Electronic Companion: Yield Optimization of Display Advertising with Ad Exchange

Santiago R. Balseiro
Fuqua School of Business, Duke University, 100 Fuqua Drive, NC 27708, srb43@duke.edu

Jon Feldman, Vahab Mirrokni, S. Muthukrishnan
Google Research, New York, NY 10011, jonfeld@google.com, mirrokni@google.com, muthu@google.com

Forthcoming in *Management Science*

EC.1. Extensions

In this section we consider a number of extensions of the model and policy.

EC.1.1. AdX with Multiple Bidders

Here we generalize our results to the case where multiple buyers participate in the Ad Exchange. We model AdX as an auction with K risk neutral buyers with individual valuations drawn independently from the same distribution with c.d.f $F(\cdot)$, density $f(\cdot)$, and support $[p_0, p_\infty]$ (to simplify the notation we drop the dependence on the user attributes). Moreover, we assume that the distribution of the values have increasing failure rates, are absolutely continuous and strictly monotonic.

Myerson (1981) argued that under our assumptions the optimal mechanism is a Vickrey or second-price sealed-bid auction. Moreover, it is known that in such auctions bidding the true valuation is a dominant strategy for the buyers, and that the optimal reservation price $p^*(c)$ is independent of the number of buyers (Laffont and Maskin 1980).

Let $B_{1:K}$ and $B_{2:K}$ be highest and the second highest bid, respectively. Given a reserve price p, the item is sold if $B_{1:K} \geq p$, i.e., there is some bid higher than the reserve price. The winning buyer pays the second highest bid, or alternatively $\max\{B_{2:K}, p\}$, since the seller should receive at least the reserve price p. Therefore, the publisher's maximization problem is

$$R(c) = \max_{p \geq 0} \mathbb{E}\left[\mathbf{1}\{B_{1:K} \geq p\} \max\{B_{2:K}, p\} + \mathbf{1}\{B_{1:K} < p\} c\right].$$

The setup of §2.2 can be consider as a particular case of a second-price auction in which we have only one bidder and $B_{2:K} = 0$.

As done previously, we cast our problem in terms of survival or winning probabilities. Letting s be the probability than the impression is sold, we have that $s = \mathbb{P}\{B_{1:K} \geq p\} = 1 - F^K(p)$ since valuations are i.i.d. Conversely, the reserve price as a function of the survival probability is given by $p(s) = \bar{F}^{-1}(1 - (1-s)^{1/K})$, which is well-defined due to the strict monotonicity of the c.d.f. In terms of survival probabilities, the problem is now

$$R(c) = \max_{0 \leq s \leq 1} r(s) + (1-s)c,$$

where we defined the revenue function as $r(s) = r(p(s))$, and $r(p) = \mathbb{E}\left[\mathbf{1}\{B_{1:K} \geq p\} \max\{B_{2:K}, p\}\right]$.

The next proposition shows that the revenue function is regular, and as a consequence all previous results hold for the case with multiple bidders.

Proposition EC.1. *Under the previous assumptions the revenue function $r(s)$ is regular. Moreover, the optimal reserve price $p^*(c)$ solves*

$$\frac{\bar{F}(p)}{f(p)} = p - c,$$

when $c \in [p_0 - 1/f(p_0), p_\infty]$. When the opportunity cost is higher than the null price ($c > p_\infty$), the publisher bypasses the exchange ($p^*(c) = p_\infty$). Finally, when the opportunity cost is low enough ($c < p_0 - 1/f(p_0)$), the impression is kept by the highest bidder ($p^*(c) = p_0$).

Proof of Proposition EC.1. The joint distribution of $B_{1:K}$ and $B_{2:K}$ has a density function (Laffont and Maskin 1980)

$$f(b_1, b_2) = \begin{cases} K(K-1)F(b_2)^{K-2}f(b_1)f(b_2) & \text{if } b_1 \geq b_2 \\ 0 & \text{otherwise} \end{cases}.$$

Then, we have that

$$\begin{aligned} r(p) &= \mathbb{E}\left[\mathbf{1}\{B_{2:K} \geq p\}B_{2:K} + p\mathbf{1}\{B_{1:K} \geq p, B_{2:K} < p\}\right] \\ &= \int_p^\infty \int_p^{b_1} b_2 f(b_1, b_2)\ \mathrm{d}b_2\ \mathrm{d}b_1 + p\int_p^\infty \int_0^p f(b_1, b_2)\ \mathrm{d}b_2\ \mathrm{d}b_1 \\ &= K(K-1)\int_p^\infty b_2 F(b_2)^{K-2} f(b_2)(1-F(b_2))\ \mathrm{d}b_2 + KpF(p)^{K-1}(1-F(p)) \end{aligned}$$

Continuity of $r(s)$ follows because the p.d.f. is continuous, and $p(s)$ is continuous (if F not strictly monotone, the inverse may have jumps). Additionally, we may bound the revenue by

$$r(p) \leq \mathbb{E}\left[\mathbf{1}\{B_{1:K} \geq p\}B_{1:K}\right] \leq K\mathbb{E}\left[\mathbf{1}\{B \geq p\}B\right] \leq K\mathbb{E}B < \infty,$$

the first inequality follows because $B_{1:K}$ is the maximum, the second because any order statistic is upper bounded by the sum of the bids, and the fourth because bids are integrable. Moreover, integrability of B implies that $\lim_{p\to\infty} r(p) = 0$.

Next, we turn to the concavity of $r(s)$. Differentiating w.r.t to p we get $\frac{dr}{dp} = KF(p)^{K-1}(\bar{F}(p) - pf(p))$. Then, using the fact that $\frac{ds}{dp} = -KF(p)^{1-k}/f(p)$ we get from the composition rule that $\frac{dr}{ds} = \left.\frac{dr}{dp}\right|_{p(s)} \frac{dp}{ds} = p(s) - \frac{1}{h(p(s))}$, where $h(p) = f(p)/\bar{F}(p)$ is the hazard rate of the bidder's valuation. Because $p(s)$ is non-increasing in s and the $h(p)$ is non-decreasing in p, we conclude that $\frac{dr}{ds}$ is non-increasing. Thus, the revenue function is concave.

Finally, notice that the that derivative of the objective w.r.t to s is

$$p(s) - \frac{1}{h(p(s))} - c, \tag{EC.1}$$

which is non-increasing. When $c > p_\infty$ we have that (EC.1) is negative, so $s^*(c) = 0$ and $p^*(c) = p_\infty$. Similarly, when $c < p_0 - 1/h(p_0)$ we that (EC.1) is positive, so $s^*(c) = 1$ and $p^*(c) = p_0$. □

EC.1.2. Covering Constraints

Guaranteed contracts typically specify a lump-sum amount in return for a fixed number of impressions and the publisher is not be monetarily rewarded for delivering impressions beyond these targets. In some settings, however, the publisher may seek to exceed these contractual targets in view of attracting feature business, at the expense of reducing the revenue from the exchange.

Our model is quite general and allows to easily accommodate *covering constraints*, that is, the case where the number of impressions assigned to each contract should be greater or equal to the capacity. In this case the capacity constraint of the DAP is relaxed to $\mathbb{E}\left[\sum_{n=1}^N i_{n,a}\right] \geq N\rho_a$, for all $a \in \mathcal{A}$. The analysis proceeds as before with the only difference that now the dual variables are non-negative, that is, the publisher should solve its dual problem under the constraint that $v_a \geq 0$. Additionally, when implementing the stochastic policy the publisher should now allow contracts to exceed their capacity. This amounts to determining the maximum contract-adjusted quality between all contracts $a \in \mathcal{A}$ (when the total number of impressions left is greater than the number of impressions necessary to fulfill the contracts), or equivalently changing Line 5 of Policy 2 to $a_n^* = \arg\max_{a \in \mathcal{A} \cup \{0\}} \{Q_{n,a} - v_a\}$. Regarding the performance the bid-price control μ^B, Theorem 2 still holds in this setting.

EC.1.3. Target Quality Constraints

Some publishers might feel more comfortable specifying target quality constraint than picking a Lagrange multiplier to weight the impact of quality in the objective. In other settings the advertisers themselves might demand that certain level of quality is guaranteed. In the following, we consider the case where the publisher strives to maximize the revenue from AdX, while complying with target quality constraints and capacity constraints.

The publisher imposes that the average quality of the impressions assigned to advertiser a is larger or equal than a threshold value ℓ_a. The DAP is similar, except that the objective only accounts for AdX's revenue, and for the inclusion of the constraints

$$\mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} i_{n,a} Q_{n,a}\right] \geq \ell_a, \qquad \forall a \in \mathcal{A}. \tag{EC.2}$$

Let $\gamma_a \geq 0$ be the Lagrange multiplier associated to (EC.2). Problem (3) can be interpreted as the Lagrange relaxation of our new problem w.r.t. the target quality constraints, and the dual variables γ_a as the shadow prices of the target quality constraints. The new constraints preserve the convexity of the primal program, and strong duality still holds. Following the same steps, we obtain the new dual problem

$$N \min_{\gamma \geq 0, v} \left\{ \mathbb{E}R\left(\max_{a \in \mathcal{A}_0}\{\gamma_a Q_a - v_a\}\right) + \sum_{a \in \mathcal{A}} v_a \rho_a - \gamma_a \ell_a \right\},$$

which still is a convex minimization problem. The publisher now jointly optimizes over v and γ to determine the bid-prices of the stochastic policy.

Regarding the performance of bid-price control, one can reproduce the steps of Theorem 2's proof to show that the policy asymptotically attains the optimal revenue from AdX, while complying with the delivery targets. Additionally, from the same asymptotic analysis one obtains that the expected average quality for contract a is lower bounded by $\left(1 - K/\sqrt{N}\right) \mathbb{E}\left[i_a^* Q_a\right]$. Hence, for advertisers with binding constraint (EC.2), albeit not feasible, the expected average quality becomes arbitrary close to the threshold value as the number of impressions in the horizon increases. For the remaining advertisers whose target quality constraint is not binding, the expected average quality will surpass the threshold for suitably large N.

EC.2. Additional Proofs

Given a subset of the quality space $D \subseteq \Omega$, we define the measure $\mathbb{P}_R(D)$ as the probability that the quality vector belongs to that subset and the impression is rejected by the Ad Exchange when the optimal survival probability is used. More formally,

$$\mathbb{P}_R(D) = \mathbb{E}\left[\left(1 - s^*\left(\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}; U\right)\right)\mathbf{1}\{Q \in D\}\right].$$

Notice that the latter is not a probability measure since $\mathbb{P}_R(\Omega) \le 1$. Proposition EC.2 characterizes the directional derivative of the objective function of the dual along some directions that, as we will show later, are of particular interest. Results are given in terms of the measure $\mathbb{P}_R$.

Proposition EC.2. *Given a subset $\alpha \in \mathcal{A}$, the directional derivative of the objective function of the dual w.r.t. directions $\mathbf{1}_\alpha$ and $-\mathbf{1}_\alpha$ are respectively*

$$\nabla_{\mathbf{1}_\alpha}\bar\psi(v) = -\mathbb{P}_R\left\{\max_{a\in\alpha}\{\gamma Q_a - v_a\} > \max_{a\in\mathcal{A}_0\setminus\alpha}\{\gamma Q_a - v_a\}\right\} + \sum_{a\in\alpha}\rho_a,$$
$$\nabla_{-\mathbf{1}_\alpha}\bar\psi(v) = \mathbb{P}_R\left\{\max_{a\in\alpha}\{\gamma Q_a - v_a\} \ge \max_{a\in\mathcal{A}_0\setminus\alpha}\{\gamma Q_a - v_a\}\right\} - \sum_{a\in\alpha}\rho_a,$$

where $\bar\psi(v) = \psi(v)/N$ is the normalized dual objective.

Proof of Proposition EC.2. We consider first the direction $\mathbf{1}_\alpha$. Notice that the random function $R\left(\max_{a\in\mathcal{A}_0}\{Q_a - v_a\}; U\right)$ is convex, and thus directionally differentiable. We first show that $\bar\psi(v)$ is finite. From Assumption 1 we have that the revenue function is bounded by $r(s;u) \le M$, and thus $R(c;u) \le M + \max(c,0) \le M + |c|$. Therefore, using the triangle inequality we obtain that

$$\bar\psi(v) \le M + \mathbb{E}|\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}| + \sum_{a\in\mathcal{A}}|v_a| \le M + \sum_{a\in\mathcal{A}}\gamma\mathbb{E}|Q_a| + 2|v_a| < \infty$$

We can now apply Theorem 7.46 in Shapiro et al. (2009) and obtain that $\bar\psi(v)$ is directionally differentiable at v and that one can exchange expectation and directional derivative. Putting all together we get that

$$\begin{aligned}\nabla_{\mathbf{1}_\alpha}\bar\psi(v) &= \mathbb{E}\left[\nabla_{\mathbf{1}_\alpha}R\left(\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}; U\right)\right] + \sum_{a\in\alpha}\rho_a \\ &= \mathbb{E}\left[\frac{\mathrm{d}R}{\mathrm{d}c}\left(\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}; U\right)\nabla_{\mathbf{1}_\alpha}\left\{\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}\right\}\right] + \sum_{a\in\alpha}\rho_a,\end{aligned}$$

where the second equation follows from the chain rule. We conclude by the fact that $\frac{\mathrm{d}R}{\mathrm{d}c}(c;u) = 1 - s^*(c;u)$ and $\nabla_{\mathbf{1}_\alpha}\left\{\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}\right\} = -\mathbf{1}\left\{\max_{a\in\alpha}\{\gamma Q_a - v_a\} > \max_{a\in\mathcal{A}_0\setminus\alpha}\{\gamma Q_a - v_a\}\right\}$. A similar result follows for the opposite direction $-\mathbf{1}_\alpha$ from the fact that $\nabla_{-\mathbf{1}_\alpha}\left\{\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\}\right\} = \mathbf{1}\left\{\max_{a\in\alpha}\{\gamma Q_a - v_a\} \ge \max_{a\in\mathcal{A}_0\setminus\alpha}\{\gamma Q_a - v_a\}\right\}$. $\square$

EC.3. A Representative Publisher

In this section we describe in detail the structure of Publisher 3. This medium-size publisher has 17 guaranteed contacts, 13 types, and it is moderately constrained with $\sum_{a\in\mathcal{A}}\rho_a = 43\%$.

