
<th>Policy</th>
<th>$p$</th>
<th>$d$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA</td>
<td></td>
<td></td>
<td>1.96</td>
</tr>
<tr>
<td>BIN-TAC (incomplete information)</td>
<td>2.60</td>
<td>3</td>
<td>0.43</td>
</tr>
<tr>
<td>BIN-TAC (complete information)</td>
<td>1.95</td>
<td>3</td>
<td>0.65</td>
</tr>
<tr>
<td>BIN-TAC (rationalizable worst case)</td>
<td>2.10</td>
<td>4</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Revenue-maximizing parameter choices. For each of the above mechanisms, we find these by maximizing the revenue functions defined in the main text over the available parameters numerically using a grid search.

Table 5: Counterfactual Revenues and Welfare

<table>
<thead>
<tr>
<th>Policy</th>
<th>Revenue</th>
<th>Consumer Surplus</th>
<th>Total Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA (no reserve)</td>
<td>0.983</td>
<td>1.974</td>
<td>2.957</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>SPA (optimal reserve)</td>
<td>1.028</td>
<td>1.471</td>
<td>2.499</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>BIN-TAC</td>
<td>1.075</td>
<td>1.633</td>
<td>2.708</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Bundling by Product-User Region</td>
<td>0.644</td>
<td>0.730</td>
<td>1.374</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.016)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Robustness to Informational Assumptions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIN-TAC (complete information)</td>
<td>1.072</td>
<td>1.589</td>
<td>2.661</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>BIN-TAC (rationalizable worst case)</td>
<td>1.066</td>
<td>1.530</td>
<td>2.596</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

Counterfactual simulations of average advertiser revenues, consumer surplus and total welfare (sum of producer and consumer surplus). All statistics reported outside parentheses are averages across impressions; those in parentheses are standard errors computed by bootstrapping the full dataset (i.e. they reflect uncertainty over the true DGP). Six different simulations are run. The first is of a second price auction without reserve, while the second is of a second price auction with optimal (revenue-maximizing) reserve. The third is of the BIN-TAC mechanism, under the incomplete information structure outlined in the text. The fourth is a no-targeting counterfactual where the impressions are bundled according to the product (i.e. URL and ad size) and user region, and sold by second-price auction. The last two are robustness checks, varying the informational assumptions made for BIN-TAC. In the complete information case, bidders know the valuations of the other participants, and made BIN decisions accordingly. In the rationalizable worst-case model, bidders assume they will only have to pay the reserve price in TAC auction, and therefore take the BIN option more rarely. Where applicable, the parameters used are the optimal parameters from Table 4.