The next figure shows the type-advertiser graph of the publisher. This bipartite graph has one node for each type T with a supply of $\pi(T)$ on the left side; one node for each advertisers $a\in\mathcal{A}$ with a demand ρ_a on the right side; and one arc joining user type T with advertisers a if and only if $a\in T$. In this publisher most advertisers have distinct targeting criteria, and few compete for the same inventory: on average each type can be assigned to 1.3 advertiser. The average number of advertisers per type is typically higher for the other publishers, with the highest equal to 6.6 in Publisher 6.

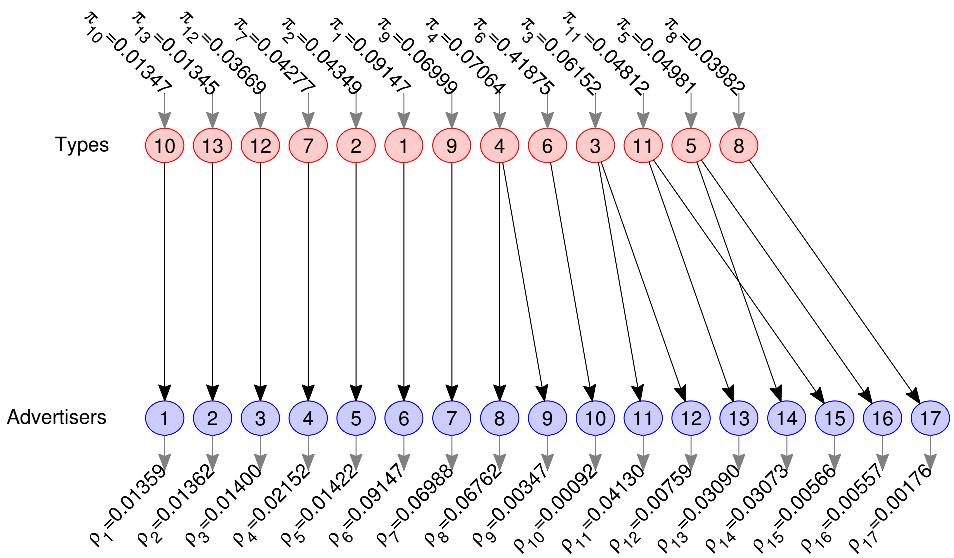


Figure EC.1 User type-advertiser graph for Publisher 3 (rotated 90 degrees clock-wise).

The next figure exhibits the estimated AdX's survival probability and revenue function for Publisher 3, which has an optimal acceptance probability of 74%.

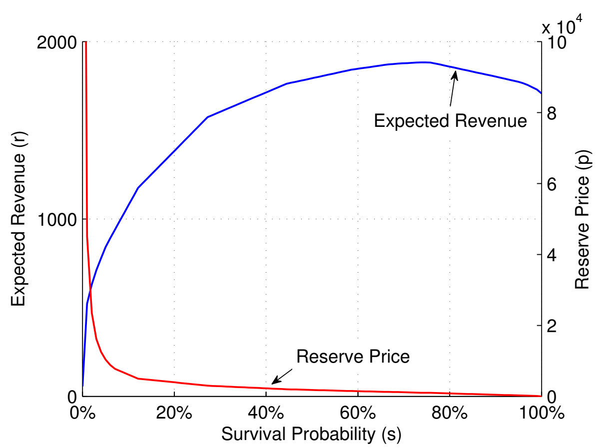
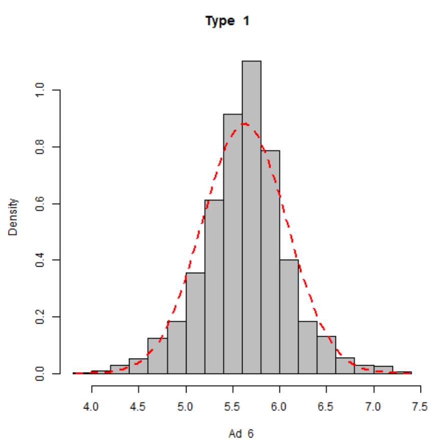
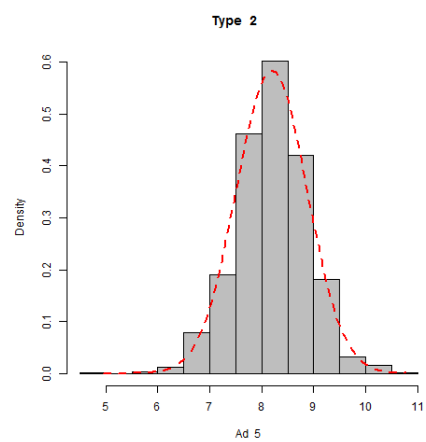


Figure EC.2 Estimated survival probability and revenue function from AdX for Publisher 3.

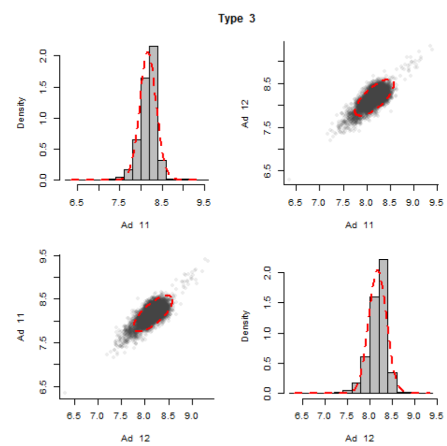
We conclude by showing the empirical distributions of log-quality for the 13 types in Publisher 3. Each type has its own subfigure with a matrix plot of the placement qualities of the advertisers that belong to the type. On the diagonals we show histograms of the empirical marginal distributions of log-quality together with the fitted normals distributions. The marginal log-quality of each advertiser approximately resembles a normal curve. On the off-diagonals, we show scatter plots of the correlation between advertisers, and the ellipse-like level curve for the fitted distributions at the confidence level of 90%. The scatter plots show that the correlation between advertisers is strongly positive, which may be a result of advertisers having similar targeting criteria.



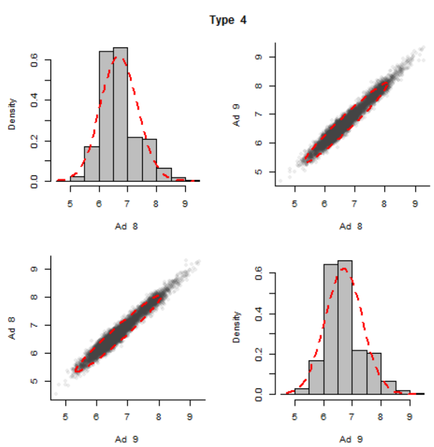
(a) Type 1



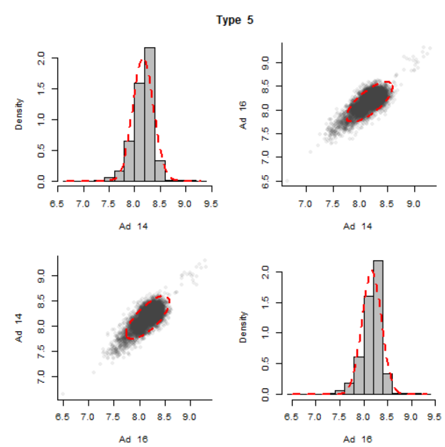
(b) Type 2



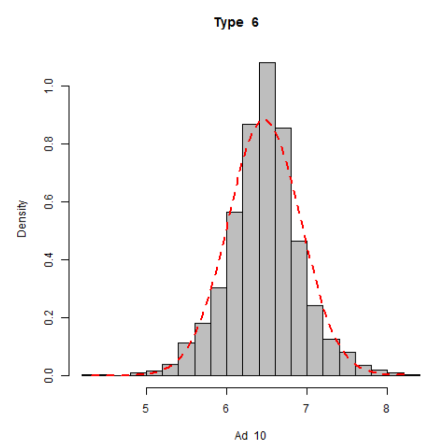
(c) Type 3



(d) Type 4

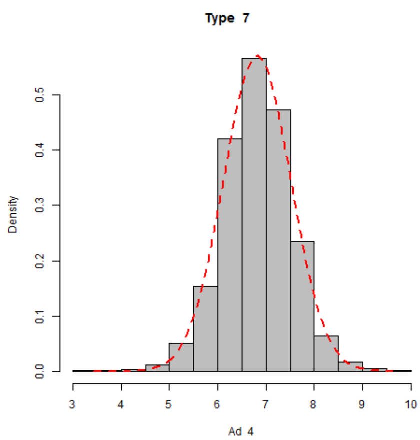


(e) Type 5

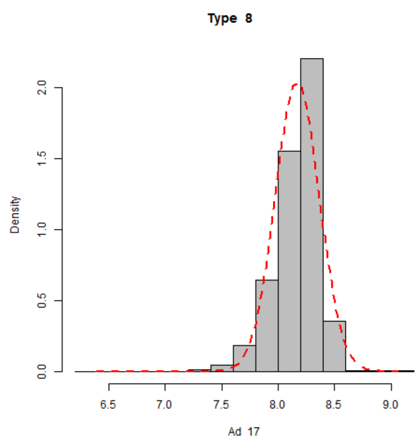


(f) Type 6

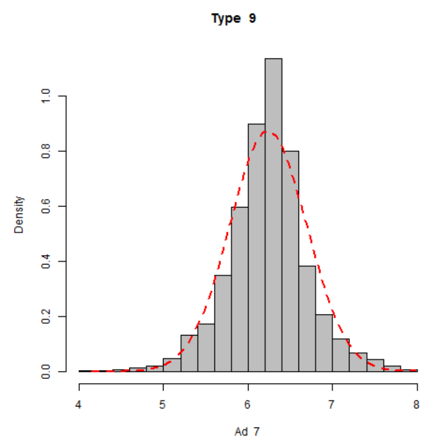
Figure EC.3 Empirical distribution of log-quality for types 1 to 6.



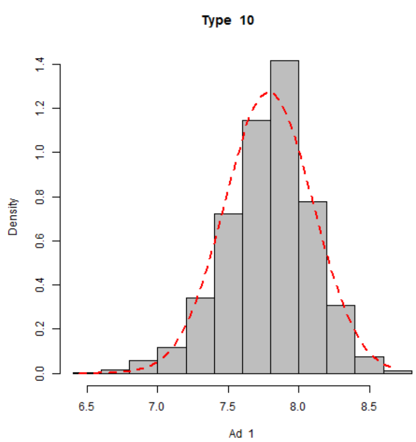
(a) Type 7



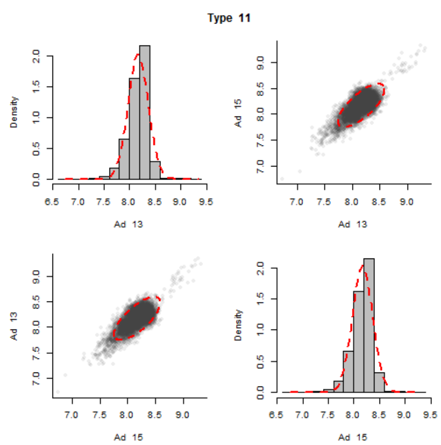
(b) Type 8



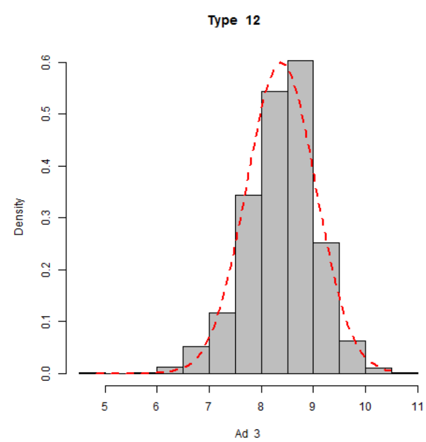
(c) Type 9



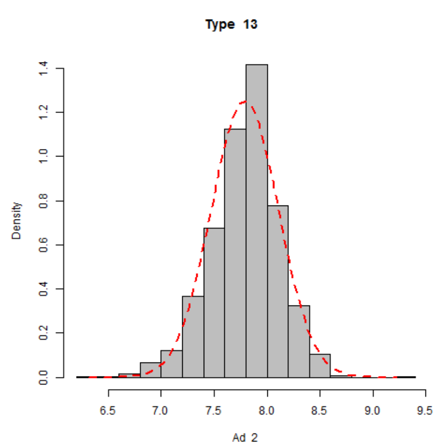
(d) Type 10



(e) Type 11



(f) Type 12



(g) Type 13

Figure EC.4 Empirical distribution of log-quality for types 7 to 13.

EC.4. Numerical Results

Publisher 1

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	622.1	622.1	621.8	619.8	618.2	612.9	604.4	592.5	569.1	546.3	526.2	499.1	494.1
	quality ($J_C^{D(\gamma)}$)	673.0	673.1	720.9	804.0	825.7	857.5	881.4	898.3	913.1	919.7	922.6	923.9	923.9
	optimality gap	0.016%	0.000%	0.000%	0.001%	0.001%	0.002%	0.003%	0.005%	0.006%	0.005%	0.007%	0.019%	0.023%
Policy	revenue ($J_A^{B(\gamma)}$)	621.8	621.8	621.3	618.4	616.6	611.3	602.8	591.6	568.4	546.1	526.2	499.1	494.1
	quality ($J_C^{B(\gamma)}$)	669.6	669.7	716.2	797.0	818.8	849.2	871.5	888.4	902.7	908.8	912.1	913.6	913.8

Publisher 2

	γ	0	0.001	0.01	0.1	0.25	0.5	1	2.5	5	10	25	50	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	163.0	163.1	163.0	163.1	163.1	163.0	162.7	161.7	159.9	155.2	139.8	117.0	89.8	49.2
	quality ($J_C^{D(\gamma)}$)	60.3	60.5	60.5	60.7	60.8	60.9	61.3	62.0	62.5	63.1	64.0	64.7	65.0	65.3
	optimality gap	0.045%	0.012%	0.018%	0.000%	0.000%	0.000%	0.000%	0.008%	0.003%	0.010%	0.013%	0.017%	0.053%	0.019%
Policy	revenue ($J_A^{B(\gamma)}$)	162.7	163.1	163.1	163.1	163.0	163.0	162.7	161.6	159.7	155.0	139.5	116.8	89.8	49.2
	quality ($J_C^{B(\gamma)}$)	60.0	60.2	60.2	60.4	60.5	60.6	61.0	61.6	62.1	62.8	63.7	64.3	64.7	64.9

Publisher 3

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	1826.9	1830.0	1830.0	1829.9	1823.8	1806.4	1790.8	1755.1	1662.9	1562.6	1448.8	1186.3	1066.0
	quality ($J_C^{D(\gamma)}$)	551.7	620.4	620.4	623.5	649.3	779.6	826.3	880.2	934.8	964.3	981.9	994.0	994.5
	optimality gap	0.225%	0.015%	0.015%	0.020%	0.282%	0.071%	0.063%	0.098%	0.218%	0.166%	0.060%	0.023%	0.030%
Policy	revenue ($J_A^{B(\gamma)}$)	1826.8	1828.8	1828.8	1828.7	1824.3	1802.9	1788.6	1750.8	1660.4	1558.5	1445.2	1184.4	1066.0
	quality ($J_C^{B(\gamma)}$)	541.5	606.9	606.9	610.3	636.2	764.9	810.1	863.8	916.3	945.1	961.5	973.9	974.6

Publisher 4

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	1319.4	1319.3	1318.3	1317.7	1315.2	1302.7	1269.4	1216.8	1129.1	1072.6	1028.8	974.2	954.9
	quality ($J_C^{D(\gamma)}$)	331.1	461.8	514.1	535.9	567.3	645.7	735.0	808.1	868.6	884.9	892.5	895.2	895.3
	optimality gap	0.125%	0.057%	0.128%	0.125%	0.149%	0.103%	0.170%	0.203%	0.080%	0.116%	0.025%	0.033%	0.029%
Policy	revenue ($J_A^{B(\gamma)}$)	1319.7	1319.6	1318.2	1317.3	1315.0	1302.1	1268.4	1215.6	1128.2	1070.8	1027.9	973.8	954.9
	quality ($J_C^{B(\gamma)}$)	328.7	458.9	510.4	531.2	563.2	635.4	726.1	800.6	856.5	873.0	879.1	881.5	881.4

Publisher 5

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	828.7	837.2	837.3	836.3	833.3	822.9	805.9	771.8	702.1	633.2	562.3	428.0	383.1
	quality ($J_C^{D(\gamma)}$)	1526.5	1539.3	1550.2	1564.2	1611.0	1677.4	1723.0	1770.5	1815.2	1835.2	1845.6	1849.9	1853.5
	optimality gap	0.428%	0.045%	0.035%	0.104%	0.081%	0.059%	0.087%	0.104%	0.115%	0.092%	0.075%	0.213%	0.028%
Policy	revenue ($J_A^{B(\gamma)}$)	831.0	837.5	837.5	836.4	833.1	822.6	805.0	770.3	700.4	631.6	560.7	427.0	383.1
	quality ($J_C^{B(\gamma)}$)	1527.1	1539.4	1551.7	1563.8	1609.6	1672.4	1718.4	1766.4	1810.9	1830.1	1840.2	1843.5	1847.7

Publisher 6

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	2050.7	2052.8	2052.9	2052.3	2050.9	2032.3	1976.6	1907.5	1805.8	1727.2	1658.5	1539.7	1492.4
	quality ($J_C^{D(\gamma)}$)	525.1	678.6	679.2	696.6	716.3	797.8	956.7	1056.4	1120.8	1142.8	1154.5	1161.4	1161.4
	optimality gap	0.150%	0.039%	0.035%	0.044%	0.049%	0.171%	0.152%	0.196%	0.329%	0.454%	0.383%	0.344%	0.357%
Policy	revenue ($J_A^{B(\gamma)}$)	2050.7	2052.8	2052.7	2052.4	2050.9	2032.3	1976.3	1907.2	1805.3	1726.8	1658.0	1539.1	1492.4
	quality ($J_C^{B(\gamma)}$)	525.8	678.8	679.8	699.6	721.6	803.3	959.2	1054.2	1118.9	1140.3	1149.5	1153.7	1153.8

Publisher 7

	γ	0	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100	Inf
DAP	revenue ($J_A^{D(\gamma)}$)	1376.2	1376.1	1376.6	1374.5	1373.7	1371.3	1366.8	1357.4	1330.3	1301.0	1269.9	1194.6	1155.5
	quality ($J_C^{D(\gamma)}$)	122.5	126.5	147.0	192.4	207.2	219.1	232.9	247.2	263.0	272.6	277.3	281.6	281.8
	optimality gap	0.199%	0.161%	0.116%	0.188%	0.171%	0.224%	0.223%	0.215%	0.524%	0.504%	0.563%	0.424%	0.419%
Policy	revenue ($J_A^{B(\gamma)}$)	1376.5	1376.5	1376.2	1375.1	1374.2	1371.6	1367.0	1356.2	1328.9	1298.9	1268.0	1193.2	1155.5
	quality ($J_C^{B(\gamma)}$)	123.5	127.4	148.1	194.8	209.0	221.6	234.1	248.0	262.9	270.0	275.0	276.9	276.6

Table EC.1 AdX's revenue, contracts' quality, and optimality for the DAP, together with the simulated AdX's revenue and contracts' quality from the policy μ^B for 7 publishers, and different choices of γ.

Publisher 1

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	622.4	628.4	658.3	698.5	823.6	1038.5	1480.0	2825.3	5090.2	9647.0	91856.7
Greedy Pol.	yield	622.6	627.1	645.2	670.6	740.4	860.0	1117.4	1888.4	3223.0	5934.9	54860.9
	gap	0.00%	-0.21%	-1.99%	-3.99%	-10.10%	-17.19%	-24.50%	-33.16%	-36.68%	-38.48%	-40.28%
Static Price Pol.	yield	622.4	628.5	655.3	689.2	796.2	994.0	1427.9	2784.7	5063.4	9633.8	91873.2
	gap	-0.00%	0.00%	-0.46%	-1.32%	-3.33%	-4.29%	-3.52%	-1.44%	-0.53%	-0.14%	0.00%

Publisher 2

	γ	0.001	0.01	0.1	0.25	0.5	1	2.5	5	10	25	50	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	163.1	163.7	169.1	178.2	193.3	223.7	315.6	470.2	782.6	1731.6	3334.0	6563.6
Greedy Pol.	yield	51.0	51.5	56.4	64.8	78.8	106.9	192.2	338.0	631.9	1495.2	2830.1	5519.0
	gap	-68.72%	-68.53%	-66.64%	-63.63%	-59.25%	-52.22%	-39.08%	-28.12%	-19.26%	-13.65%	-15.12%	-15.91%
Static Price Pol.	yield	51.0	51.6	57.5	67.2	83.4	115.9	213.2	375.5	700.1	1673.7	3296.5	6542.1
	gap	-68.71%	-68.46%	-66.02%	-62.28%	-56.84%	-48.19%	-32.43%	-20.14%	-10.54%	-3.34%	-1.13%	-0.33%

Publisher 3

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	1829.4	1834.9	1859.2	1887.9	1994.2	2193.7	2614.6	3951.1	6283.9	11060.4	98570.7
Greedy Pol.	yield	1443.1	1446.6	1485.4	1536.7	1706.8	1896.2	2228.1	3186.2	4727.9	7820.8	64573.5
	gap	-21.12%	-21.16%	-20.11%	-18.60%	-14.41%	-13.56%	-14.78%	-19.36%	-24.76%	-29.29%	-34.49%
Static Price Pol.	yield	1459.8	1468.0	1504.6	1550.4	1687.8	1916.8	2374.7	3747.4	6057.4	10857.0	98527.3
	gap	-20.21%	-19.99%	-19.07%	-17.88%	-15.36%	-12.62%	-9.18%	-5.15%	-3.60%	-1.84%	-0.04%

Publisher 4

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	1320.1	1323.3	1343.9	1371.3	1461.0	1631.5	2016.2	3269.4	5435.9	9819.3	89121.3
Greedy Pol.	yield	1295.9	1323.5	1337.6	1355.6	1410.5	1503.7	1699.7	2325.5	3424.8	5688.9	46954.2
	gap	-1.83%	0.00%	-0.47%	-1.14%	-3.45%	-7.83%	-15.70%	-28.87%	-37.00%	-42.06%	-47.31%
Static Price Pol.	yield	1262.8	1267.8	1290.3	1318.3	1402.1	1544.6	1914.5	3191.1	5383.2	9787.5	89095.3
	gap	-4.34%	-4.19%	-3.99%	-3.86%	-4.03%	-5.32%	-5.05%	-2.39%	-0.97%	-0.32%	-0.03%

Publisher 5

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	839.1	853.0	914.6	994.1	1240.7	1664.2	2536.6	5227.6	9782.0	18963.0	184774.0
Greedy Pol.	yield	400.5	412.0	469.5	543.7	759.1	1093.4	1719.5	3581.6	6679.0	12915.2	124740.4
	gap	-52.27%	-51.70%	-48.66%	-45.30%	-38.82%	-34.30%	-32.21%	-31.49%	-31.72%	-31.89%	-32.49%
Static Price Pol.	yield	401.3	417.9	491.8	584.2	861.4	1323.3	2247.1	5018.7	9637.9	18876.5	185172.1
	gap	-52.18%	-51.01%	-46.23%	-41.23%	-30.57%	-20.48%	-11.41%	-4.00%	-1.47%	-0.46%	0.00%

Publisher 6

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	2053.4	2059.5	2087.3	2123.0	2233.1	2455.9	2961.4	4602.6	7428.2	13153.5	116911.0
Greedy Pol.	yield	1965.1	1968.0	1983.2	2011.5	2133.6	2270.9	2497.0	3220.8	4475.2	7019.2	53457.0
	gap	-4.30%	-4.44%	-4.99%	-5.25%	-4.45%	-7.53%	-15.68%	-30.02%	-39.75%	-46.64%	-54.28%
Static Price Pol.	yield	1997.9	2005.1	2037.2	2077.2	2197.2	2397.1	2877.8	4518.2	7352.5	13084.8	116896.0
	gap	-2.71%	-2.64%	-2.40%	-2.16%	-1.61%	-2.39%	-2.82%	-1.83%	-1.02%	-0.52%	-0.01%

Publisher 7

	γ	0.001	0.01	0.05	0.1	0.25	0.5	1	2.5	5	10	100
Bid-price Pol.	yield ($J^{B(\gamma)}$)	1376.6	1377.7	1384.9	1395.1	1427.0	1484.0	1604.2	1986.1	2649.1	4018.3	28887.1
Greedy Pol.	yield	1378.3	1378.8	1381.2	1384.7	1395.7	1414.4	1451.5	1559.3	1736.7	2084.0	8478.1
	gap	0.00%	0.00%	-0.26%	-0.75%	-2.19%	-4.69%	-9.52%	-21.49%	-34.44%	-48.14%	-70.65%
Static Price Pol.	yield	1376.6	1377.7	1383.8	1393.8	1425.1	1479.3	1589.3	1956.9	2614.5	3977.5	28893.8
	gap	-0.00%	0.00%	-0.08%	-0.10%	-0.13%	-0.31%	-0.93%	-1.47%	-1.31%	-1.02%	0.00%

Table EC.2 Comparison of the yield of the optimal policy with the yield of the Greedy and Static Price policy for 7 publishers and different choices of γ.

EC.5. Non-parametric Learning Policy

In this section we present a policy that learns in a non-parametric fashion the parameters of the modified bid-price policy with tie-breaking. Similar to the work of Devanur and Hayes (2009), Feldman et al. (2010), Vee et al. (2010), Agrawal et al. (2009) our policy determines a dual variable for each contract and then assigns the impressions to the advertiser with maximum contract adjusted quality. The policy presented here, however, extends their work in two directions. First, previous work assumes that inputs are in general position so that ties can be effectively discarded. We have seen that ties contribute to a sizable fraction of the yield, so these need to be carefully dealt with. Our new policy explicitly handles ties by implementing the novel tie-breaking rule introduced in Section 4.3, and determines tie-breaking probabilities by solving a feasible flow problem. Additionally, we present a new theoretical analysis on the expected performance of the policy by borrowing tools from statistical learning theory and introduce a bound that accounts for the number of potential ties in the problem. Second, our policy accounts for the pricing dimension of the problem by dynamically determining a reserve price for the AdX auction based on the opportunity cost of not assigning the impression to a contract.

A limitation of our policy is that it assumes that the revenue function and distribution of the highest bid in the exchange are known in advance. The pricing dimension of the problem introduces some complexities that are beyond the scope of this paper. The first issue is the observability of the rewards in the pricing problem. The literature on prior-free allocation problems for online advertising assumes that rewards are observed –before– the decision is made. However, a distinct feature of the dynamic pricing problem is that rewards are observed –after– the decision is made since the publisher observes the result of the auction after posting the reserve price, which is binding. As a result, a prior-free method would need to carefully explore the price space in the AdX problem and the usual exploration vs. exploitation trade-off of the learning literature for pricing applies here (see, e.g, Broder and Rusmevichientong (2012) and Besbes and Zeevi (2009)). Note that this trade-off is absent in the allocation problem. The second issue is that a prior-free approach for the AdX pricing problem needs to estimate the whole demand curve. In the standard dynamic pricing problem the firm aims to estimate a single parameter: the revenue maximizing price. In our problem, however, the optimal reserve price for AdX should take into account the opportunity cost of assigning the impression to an advertiser (as given by $\max_{a\in\mathcal{A}_0}\{\gamma q_{n,a} - v_a\}$), which is potentially different in each auction. As a result the publisher needs to estimate the whole pricing curve to price efficiently in the presence of the guaranteed contracts, which calls for more involved methods.

EC.5.1. The Learning Policy

The policy has a *learning phase* followed by a *implementation phase*. In the first phase an i.i.d. sample of M vectors of user attributes is used to learn the parameters of the policy. In the second phase the obtained policy is implemented in an independent horizon of N impressions, and the impressions are delivered to the contracts.

Learning Phase. In the learning phase an i.i.d. sample of M vectors of user attributes $\{u_m\}_{m=1}^M$ is employed to learn the parameters of the modified bid-price policy with tie-breaking, which is given by a pair $(\hat{v},\hat{p})$ with $\hat{v}\in\mathbb{R}^A$ the vector of dual variables and $\hat{p}:2^{\mathcal{A}_0}\to[0,1]^{A+1}$ the tie-breaking rule. The dual variables are obtained by solving the Sample Average Approximation version of the dual

$$\min_v \frac{1}{M}\sum_{m=1}^M R\Big(\max_{a\in\mathcal{A}_0}\{\gamma q_{m,a}-v_a\};u_n\Big)+\sum_{a\in\mathcal{A}}\rho_a v_a, \tag{EC.3}$$

which is a non-differentiable convex minimization problem.

While one could potentially determine the tie-breaking probabilities for all possible ties, in practice most ties do not occur with positive probability and can be effectively discarded. Our policy exploits prior knowledge about the ties that have positive probability with respect to the underlying unknown probability measure, by determining the tie-breaking probabilities only for these. Let $\mathcal{S}_a=\big\{S\subseteq\mathcal{A}_0: a\in S \text{ and } \mathbb{P}\big\{S=\arg\max_{a\in\mathcal{A}_0}\{\gamma Q_a-v_a\}\big\}>0 \text{ for some } v\in\mathbb{R}^A\big\}$ be the set of potential ties involving contract $a\in\mathcal{A}_0$ that occur with positive probability for some vector of dual variables. Similarly, let $\mathcal{S}_a^c=\{S\subseteq\mathcal{A}_0: a\in S, S\notin\mathcal{S}_a\}$ be the subsets containing a for which the probability of a tie occurring is zero. The tie-breaking rule is obtained by solving an assignment problem over sets in $\mathcal{S}_a$

$$\sum_{S\in\mathcal{S}_a}\hat{\mathbb{P}}(S\text{-tie})p_a(S)\le\rho_a,\qquad \forall a\in\mathcal{A}, \tag{EC.4a}$$

$$\sum_{S\in\mathcal{S}_a}\hat{\mathbb{P}}(S\text{-tie})p_a(S)\ge\rho_a-\sum_{S\in S_a^c}\hat{\mathbb{P}}(S\text{-tie}),\qquad \forall a\in\mathcal{A}, \tag{EC.4b}$$

$$\sum_{a\in S}p_a(S)=1,\qquad \forall S\in\cup_{a'\in\mathcal{A}}\mathcal{S}_{a'}, \tag{EC.4c}$$

$$p_a(S)\ge 0,\qquad \forall S\in\cup_{a'\in\mathcal{A}}\mathcal{S}_{a'}, a\in S,$$

$$p_a(S)=0,\qquad \forall S\notin\cup_{a'\in\mathcal{A}}\mathcal{S}_{a'}, a\in S,$$

where we define $\hat{\mathbb{P}}(S\text{-tie})=\frac{1}{M}\sum_{m=1}^M\big(1-s^*(\max_{a\in\mathcal{A}_0}\{\gamma q_{a,m}-\hat{v}_a\};u_m)\big)\mathbf{1}\big\{S=\arg\max_{a\in\mathcal{A}_0}\{\gamma q_{a,m}-\hat{v}_a\}\big\}\big)$ as the empirical probability that the impression is rejected by AdX and the maximum is verified by contracts in the set S. Because the tie-breaking policy is restricted to the sets with positive probability, the fraction of impressions assigned to contract a could be different than ρ_a even over the sample $\{u_m\}_{m=1}^M$. Equations (EC.4a) and (EC.4b) guarantee that contract a is assigned a fraction of impressions "close" to ρ_a. Equation (5b) guarantees that for each tie S the probabilities sum up to one.

Implementation Phase. In the implementation phase the policy prices in the exchange according to the acceptance probability $s^{(\hat{v},\hat{p})}(u)=s^*(\max_{a\in\mathcal{A}_0}\{\gamma q_a-\hat{v}_a\};u)$, and, if rejected, assigns the impression to contract a with probability $I_a^{(\hat{v},\hat{p})}(u)=\sum_{S\subseteq\mathcal{A}_0:a\in S}\hat{p}_a(S)\mathbf{1}\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma q_{a'}-\hat{v}_{a'}\}\}$. The corrections of Policy 2 are implemented to guaranteed that contracts are satisfies almost surely. Policy 1 describes the algorithm in detail.

Policy 1 Learning Policy μ^L.

```
Observe samples {u_m}_{m=1}^M.
Solve problem (EC.3) to determine dual variables v̂.
Solve problem (EC.4) to determine tie-breaking rule p̂.
for n = 1, ..., N do
    Observe state (n, X), and the vector of attributes u_n.
    Determine the vector of placement qualities q_n.
    Let A_n = {a ∈ A : x_{n,a} > 0} be the set of ads yet to be satisfied.
    if Σ_{a∈A} x_{n,a} < n then
        Let S_n = arg max_{a∈A_n∪{0}} {γ q_{n,a} − v̂_a}
        Submit to AdX with price p*(max_{a∈A_n∪{0}} {γ q_{n,a} − v̂_a} ; u_n).
        if impression rejected by AdX then assign to advertiser a ∈ S_n with probability p̂_a(S_n).
    else
        Let S_n = arg max_{a∈A_n} {γ q_{n,a} − v̂_a}
        Assign to advertiser a with probability p̂_a(S_n).
    end if
end for
```

EC.5.2. Analysis

In this section we analyze the expected performance of the learning policy described in Policy 1. The main result extends the asymptotic analysis of §4.4 to take into account the learning of the algorithm. We prove the main result under the following assumption on the primitives.

ASSUMPTION EC.1. *The exchange's bids lie in $[0, \bar{B}]$ with $\bar{B} < \infty$; placement qualities lie in $[\underline{Q}, \bar{Q}]$ with $\infty < \underline{Q} \le 0 < \bar{Q} < \infty$; and the minimum capacity-to-impression ratio $\underline{\rho} = \min_{a \in \mathcal{A}_0} \rho_a$ is strictly positive, that is, $\underline{\rho} > 0$.*

The previous assumption imposes that all the primitives are bounded both from below and above, which is reasonable for most settings. Let $\|\mathcal{S}\|_1 = \sum_{a \in \mathcal{A}} |\mathcal{S}_a|$ be the total number of potential ties. Our main result bounds the expected regret of the learning policy w.r.t. the optimal online policy that has full knowledge of the probabilistic distributions of the primitives.

THEOREM EC.1. *Suppose that Assumptions 1 and EC.1 hold. Let J^L the expected performance of the learning policy with a training set of $M > 2$ samples on an independent horizon of length N. The expected regret of the learning policy w.r.t. the optimal online policy is bounded by*

$$\frac{J^* - J^L}{N} \le O\left(\frac{1}{\sqrt{N}} + \|\mathcal{S}\|_1 \sqrt{\frac{A \log(M)}{M}} \right),$$

whenever $N \ge K^2$ and $M > A$.

The first term of the bound of Theorem EC.1 controls for the error introduced by the stationary nature of the policy during the implementation phase of the campaigns. This term depends exclusively on the number of impressions of the horizon, and is of order $\sqrt{1/N}$ as in Theorem 2. The second term of the bound controls for the error introduced during the learning phase and its impact on the delivery of the campaigns. This term depends on the number of samples M and advertisers A and is of order $\sqrt{A\log(M)/M}$.

The learning error term also depends linearly on the number of ties $\|\mathcal{S}\|_1$, which is upper bounded by $A2^A$. The latter bound is conservative and in most applications one can obtain better bounds by exploiting the known structure of the distribution of placement qualities. For example, when placement qualities are absolutely continuous all ties are singletons, and thus $\|\mathcal{S}\|_1 = A$. In the user type model presented in §5 the number of ties is at most $\|\mathcal{S}\|_1 \le TA$, where T is the number of types.

EC.5.3. Proof of Main Result

We prove the result in five steps. First, we show that the dual variables can be restricted to a compact domain. Second, we prove that the problem (EC.4) admits a feasible tie-breaking rule when the optimal dual variables of (EC.3) are used. Third, we quantify the impact of the learning error on the expected performance of the algorithm during the implementation phase. In the last two steps we bound the expected error of the learning phase in terms of the number of samples in the training set.

Lemma EC.1. *Suppose that EC.1 holds. Then the optimal dual variables of problem* (EC.3) *lie in the compact set* $[-V,V]^A$ *with* $V \triangleq (\bar{B} + \gamma A\bar{Q} - \gamma \underline{Q})/\underline{\rho}$.

For the following result we first show that if we consider every possible tie in problem (EC.4), then the resulting assignment problem admits a feasible solution. The latter follows from repeating the steps of Proposition 2's proof with the empirical measure $\hat{\mathbb{P}}$ instead of unknown probability measure $\mathbb{P}$, and using the SAA version of the dual (EC.3) instead of (4). Then we proceed by constructing a feasible solution to our problem by setting to zero the tie-breaking probabilities associated to ties with zero probability.

Lemma EC.2. *The assignment problem EC.4 admits a feasible tie-breaking rule.*

Let $\xi_a^{(v,p)} = \mathbb{E}\big[(1 - s^{(v,p)}(U))I_a^{(v,p)}(U)\big] - \rho_a$ be the error incurred in the expected fraction of impressions assigned to contract $a \in \mathcal{A}$ by a bid-price control that employs dual variables v and the tie-breaking rule p, where the expectation is taken over the unknown distribution of user's attributes. The second result shows that the expected performance of any policy can be lower bounded in terms of the error of the policy.

Proposition EC.3. *Suppose that Assumptions 1 and EC.1 hold. Let $J^{(v,p)}$ be the expected performance in the implementation phase of a stochastic policy with dual variables v and the tie-breaking rule p in an horizon of length N. The performance is lower bounded by*

$$J^{(v,p)} \ge J^D - V\underline{\rho}K\sqrt{N} - (N+1)V\|\xi^{(v,p)}\|_1,$$

whenever $N \geq K^2$.

In the last two steps we bound the expected error of the learning phase by decomposing it into an *estimation error* and an *empirical assignment error*. The empirical assignment error is given by $\hat{\xi}_a^{(\hat{v},\hat{p})} = \frac{1}{M}\sum_{m=1}^{M}(1 - s^{(\hat{v},\hat{p})}(u_m))I_a^{(v,p)}(u_m) - \rho_a$, and is equal to the error incurred in the fraction of impressions assigned to contract $a \in \mathcal{A}$ by a bid-price control that employs dual variables $\hat{v}$ and the tie-breaking rule $\hat{p}$, which are computed by solving problems (EC.3) and (EC.4) using the sample $\{u_m\}_{m=1}^{M}$. Because the tie-breaking probabilities are determined by considering only the subsets with positive probability $\cup_{a\in\mathcal{A}}\mathcal{S}_a$, even over the sample $\{u_m\}_{m=1}^{M}$ the fraction of impressions assigned to contract a could be different than ρ_a. The next results bounds the empirical assignment error by showing that the contribution of the ties that are not considered is limited because these essentially occur with zero probability.

Proposition EC.4. *Suppose that $M > A$. Let $(\hat{v},\hat{p})$ be the dual variables and tie-breaking probabilities obtained by solving problem* (EC.3) *and* (EC.4) *on a sample $\{u_m\}_{m=1}^{M}$. The expected empirical assignment error for contract $a \in \mathcal{A}$ is upper bounded by*

$$\mathbb{E}_M|\hat{\xi}_a^{(\hat{v},\hat{p})}| \leq \frac{A}{M},$$

where the expectation is taken over i.i.d. samples of length M.

The next result bounds the expected error in the assignment incurred by the learning phase of the algorithm by using tools from statistical learning theory.

Proposition EC.5. *Suppose that Assumption 1 holds and $M > A$. Let $(\hat{v},\hat{p})$ be the dual variables and tie-breaking probabilities obtained by solving problem* (EC.3) *and* (EC.4) *on a sample $\{u_m\}_{m=1}^{M}$ of length $M > 2$. The expected error is upper bounded by*

$$\mathbb{E}_M\|\xi^{(\hat{v},\hat{p})}\|_1 \leq 17\|\mathcal{S}\|_1\sqrt{\frac{A\log(M)}{M}},$$

where the expectation is taken over i.i.d. samples of length M.

The proof of the main result follows from noting that the expected performance of the algorithm is given by $J^L = \mathbb{E}_M\left[J^{(\hat{v},\hat{p})}\right]$ where the expectation is taken over i.i.d. samples of length M. Hence one obtains that

$$\frac{J^* - J^L}{N} = \frac{J^* - \mathbb{E}_M\left[J^{(\hat{v},\hat{p})}\right]}{N} \leq \frac{V\underline{\rho}K}{\sqrt{N}} + 2V\mathbb{E}_M\|\xi^{(\hat{v},\hat{p})}\|_1 \leq \frac{V\underline{\rho}K}{\sqrt{N}} + 34V\|\mathcal{S}\|_1\sqrt{\frac{A\log(M)}{M}},$$

where the first inequality follows from taking expectations in the bound of Proposition EC.3 and that $J^* \leq J^D$ from Theorem 3; and the second inequality from Proposition EC.5.

EC.5.4. Optimal Sample Size

In Policy 1 the parameters of the policy are learned on an independent set $\{u_m\}_{m=1}^M$ of length M. In practice one would employ ϵN of the horizon to learn the parameters and then implement the resulting policy in the remaining $(1-\epsilon)N$ of the horizon. In choosing the fraction ϵ, the decision-maker faces the typical trade-off between exploration to reduce the estimation error and exploitation of the best estimated parameters. The bound of Theorem EC.1 allows one to determine the optimal fraction ϵ that balances these two effects.

Let $\mu^{L(\epsilon)}$ be a policy with an exploration phase of ϵN impressions followed by an exploitation phase $(1-\epsilon)N$ impressions. During the exploration phase impressions are assigned arbitrarily to the different contracts so that each contract receives exactly ϵC_a impressions (to simplify the argument we ignore rounding issues). After learning the dual variables $\hat{v}$ and tie-breaking probabilities $\hat{p}$, the resulting policy is implemented during the exploitation phase. Let $J^{L(\epsilon)}$ denote the expected performance of this algorithm, which takes into account the yield loss during the exploration phase.

The expected regret of this algorithm w.r.t. the optimal policy is upper bounded by

$$\begin{aligned}\frac{J^* - J^{L(\epsilon)}}{N} &\leq \frac{J^* - \mathbb{E}_{\epsilon N}\left[J^{(\hat{v},\hat{p})}_{(1-\epsilon)N}\right]}{N} \leq \frac{\epsilon J^D}{N} + V\underline{\rho}K\sqrt{\frac{1-\epsilon}{N}} + \frac{(1-\epsilon)N+1}{N}V\mathbb{E}_{\epsilon N}\left[\|\xi^{(v,p)}\|\right] \\ &\leq \epsilon V\underline{\rho} + V\underline{\rho}K\sqrt{\frac{1-\epsilon}{N}} + 34(1-\epsilon)V\|\mathcal{S}\|_1\sqrt{\frac{A\log(\epsilon N)}{\epsilon N}}\,, \end{aligned} \tag{EC.5}$$

where the first inequality follows from discarding the yield of the exploration phase to obtain $J^{L(\epsilon)} \geq \mathbb{E}_{\epsilon N}\left[J^{(\hat{v},\hat{p})}_{(1-\epsilon)N}\right]$, where $J^{(\hat{v},\hat{p})}_{(1-\epsilon)N}$ is the expected yield of a policy with parameters $(\hat{v},\hat{p})$ in an horizon of length $(1-\epsilon)N$ with capacities $(1-\epsilon)C_a$; and the expectation is taken over the ϵN samples in the exploration phase; the second inequality follows from Proposition EC.3; and the third inequality from the fact that $J^D \leq NV\underline{\rho}$ and Proposition EC.5.

The optimal fraction is chosen by minimizing the bound EC.5 over $\epsilon \in (0,1)$. For example by choosing $\epsilon \sim N^{-1/3}$ we obtain that the regret is of order $O(\sqrt{\log N}N^{-1/3})$.

EC.5.5. Proof of Auxiliary Results

Proof of Lemma EC.1. Let $\hat{\psi}(v)$ be the objective function of the dual SAA problem (EC.3). We show that the dual variables are bounded by proving that if for any $a \in \mathcal{A}$ the dual variables lie outside $[-V, V]$ then the objective value is largen than $\hat{\psi}(0)$ regardless of the value of the other variables.

Using that $R(c) \leq \bar{B} + c$ we can upper bound the objective value at zero by

$$\hat{\psi}(0) = \frac{1}{M}\sum_{m=1}^M R\Big(\max_{a\in\mathcal{A}_0}\{\gamma q_{m,a}\}; u_n\Big) \leq \bar{B} + \frac{1}{M}\sum_{m=1}^M \max_{a\in\mathcal{A}_0}\{\gamma q_{m,a}\} \leq \bar{B} + \gamma A\bar{Q}\,,$$

where the last inequality follows from the fact that qualities are bounded by $\bar{Q}$.

Fix $a \in \mathcal{A}$. We first prove the upper bound by showing that $\bar{\psi}(v) \geq \bar{\psi}(0)$ whenever the a^{th}-component of the vector satisfies $v_a \geq V$. Note that $\hat{\psi}(v) \geq \rho \cdot v$ and $\hat{\psi}(v) \geq \gamma\underline{Q} - v_{a'} + \rho \cdot v$ for $a' \in \mathcal{A} \setminus \{a\}$

because $R(c) \geq c$. Taking the point-wise maximum of these inequalities implies that $\hat{\psi}(v) \geq \rho_a v_a + \left(\sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'} v_{a'} + \max_{a' \in \mathcal{A}\setminus\{a\}} \left\{\gamma \underline{Q} - v_{a'}\right\}^+\right)$. The term in parenthesis on the right hand-side can be understood as the dual objective of an assignment problem without contract a, with weights $\gamma \underline{Q}$ and capacity constraints exactly equal to $\rho_{a'}$. It is not hard to show that the optimal objective value of the latter is $\gamma \underline{Q} \sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'}$. By minimizing over $v_{a'}$ for $a' \neq a$ we obtain that

$$\hat{\psi}(v) \geq \rho_a v_a + \gamma \underline{Q} \sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'} \geq \underline{\rho} V + \gamma \underline{Q} = \bar{B} + \gamma A \bar{Q} \geq \hat{\psi}(0)\,,$$

where the second inequality follows from the fact that $\rho_a \geq \underline{\rho}$, $v_a \geq V$ and $\sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'} \leq 1$ together with $\underline{Q} \leq 0$; the equality from our definition for V; and the last from the upper bound on $\hat{\psi}(0)$.

For the lower bound we prove that $\bar{\psi}(v) \geq \bar{\psi}(0)$ whenever $v_a \leq -V$. Similarly to the previous case we have that $\hat{\psi}(v) \geq \gamma \underline{Q} - v_{a'} + \rho \cdot v$ for all $a' \in \mathcal{A}$. Taking the point-wise maximum implies that $\hat{\psi}(v) \geq \rho_a v_a + \left(\sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'} v_{a'} + \max_{a' \in \mathcal{A}} \left\{\gamma \underline{Q} - v_{a'}\right\}^+\right)$. By minimizing over $v_{a'}$ for $a' \neq a$ we obtain that

$$\hat{\psi}(v) \geq \gamma \underline{Q} - v_a \left(1 - \sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'}\right) \geq \gamma \underline{Q} + \underline{\rho} V = \bar{B} + \gamma A \bar{Q} \geq \hat{\psi}(0)\,,$$

where the second inequality follows from the fact that $-v_a \geq V$ and $1 - \sum_{a' \in \mathcal{A}\setminus\{a\}} \rho_{a'} = \rho_a + \rho_0 \geq \underline{\rho}$; the equality from our definition for V; and the last from the upper bound on $\hat{\psi}(0)$. $\square$

Proof of Lemma EC.2. We prove the result by following closely the proof of Proposition 2. Consider problem (EC.4) without being restricted to the subsets with positive probability, which can be written as

$$\sum_{S \subseteq \mathcal{A}_0 : a \in S} \hat{\mathbb{P}}(S\text{-tie}) p_a(S) = \rho_a, \qquad \forall a \in \mathcal{A}\,, \tag{EC.6a}$$

$$\sum_{a \in S} p_a(S) = 1, \qquad \forall S \subseteq \mathcal{A}_0\,, \tag{EC.6b}$$

$$p_a(S) \geq 0, \qquad \forall S \subseteq \mathcal{A}_0, a \in S\,. \tag{EC.6c}$$

The latter is identical to the assignment problem (5) with the unknown probability measure $\mathbb{P}(U \in B)$ replaced by the empirical measure $\hat{\mathbb{P}}(U \in B) = \frac{1}{M}\sum_{m=1}^{M} \mathbf{1}\{u_m \in B\}$. Repeating the steps of Proposition 2's proof with the empirical measure $\hat{\mathbb{P}}$ instead of unknown probability measure $\mathbb{P}$, and using the SAA version of the dual (EC.3) instead of (4), one obtains that problem (EC.6) admits a feasible solution. Using a solution of the latter as a starting point we shall construct a solution to our problem.

Let $\hat{p}_a(S)$ be a solution of problem (EC.6). Consider a solution $\hat{p}'_a(S)$ given by

$$\hat{p}'_a(S) = \begin{cases} \hat{p}_a(S)\,, & \text{if } S \in \cup_{a \in \mathcal{A}} \mathcal{S}_a\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

We need to show that this solution is feasible for (EC.4). Non-negativity follows trivially, equation (EC.4c) follows directly from (EC.6b), and by construction we have that $\hat{p}'_a(S)=0$ for all subsets with zero probability. For the capacity constraints we obtain by dividing the sum over the subsets $S\subseteq\mathcal{A}_0 : a\in S$ into $\mathcal{S}_a$ and $\mathcal{S}_a^c$, and using the definition of $\hat{p}'_a(S)$ that

$$\sum_{S\in\mathcal{S}_a}\hat{\mathbb{P}}(S\text{-tie})\hat{p}'_a(S)\le\sum_{S\subseteq\mathcal{A}_0:a\in S}\hat{\mathbb{P}}(S\text{-tie})\hat{p}_a(S)\le\sum_{S\in\mathcal{S}_a}\hat{\mathbb{P}}(S\text{-tie})\hat{p}'_a(S)+\sum_{S\notin\mathcal{S}_a}\hat{\mathbb{P}}(S\text{-tie})$$

where the inequality to the left follows from discarding the terms in $\mathcal{S}_a^c$ because $\hat{p}_a(S)\ge 0$; while the equation to the right follows from the fact that $\hat{p}_a(S)\le 1$. From equation (EC.6a) we have that $\rho_a=\sum_{S\subseteq\mathcal{A}_0:a\in S}\hat{\mathbb{P}}(S\text{-tie})\hat{p}_a(S)$, which implies equations (EC.4a) and (EC.4b), respectively. $\square$

Proof of Proposition EC.3. We prove the result by extending the proof of Theorem 2 to take into account the errors in the assignment. In the remainder of the proof time periods are indexed *forward in time*.

Let $S^\mu_{n,a}=\sum_{i=1}^n\left(1-X_i(s^\mu_i(U_i))\right)I^\mu_{i,a}(U_i)$ be the total number of impressions assigned to advertiser a by time n when following the stochastic policy μ that employs dual variables v and the tie-breaking rule p. Additionally, we denote by $S^\mu_n=\{S^\mu_{n,a}\}_{a\in\mathcal{A}}$ the random vector of impressions assigned to advertisers (with the case of $a=0$ defined as before). Let the stopping time $N^\mu=\inf\left\{1\le n\le N : S^\mu_{n,a}=C_a \text{ for some } a\in\mathcal{A}_0\right\}$ be the first time that any advertiser's contract is fulfilled or the point is reached where all arriving impressions need to be assigned to the advertisers. Let Y^μ_n be the yield from impression n under policy μ.

As in the proof of Theorem 2 we define by $S_{n,a}$ the number of impressions assigned to contract $a\in\mathcal{A}_0$ by time n and Y_n the yield when following the deterministic controls in an alternate coupled system with no capacity constraints. It is the case that $S_{n,a}=S^\mu_{n,a}$ and $Y_n=Y^\mu_n$ for $n<N^\mu$. We get that the expected yield is lower bounded by

$$J^{(v,p)}=\mathbb{E}\left[\sum_{n=1}^N Y^\mu_n\right]\ge\mathbb{E}\left[\sum_{n=1}^{N^\mu}Y_n\right]=\mathbb{E}[N^\mu]\mathbb{E}[Y_1],$$

where the inequality follows from discarding terms after the stopping time, and the last equality from Wald's equation. We next bound each term at a time.

Expected Yield. Before proceeding to lower bound the expected yield of a single impression, we observe that the total contract adjusted quality satisfies

$$\begin{aligned}\sum_{a\in\mathcal{A}}I^{(v,p)}_a(U)(\gamma Q_a-v_a)&=\sum_{a\in\mathcal{A}}\sum_{S\subseteq\mathcal{A}_0:a\in S}p_a(S)(\gamma Q_a-v_a)\mathbf{1}\left\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma Q_{a'}-v_{a'}\}\right\}\\&=\max_{a\in\mathcal{A}_0}\{\gamma Q_a-v_a\}\sum_{S\subseteq\mathcal{A}_0}\left(\mathbf{1}\left\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma Q_{a'}-v_{a'}\}\right\}\sum_{a\in S}p_a(S)\right)\\&=\max_{a\in\mathcal{A}_0}\{\gamma Q_a-v_a\}\mathbf{1}\left\{\mathcal{S}\ni\arg\max_{a'\in\mathcal{A}_0}\{\gamma Q_{a'}-v_{a'}\}\right\}\overset{(a.s.)}{=}\max_{a\in\mathcal{A}_0}\{\gamma Q_a-v_a\},\end{aligned}$$

where the first equation follows from the definition of the bid-price control $I_a^{(v,p)}(U)$; the second from exchanging summations and using that $\gamma Q_a - v_a$ achieves the maximum whenever $a \in S$; the third equation because the tie-breaking rule satisfies that $\sum_{a\in S} p_a(S) = 1$ for $S \in \mathcal{S}$ and at most one subset in $\mathcal{S}$ verifies the maximum because $p_a(S) = 0$ for $S \notin \mathcal{S}$; and the last because with probability one the maximum the verified by one subset in $\mathcal{S}$.

Let $\bar{J}^D = J^D/N$ and $\bar{\psi}(v) = \psi(v)/N$ be the normalized optimal yield from the DAP and dual function, respectively. Using the previous remark the yield of a single impression can be written as a function of the dual objective as follows

$$
\begin{aligned}
\mathbb{E}[Y_1] &= \mathbb{E}\left[r(s^{(v,p)}(U);U) + \gamma\left(1 - s^{(v,p)}(U)\right) \sum_{a\in\mathcal{A}} I_a^{(v,p)}(U) Q_a \right] \\
&= \mathbb{E}\left[r(s^{(v,p)}(U);U) + \left(1 - s^{(v,p)}(U)\right) \sum_{a\in\mathcal{A}} I_a^{(v,p)}(U)(\gamma Q_a - v_a) \right] + \sum_{a\in\mathcal{A}} \rho_a v_a + \sum_{a\in\mathcal{A}} v_a \xi_a^{(v,p)} \\
&= \mathbb{E}\left[R\left(\max_{a\in\mathcal{A}_0}\{\gamma Q_a - v_a\} \right) \right] + \sum_{a\in\mathcal{A}} \rho_a v_a + \sum_{a\in\mathcal{A}} v_a \xi_a^{(v,p)} \\
&= \bar{\psi}(v) + \sum_{a\in\mathcal{A}} v_a \xi_a^{(v,p)} \geq \bar{J}^D + \sum_{a\in\mathcal{A}} v_a \xi_a^{(v,p)},
\end{aligned}
$$

where the third equation follows the previous remark and because the acceptance probability is chosen optimality with respect to that opportunity cost; the fourth equation follows from the definition of the dual function, and the inequality from weak duality for the DAP. Using the fact that dual variables satisfy $|v_a| \leq V$ we obtain the final bound $\mathbb{E}[Y_1^{(v,p)}] \geq \bar{J}^D - V||\xi^{(v,p)}||_1$.

Expected Stopping Time. For the stopping time recall that we have that $N^\mu \stackrel{(d)}{=} \min_{a\in\mathcal{A}_0}\{N_a\}$ where we define $N_a = \inf\{n \geq 1 : S_{n,a} = C_a\}$ as the time when the contract of advertiser $a \in \mathcal{A}_0$ is fulfilled. For $a \in \mathcal{A}_0$, the summands of $S_{n,a}$ are independent Bernoulli random variables with success probability $\rho_a + \xi_a^{(v,p)}$. Hence, $N_a - C_a$ is distributed as a negative binomial random variable with C_a successes and success probability $\rho_a + \xi_a^{(v,p)}$. The mean and variance are given by $\mathbb{E}N_a = N\frac{\rho_a}{\rho_a + \xi_a^{(v,p)}}$, and $\mathrm{Var}[N_a] = N\rho_a \frac{1-\rho_a-\xi_a^{(v,p)}}{(\rho_a+\xi_a^{(v,p)})^2}$, where we used that $\rho_a = C_a/N$. In the latter we set $\xi_0^{(v,p)} = 1 - \sum_{a\in\mathcal{A}} \mathbb{E}\big[(1 - s^{(v,p)}(U)) I_a^{(v,p)}(U)\big] - \rho_0 = -\sum_{a\in\mathcal{A}} \xi_a^{(v,p)}$ because $\rho_0 = 1 - \sum_{a\in\mathcal{A}} \rho_a$.

Consider yet another alternate system in which we truncate the error term to its positive part $(\xi_a^{(v,p)})^+$, and let $\tilde{N}_a^\mu$ be the stopping time for contract a in this alternate system. That is, $\tilde{N}_a - C_a$ is distributed as a negative binomial random variable with C_a successes and success probability $\rho_a + (\xi_a^{(v,p)})^+$. Clearly we have that N_a stochastically dominates $\tilde{N}_a$ because successes are less likely in the former and more trials are needed to reach the capacity C_a. Additionally, the mean is lower bounded by $\mathbb{E}\tilde{N}_a^\mu = N\frac{\rho_a}{\rho_a+(\xi_a^{(v,p)})^+} \geq N\left(1 - \frac{(\xi_a^{(v,p)})^+}{\rho_a}\right)$ and the variance is upper bounded by $\mathrm{Var}[\tilde{N}_a] = N\rho_a\frac{1-\rho_a-(\xi_a^{(v,p)})^+}{(\rho_a+(\xi_a^{(v,p)})^+)^2} \leq N\frac{1-\rho_a}{\rho_a}$ because $(\xi_a^{(v,p)})^+ \geq 0$. Using the fact that the minimum is non-decreasing together with the lower bound on the mean of the minimum of a number of random variables of Aven (1985) we get that

$$
\mathbb{E}N^\mu = \mathbb{E}\min_{a\in\mathcal{A}_0}\{N_a\} \geq \mathbb{E}\min_{a\in\mathcal{A}_0}\{\tilde{N}_a\} \geq \min_{a\in\mathcal{A}_0} \mathbb{E}\tilde{N}_a - \sqrt{\frac{A}{A+1}\sum_{a\in\mathcal{A}_0} \mathrm{Var}[\tilde{N}_a]}
$$

$$\geq N \min_{a\in\mathcal{A}_0}\left(1-\frac{(\xi_a^{(v,p)})^+}{\rho_a}\right)-\sqrt{\frac{A}{A+1}}\sqrt{\sum_{a\in\mathcal{A}_0} N\frac{1-\rho_a}{\rho_a}}$$
$$= N\left(1-\max_{a\in\mathcal{A}_0}\frac{(\xi_a^{(v,p)})^+}{\rho_a}\right)-\sqrt{N}K\,,$$

where the third inequality follows from the bounds on the means and variances, and the last from our definition of K in the statement of Theorem 2. Using the fact that $\rho_a \geq \underline{\rho}$ and that $\max_{a\in\mathcal{A}_0}(\xi_a^{(v,p)})^+ \leq \|\xi^{(v,p)}\|_1$ because $\xi_0^{(v,p)} = -\sum_{a\in\mathcal{A}}\xi_a^{(v,p)}$ we obtain the final bound $\mathbb{E}N^\mu \geq N-\sqrt{N}K-\|\xi^{(v,p)}\|_1/\underline{\rho}$.

Putting it all together. Combining the previous bounds we obtain that

$$J^{(v,p)} \geq \mathbb{E}[N^\mu]\mathbb{E}[Y_1] \geq \left(\bar{J}^D - V\|\xi^{(v,p)}\|_1\right)^+\left(N-\sqrt{N}K-\|\xi^{(v,p)}\|_1/\underline{\rho}\right)^+$$
$$\geq J^D - \bar{J}^D K\sqrt{N} - (NV+\bar{J}_D/\underline{\rho})\|\xi^{(v,p)}\|_1\,,$$

where the last inequality follows from the identity $(a-b|x|)^+(c-d|x|)^+ \geq ac-(ad+bc)|x|$ for $a,b,c,d\geq 0$ because $N\geq K^2$ and dropping one term. The result follows from $\bar{J}^D \leq \bar{\psi}(0) = V\underline{\rho}$ from the proof of the previous lemma. □

Proof of Proposition EC.4. Fix the sample $\{u_m\}_{m=1}^M$. From Lemma EC.2 we know that problem (EC.4) admits a feasible solution. We have that the empirical assignment error can be equivalently written as

$$\hat{\xi}_a^{(\hat{v},\hat{p})} = \frac{1}{M}\sum_{m=1}^M (1-s^{(\hat{v},\hat{p})}(u_m))I_a^{(\hat{v},\hat{p})}(u_m)-\rho_a = \sum_{S\subseteq\mathcal{A}_0: a\in S}\hat{p}_a(S)\hat{\mathbb{P}}(S\text{-tie})-\rho_a = \sum_{S\in\mathcal{S}_a}\hat{p}_a(S)\hat{\mathbb{P}}(S\text{-tie})-\rho_a\,,$$

where the second equality follows from the definition of the controls $s^{(\hat{v},\hat{p})}(u)$ and $I_a^{(\hat{v},\hat{p})}(u)$, exchanging summations, and the definition of $\hat{\mathbb{P}}(S\text{-tie})$; and the last from the fact that $\hat{p}_a(S)=0$ for $S\in\mathcal{S}_a^c$.

The capacity constraint (EC.4a) implies that $\hat{\xi}_a^{(\hat{v},\hat{p})}\leq 0$, while the capacity constraint (EC.4b) implies that $\hat{\xi}_a^{(\hat{v},\hat{p})} \geq -\sum_{S\in\mathcal{S}_a^c}\hat{\mathbb{P}}(S\text{-tie})$. Taking expectations w.r.t. the samples we obtain that

$$E_M\left|\hat{\xi}_a^{(\hat{v},\hat{p})}\right| \leq E_M\left[\sum_{S\in\mathcal{S}_a^c}\hat{\mathbb{P}}(S\text{-tie})\right] \leq E_M\left[\frac{1}{M}\sum_{m=1}^M\sum_{S\in\mathcal{S}_a^c}\mathbf{1}\left\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma Q_{m,a'}-\hat{v}_{a'}\}\right\}\right]$$
$$\leq \frac{1}{M}\mathbb{E}_M\Bigg[\underbrace{\sup_{v\in\mathbb{R}^A}\sum_{m=1}^M\mathbf{1}\left\{\mathcal{S}_a^c\ni\arg\max_{a'\in\mathcal{A}_0}\{\gamma Q_{m,a'}-v_{a'}\}\right\}}_{(\triangle)}\Bigg]\,,$$

where the second inequality follows because $\hat{\mathbb{P}}(S\text{-tie}) \leq \frac{1}{M}\sum_{m=1}^M\mathbf{1}\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma q_{m,a'}-\hat{v}_{a'}\}\}$ and exchanging summations, and the last from taking the supremum over all possible dual variables $v\in\mathbb{R}^A$ and using the fact that at most subset in $\mathcal{S}_a^c$ verifies the maximum.

Let $\nu(q,S)$ be the set of dual variables for which the maximum contract adjusted quality of $q\in\Omega$ is verified by all contracts in the subset $S\subseteq\mathcal{A}_0$. Let $B=\left\{(Q_m)_{m=1}^M : \bigcap_{n=1}^N\bigcup_{S\in\mathcal{S}_a^c}\nu(Q_m,S)=\emptyset \text{ for all subsequence } \{m_n\}_{n=1}^N \text{ with } N\geq A+1\right\}$ be the event that there is no dual variable v that guarantees that the impression is assigned to more than $A+1$ samples when the control is restricted to ties in the set S_a^c. By conditioning on this event we obtain that

$$E_M\left|\hat{\xi}_a^{(\hat{v},\hat{p})}\right| \le \frac{1}{M}\mathbb{E}_M[(\triangle)\mathbf{1}\{B\}+(\triangle)\mathbf{1}\{\bar{B}\}] \le \frac{1}{M}\mathbb{E}_M[A\mathbf{1}\{B\}+M\mathbf{1}\{\bar{B}\}] = \frac{A}{M}\mathbb{P}\{B\}+(1-\mathbb{P}\{B\}) = \frac{A}{M}\,,$$

where the second inequality follows from the fact that when the event B is true the supremum $(\triangle)$ is at most A and the supremum $(\triangle)$ is always less than M. The last equality follows from the fact that Lemma EC.5 implies that $\mathbb{P}\{B\}=1$ whenever $M\ge A+1$.

Proof of Proposition EC.5. We proceed to bound the error term of each contract $a\in\mathcal{A}$. Let $i_a^{(v,p)}(u) = \left(1-s^{(v,p)}(u)\right) I_a^{(v,p)}(u) = \sum_{S\in\mathcal{S}_a} p_a(S)\left(1-s^*(\gamma q_a - v_a;u)\right)\mathbf{1}\left\{S=\arg\max_{a'\in\mathcal{A}_0}\{\gamma q_{a'}-v_{a'}\}\right\}$ be the total probability that the impression is assigned to contract a when the user attribute is $u\in\mathcal{U}$, where we used that only probabilities in the set $\mathcal{S}_a$ have positive mass. Using the triangle inequality, the error incurred by the learning phase can be decomposed into an estimation error and an empirical assignment error as follows

$$\begin{aligned}
\mathbb{E}_M\left|\xi_a^{(\hat{v},\hat{p})}\right| &= \mathbb{E}_M\left|E_U\left[i_a^{(\hat{v},\hat{p})}(U)|(\hat{v},\hat{p})\right]-\rho_a\right| \\
&\le \mathbb{E}_M\left|E_U\left[i_a^{(\hat{v},\hat{p})}(U)|(\hat{v},\hat{p})\right]-\frac{1}{M}\sum_{m=1}^{M} i^{(\hat{v},\hat{p})}(U_m)\right| + \mathbb{E}_M\left|\frac{1}{M}\sum_{m=1}^{M} i^{(\hat{v},\hat{p})}(U_m)-\rho_a\right| \\
&= \mathbb{E}_M\left|\chi_a^{(\hat{v},\hat{p})}\right| + \mathbb{E}_M\left|\hat{\xi}_a^{(\hat{v},\hat{p})}\right|,
\end{aligned}$$

where we denote by $\chi_a^{(\hat{v},\hat{p})} = E_U\left[i_a^{(\hat{v},\hat{p})}(U)|(\hat{v},\hat{p})\right] - \frac{1}{M}\sum_{m=1}^{M} i^{(\hat{v},\hat{p})}(u_m)$ the estimation error. The second term is controlled in Proposition EC.4. In the remainder of this proof we bound the first term.

We have that

$$\begin{aligned}
\mathbb{E}_M\left|\chi_a^{(\hat{v},\hat{p})}\right| &= \mathbb{E}_M\left|E_U\left[i_a^{(\hat{v},\hat{p})}(U)|(\hat{v},\hat{p})\right]-\frac{1}{M}\sum_{m=1}^{M} i^{(\hat{v},\hat{p})}(u_m)\right| \\
&\le \mathbb{E}_M\left[\sup_{(v,p)\in\mathcal{E}}\left|E_U\left[i_a^{(v,p)}(U)|(v,p)\right]-\frac{1}{M}\sum_{m=1}^{M} i_a^{(v,p)}(U_m)\right|\right] \\
&\le 2\mathbb{E}_{M,\sigma}\left[\sup_{(v,p)\in\mathcal{E}}\left|\frac{1}{M}\sum_{m=1}^{M}\sigma_m i_a^{(v,p)}(U_m)\right|\right] = 2E_M\left[R_M(\mathcal{F}_a,\vec{U})\right]\,,
\end{aligned}$$

where first inequality follows from taking the supremum over the space of dual variables v and tie-breaking probabilities p given by $\mathcal{E}=\mathbb{R}^A\times(2^{\mathcal{A}_0}\to\Delta^{A+1})$ where $\Delta^{A+1}=\{x\in\mathbb{R}^{A+1}: x_a\ge 0, \sum x_a = 1\}$ is the probability simplex of dimension $A+1$, and the second inequality from a symmetrization argument where $\{\sigma_m\}_{m=1}^M$ are i.i.d. Rademacher random variables (see, e.g., Mendelson (2003)). The last expression was bounded by $2E_M\left[R_M(\mathcal{F}_a,\vec{U})\right]$ where $R_M(\mathcal{F}_a,\vec{u})$ is the Rademacher complexity of the function class $\mathcal{F}_a=\{i_a^{(v,p)}:\mathcal{U}\to[0,1]:\forall(v,p)\in\mathcal{E}\}$ with each element a function in the space $\mathcal{U}\to[0,1]$.

We bound the Rademacher complexity by considering a larger class of functions. Let $\mathcal{F}_s=\{1-s^*(\gamma q_a - v_a;u): v_a\in\mathbb{R}\}$ be the function class of exchange acceptance probabilities parameterized by the dual variable $v_a \in \mathbb{R}$. For every subset $S \subseteq \mathcal{A}_0$ such that $a \in S$ we consider the class $\mathcal{F}_a(S) = \left\{\mathbf{1}\left\{S = \arg\max_{a' \in \mathcal{A}_0}\{\gamma q_{a'} - v_{a'}\}\right\} : v \in \mathbb{R}^A\right\}$ of indicator functions that contracts in the subset S verify the maximum contract adjusted quality parameterized by dual variables $v \in \mathbb{R}^A$. We consider the class of functions $\mathcal{H} = \overline{(\mathcal{F}_s \cdot \mathcal{F}_a(S))}_{S \in \mathcal{S}_a}$ given by the weighted sum over subsets $S \in \mathcal{S}_a$ of the product of functions from $\mathcal{F}_s$ and $\mathcal{F}_a(S)$ (see Lemma EC.3 and Lemma EC.4 for notation). This class verifies that $\mathcal{F}_a \subseteq \mathcal{H}$ since we allow different dual variables for each tie and pricing in the exchange.

Using Lemma EC.4 together with the fact that $\mathcal{F}_a \subseteq \mathcal{H}$ we obtain that for every sample $\vec{u} \in \mathcal{U}^M$

$$R_M(\mathcal{F}_a, \vec{u}) \le R_M(\mathcal{H}, \vec{u}) \le \sum_{S \subseteq \mathcal{S}_a} R_M(\mathcal{F}_s \cdot \mathcal{F}_a(S), \vec{u}) \,.$$

By the Discretization Theorem (see, e.g., Mendelson (2003)) we can upper bound the Rademacher complexity of each product class in terms of its α-covering number $\mathcal{N}(\alpha, \mathcal{F}_s \cdot \mathcal{F}_a(S), M)$ (precise definitions are provided in Section EC.5.6) as follows

$$R_M(\mathcal{F}_s \cdot \mathcal{F}_a(S), \vec{u}) \le \inf_{\alpha > 0} \alpha + \sqrt{\frac{2\log \mathcal{N}(\alpha, \mathcal{F}_s \cdot \mathcal{F}_a(S), M)}{M}} \,. \tag{EC.7}$$

Next we proceed to bound the α-covering number of the function class.

For the class $\mathcal{F}_s$ we have that for a fixed sample $\vec{u} \in \mathcal{U}^M$ the set $\mathcal{F}|_{\vec{u}}$ is a curve in the space $[0,1]^M$ parameterized by the dual variable $v_a \in \mathbb{R}$. Because $s^*(\cdot)$ is non-increasing, the curve is non-increasing and has a length of at most M w.r.t. the ℓ_1 norm. Performing a natural parametrization w.r.t. the ℓ_1 norm, this curve can be equivalently written as $\phi : [0, M] \to [0,1]^M$, where we interpolate linearly through discontinuities. For any $\alpha > 0$ one can split the domain in M/α points to obtain an α-covering. Thus, $\mathcal{N}(\alpha, \mathcal{F}_s, M) \le M/\alpha$. For the class $\mathcal{F}_a(S)$ it is not hard to see that the Vapnik-Chervonenkis dimension is $d \le A$, which implies by the Sauer-Shelah lemma that for all $M \ge d$ we have that the covering number is bounded as $\mathcal{N}(\alpha, \mathcal{F}_a(S), M) \le (eM/d)^d$ for any $\alpha \ge 0$. Using Lemma EC.3 we obtain that

$$\mathcal{N}(\alpha, \mathcal{F}_s \cdot \mathcal{F}_a(S), M) \le \mathcal{N}(\alpha, \mathcal{F}_s, M)\mathcal{N}(0, \mathcal{F}_a(S), M) \le \frac{M}{\alpha}\left(\frac{eM}{d}\right)^d \,.$$

Using the equation (EC.7) we get that

$$R_M(\mathcal{F}_s \cdot \mathcal{F}_a(S), \vec{u}) \le \inf_{\alpha > 0} \alpha + \sqrt{\frac{2\log \mathcal{N}(\alpha, \mathcal{F}_s \cdot \mathcal{F}_a(S), M)}{M}} \le 2\sqrt{\frac{\log(M^{3+2d}e^{2d}d^{-2d})}{M}} \le 8\sqrt{\frac{A\log(M)}{M}} \,,$$

where the second inequality follows from using the parameter balancing technique of Ahuja et al. (1993, page 65) and choosing α equal to the second term in the infimum to obtain a 2-approximation, and the last is a loose bound for $M > 2$.

Putting it all together. Summing over all subsets in $\mathcal{S}_a$ we obtain that $\mathbb{E}_M\left|\chi_a^{(\hat{v},\hat{p})}\right| \le 2E_M\left[R_M(\mathcal{F}_a, \vec{U})\right] \le 16|\mathcal{S}_a|\sqrt{\frac{A\log(M)}{M}}$. Summing over all contracts $a \in \mathcal{A}$ and using Proposition EC.4 we get that

$$\mathbb{E}_M\|\xi^{(\hat{v},\hat{p})}\|_1 \le \mathbb{E}_M\|\chi^{(\hat{v},\hat{p})}\|_1 + \mathbb{E}_M\|\hat{\xi}^{(\hat{v},\hat{p})}\|_1 \le 16\|\mathcal{S}\|_1\sqrt{\frac{A\log(M)}{M}} + \frac{A^2}{M} \le 17\|\mathcal{S}\|_1\sqrt{\frac{A\log(M)}{M}} \,,$$

where we used that $A \le \|\mathcal{S}\|_1$ together with the fact that $M > A$ and $M > 2$ to bound the second term. $\square$

EC.5.6. Auxiliary Results

For a set $X \subseteq [0,1]^M$, we say that the set $Y \subseteq [0,1]^M$ is an α-cover of X if for every $x \in X$ there is a $y \in Y$ such that $\|x-y\|_1 \le \alpha$. For a fixed $\alpha \ge 0$ we denote by $\mathcal{N}(X,\alpha)$ the cardinality of the smallest α-cover for X, that is, $\mathcal{N}(X,\alpha) = \inf\{|Y| : Y \text{ is an } \alpha\text{-cover of } X\}$.

Let $\mathcal{F} = \{f : \mathcal{U} \to [0,1]\}$ be a class of functions that map a sample in $\mathcal{U}$ to the interval $[0,1]$. For a fixed vector of samples $\vec{u} \in \mathcal{U}^M$ and a function $f \in \mathcal{F}$ we let $f|_{\vec{u}} = \big(f(u_1), f(u_2), \ldots, f(u_M)\big)$ be the vector of values of the function at the sample, and by $\mathcal{F}_{\vec{u}} = \{f|_{\vec{u}} : f \in \mathcal{F}\} \subseteq [0,1]^M$ the set of vectors of values of all functions in the class at the sample. The uniform $\alpha-$covering number of the class $\mathcal{F}$ for samples of size M is given by $\mathcal{N}(\alpha,\mathcal{F},M) = \sup_{\vec{u}\in\mathcal{U}^M} \mathcal{N}(\mathcal{F}|_{\vec{u}},\alpha)$.

Lemma EC.3. *Let F and G be two classes of functions and consider the class of product functions $\mathcal{F}\cdot\mathcal{G} = \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ where $(fg)(u) = f(u)g(u)$. Then for every $\alpha_g \ge 0$, $\alpha_f \ge 0$, and $M \ge 1$ we have that*

$$\mathcal{N}\big(\mathcal{F}\cdot\mathcal{G}, \alpha_f\alpha_g + \alpha_f + \alpha_g, M\big) \le \mathcal{N}\big(\mathcal{F},\alpha_f,M\big)\mathcal{N}\big(\mathcal{G},\alpha_g,M\big).$$

Proof. Fix the sample $\vec{u} \in \mathcal{U}^M$. Using the claim $\mathcal{N}\big(\mathcal{F}\cdot\mathcal{G}|_{\vec{u}}, \alpha_f\alpha_g + \alpha_f + \alpha_g\big) \le \mathcal{N}\big(\mathcal{F}|_{\vec{u}},\alpha_f\big)\mathcal{N}\big(\mathcal{G}|_{\vec{u}},\alpha_g\big)$, the result follows from taking the supremum over $\vec{u} \in \mathcal{U}^M$ and using that $\sup_x r(x)s(x) \le \big(\sup_x r(x)\big)\big(\sup_x s(x)\big)$ for any two non-negative functions $r(x)$ and $s(x)$.

In the remainder of the proof we prove the claim. Let Y_f be a minimal α_f-cover of $\mathcal{F}|_{\vec{u}}$, that is, $|Y_f| = \mathcal{N}\big(\mathcal{F}|_{\vec{u}},\alpha_f\big)$. Similarly, let Y_g be a minimal α_g-cover of $\mathcal{G}|_{\vec{u}}$. We shall show that $Y = \{x \circ y : x \in Y_f, y \in Y_g\}$ is an $(\alpha_f\alpha_g + \alpha_f + \alpha_g)$-cover of $\mathcal{F}\cdot\mathcal{G}|_{\vec{u}}$, where $x \circ y \triangleq (x_1y_1, \ldots, x_My_M)$ is the Hadamard product. For every $(fg)|_{\vec{u}} \in \mathcal{F}\cdot\mathcal{G}|_{\vec{u}}$ we can find $x \in Y_f$ and $y \in Y_g$ such that $\|f|_{\vec{u}} - x\|_1 \le \alpha_f$ and $\|g|_{\vec{u}} - y\|_1 \le \alpha_g$ because Y_f and Y_g are covers. We have that

$$\begin{aligned}
\|(fg)|_{\vec{u}} - x\circ y\|_1 &= \sum_{m=1}^M |f(u_m)g(u_m) - x_my_m| = \sum_{m=1}^M \big|\big(f(u_m) - x_m + x_m\big)\big(g(u_m) - y_m + y_m\big) - x_my_m\big| \\
&\le \sum_{m=1}^M |f(u_m) - x_m|\cdot|g(u_m) - y_m| + |x_m|\cdot|g(u_m) - y_m| + |y_m|\cdot|f(u_m) - x_m| \\
&\le \alpha_f\alpha_g + \alpha_f + \alpha_g\,,
\end{aligned}$$

where the first inequality follows from canceling terms and the triangle inequality, and the second because $|f(u_m) - x_m| \le \|f|_{\vec{u}} - x\|_1 \le \alpha_f$ and the codomain is $[0,1]$. Thus Y is an $(\alpha_f\alpha_g + \alpha_f + \alpha_g)$-cover for $\mathcal{F}\cdot\mathcal{G}|_{\vec{u}}$, and we have that $\mathcal{N}\big(\mathcal{F}\cdot\mathcal{G}|_{\vec{u}}, \alpha_f\alpha_g + \alpha_f + \alpha_g\big) \le |Y| \le |Y_f|\cdot|Y_g| = \mathcal{N}\big(\mathcal{F}|_{\vec{u}},\alpha_f\big)\mathcal{N}\big(\mathcal{G}|_{\vec{u}},\alpha_G\big)$, because Y_f and Y_g are minimal. $\square$

For a class of functions $\mathcal{F}$ and a vector of samples $\vec{u} \in \mathcal{U}^M$ we denote the Rademacher complexity as $R_M(\mathcal{F},\vec{u}) = \frac{1}{M}\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{m=1}^M \sigma_m f(u_m)\right|$ where the expectation is taken over i.i.d. Rademacher random variables $\{\sigma_m\}_{m=1}^M$.

Lemma EC.4. *Let $\mathcal{F} = \{\mathcal{F}_k\}_{k=1}^K$ be a finite set of function classes and consider the weighted sum class $\bar{\mathcal{F}} = \left\{\sum_{k=1}^K p_k f_k : f_k \in \mathcal{F}_k, p_k \in [0,1]\right\}$. Then for every vector of samples $\vec{u} \in \mathcal{U}^M$ we have that*

$$R_M(\bar{\mathcal{F}}, \vec{u}) \le \sum_{k=1}^K R_M(\mathcal{F}_k, \vec{u}) .$$

Proof. Exchanging the order of the summations and using the triangle inequality we obtain that

$$\begin{aligned} R_M(\bar{\mathcal{F}}, \vec{u}) &= \frac{1}{M}\mathbb{E} \sup_{f_k \in \mathcal{F}_k, p_k \in [0,1]} \left| \sum_{m=1}^M \sigma_m \sum_{k=1}^K p_k f_k(u_m) \right| \le \frac{1}{M}\mathbb{E} \sup_{f_k \in \mathcal{F}_k, p_k \in [0,1]} \sum_{k=1}^K p_k \left| \sum_{m=1}^M \sigma_m f_k(u_m) \right| \\ &= \frac{1}{M}\mathbb{E} \sup_{f_k \in \mathcal{F}_k} \sum_{k=1}^K \sup_{p_k \in [0,1]} p_k \left| \sum_{m=1}^M \sigma_m f_k(u_m) \right| = \frac{1}{M}\mathbb{E} \sup_{f_k \in \mathcal{F}_k} \sum_{k=1}^K \left| \sum_{m=1}^M \sigma_m f_k(u_m) \right| \\ &\le \sum_{k=1}^K \frac{1}{M}\mathbb{E} \sup_{f_k \in \mathcal{F}_k} \left| \sum_{m=1}^M \sigma_m f_k(u_m) \right| = \sum_{k=1}^K R_M(\mathcal{F}_k, \vec{u}) , \end{aligned}$$

where the second equality follows from partitioning the supremum and using that the objective is separable in the p_k's; the third equality follows because $p_k = 1$ is the optimal weight; and the last inequality follows from the sub-additivity of the supremum. □

Let $\nu(q, S) = \{v \in \mathbb{R}^A : \arg\max_{a \in \mathcal{A}_0}\{\gamma q_a - v_a\} = S\}$ the set of dual variables for which the maximum contract adjusted quality of $q \in \Omega$ is verified by all contracts in the subset $S \subseteq \mathcal{A}_0$. We refer to $\nu(q, S)$ as the *trigger set* of q w.r.t. S. The next result shows that with probability one there is no dual variable that guarantees that, for more than A samples, the maximum contract adjusted quality is verified by some subset $S \in \mathcal{S}_a^c \triangleq \{S \subseteq \mathcal{A}_0 : a \in S, S \notin \mathcal{S}_a\}$ for which the probability of a tie occurring is zero.

Lemma EC.5. *Let $\{Q_m\}_{m=1}^M$ be M i.i.d. samples of placement quality. Then, for all $a \in \mathcal{A}$ and $M \ge A+1$ we have that*

$$\mathbb{P}\left\{ \bigcap_{m=1}^M \bigcup_{S \in \mathcal{S}_a^c} \nu(Q_m, S) = \emptyset \right\} = 1 .$$

Proof. Let B_M be the event $\bigcap_{m=1}^M \bigcup_{S \in \mathcal{S}_a^c} \nu(Q_m, S) = \emptyset$. We have that $B_M \subseteq B_{M+1}$, which implies that $\mathbb{P}\{B_M\} \le \mathbb{P}\{B_{M+1}\}$ and thus it suffices to prove the result for $M = A+1$.

Note that for $S \subseteq \mathcal{A}_0$ with $a \in S$ and $S \notin \mathcal{S}_a$ we have that $\arg\max_{a' \in \mathcal{A}_0}\{\gamma q_{a'} - v_{a'}\} = S$ if and only if (i) $\gamma q_{a'} - v_{a'} = \gamma q_a - v_a$ for $a' \in S \setminus \{a\}$ and (ii) $\gamma q_{a'} - v_{a'} < \gamma q_a - v_a$ for $a' \in \mathcal{A}_0 \setminus S$. Let $T : \mathbb{R}^A \to \mathbb{R}^A$ be a linear map such that $T(z)_a = -z_a$ and $T(z)_{a'} = z_{a'} - z_a$ for $a' \in \mathcal{A} \setminus \{a\}$. Denoting by $t_m = T(q_m)$, and by $z = T(v)$ the latter conditions can be equivalently written as (i) $\gamma t_{a'} = z_{a'}$ for $a' \in h(S)$, and (ii) $\gamma t_{a'} < z_{a'}$ for $a' \in \mathcal{A} \setminus h(S)$, where $h(S) = S \setminus \{0\}$ if $0 \in S$ and $h(S) = S \setminus \{a\}$ otherwise.

Therefore the result is equivalent to showing that $\mathbb{P}\left\{\bigcap_{m=1}^M \bigcup_{S \in \mathcal{T}_a} \zeta(T_m, S) = \emptyset\right\} = 1$ where $\zeta(t, S) = \{z \in \mathbb{R}^A : \gamma t_a = z_a \forall a \in S, \gamma t_a < z_a \forall a \in \mathcal{A} \setminus S\}$ and $\mathcal{T}_a = \{h(S) : S \subseteq \mathcal{A}_0, a \in S, S \notin \mathcal{S}_a\}$. We have that $\emptyset \notin \mathcal{T}_a$ because $\{a\} \notin \mathcal{S}_a^c$ since $\mathbb{P}\{a = \arg\max_{a' \in \mathcal{A}_0}\{\gamma Q_{a'} - v_{a'}\}\} > 0$ for $v_a = \underline{Q}$ and $v_{a'} = \bar{Q}$ for $a' \ne a$. The definition of the set $\mathcal{S}_a$ implies that for all $S \in \mathcal{T}_a$ we have that $\mathbb{P}\{z \in \zeta(T_m, S)\} = 0$ for all $z \in \mathbb{R}^A$.

The result follows from the next claim. □

CLAIM EC.1. *Let $\mathcal{B} \subseteq \mathcal{A}$ be a subset of contracts. Let $\zeta(t,S) = \{z \in \mathbb{R}^A : \gamma t_a = z_a \forall a \in S, \gamma t_a < z_a \forall a \in \mathcal{A} \setminus S\}$ and $\mathcal{T} = \{S \subseteq \mathcal{A} : \mathbb{P}\{z \in \zeta(T,S)\} = 0$ for all $z \in \mathbb{R}^A\}$. Suppose that $\emptyset \notin \mathcal{T}$. Then*

$$\mathbb{P}\left\{\bigcap_{m=1}^{|\mathcal{B}|+1} \bigcup_{S \in \mathcal{T}, S \subseteq \mathcal{B}} \zeta(T_m, S) = \emptyset\right\} = 1\,,$$

where $\{T_m\}_{m=1}^{A+1}$ are i.i.d. random vectors with support in $\mathbb{R}^A$.

Proof of Claim. We prove the result by induction on the cardinality of $\mathcal{B}$, that is, $|\mathcal{B}|$. For the base case of $|\mathcal{B}| = 0$ the result is trivial because $\emptyset \notin \mathcal{T}$.

Suppose that the result is true for sets of cardinality $|\mathcal{B}| - 1$, we need to prove the result for sets of cardinality $|\mathcal{B}|$. Let $\vec{S} = (S_m)_{m=1}^{|\mathcal{B}|+1} \in \mathcal{T}^{|\mathcal{B}|+1}$ be a vector of subsets and $\Pi(\vec{T}, \vec{S}) = \bigcap_{m=1}^{|\mathcal{B}|+1} \zeta(T_m, S_m)$ the intersection of the trigger sets over the samples $\vec{T}$ for the vector of subsets $\vec{S}$. We have by the commutativity of the union and intersection that

$$\bigcap_{m=1}^{|\mathcal{B}|+1} \bigcup_{S \in \mathcal{T}, S \subseteq \mathcal{B}} \zeta(T_m, S) = \bigcup_{\vec{S} \in \mathcal{T}^{|\mathcal{B}|+1} : S_m \subseteq \mathcal{B}} \bigcap_{m=1}^{|\mathcal{B}|+1} \zeta(T_m, S_m) = \bigcup_{\vec{S} \in \mathcal{T}^{|\mathcal{B}|+1} : S_m \subseteq \mathcal{B}} \Pi(\vec{T}, \vec{S})\,.$$

The previous equation together with Boole's inequality imply that

$$\mathbb{P}\left\{\bigcap_{m=1}^{|\mathcal{B}|+1} \bigcup_{S \in \mathcal{T}, S \subseteq \mathcal{B}} \zeta(T_m, S) \neq \emptyset\right\} = \mathbb{P}\left\{\bigcup_{\vec{S} \in \mathcal{T}^{|\mathcal{B}|+1} : S_m \subseteq \mathcal{B}} \left\{\Pi(\vec{T}, \vec{S}) \neq \emptyset\right\}\right\} \le \sum_{\vec{S} \in \mathcal{T}^{|\mathcal{B}|+1} : S_m \subseteq \mathcal{B}} \mathbb{P}\left\{\Pi(\vec{T}, \vec{S}) \neq \emptyset\right\}.$$

We need to show that each summand on the right-hand side is zero.

Fix the vector of subsets $\vec{S} = (S_m)_{m=1}^{|\mathcal{B}|+1}$ with $S_m \in \mathcal{T}$ and $S_m \subseteq \mathcal{B}$. Let $\pi_{\mathcal{B}} : \mathbb{R}^A \to \mathbb{R}^{|\mathcal{B}|}$ be the projection onto the $\mathcal{B}$-plane, that is, $\pi_{\mathcal{B}}(z) = (z_a)_{a \in \mathcal{B}}$. Because the projection of a set is empty if and only if the set is empty, we study whether $\Pi(\vec{T}, \vec{S})$ is non-empty by considering whether the projection onto $\mathcal{B}$ given by $\pi_{\mathcal{B}}\left(\Pi(\vec{T}, \vec{S})\right)$ is empty. To this end we study the number of points in the projection onto $\mathcal{B}$ of the trigger sets of the first $|\mathcal{B}|$ samples. That is, let C_0, C_1 and C_∞ be the event that $\pi_{\mathcal{B}}\left(\Pi\left((T_m)_{m=1}^{|\mathcal{B}|}, (S_m)_{m=1}^{|\mathcal{B}|}\right)\right)$ is empty, has one point or multiple points. Because these events are disjoint we have that

$$\begin{aligned}
\mathbb{P}\left\{\Pi(\vec{T}, \vec{S}) \neq \emptyset\right\} &= \mathbb{P}\left\{\pi_{\mathcal{B}}\left(\Pi(\vec{T}, \vec{S})\right) \neq \emptyset\right\} \\
&= \mathbb{P}\left\{C_0, \pi_{\mathcal{B}}\left(\Pi(\vec{T}, \vec{S})\right) \neq \emptyset\right\} + \mathbb{P}\left\{C_1, \pi_{\mathcal{B}}\left(\Pi(\vec{T}, \vec{S})\right) \neq \emptyset\right\} + \mathbb{P}\left\{C_\infty, \pi_{\mathcal{B}}\left(\Pi(\vec{T}, \vec{S})\right) \neq \emptyset\right\}.
\end{aligned}$$

We conclude the proof by showing that each term is zero.

When the event C_0 is true (i.e., projection onto $\mathcal{B}$ of the trigger sets of the first samples is empty), we have that $\Pi(\vec{T}, \vec{S}) = \emptyset$ because $\pi_{\mathcal{B}}\left(\Pi\left(\vec{T}, \vec{S}\right)\right) \subseteq \pi_{\mathcal{B}}\left(\Pi\left((T_m)_{m=1}^{|\mathcal{B}|}, (S_m)_{m=1}^{|\mathcal{B}|}\right)\right)$. Thus the first term is zero.

When the event C_1 is true (i.e., projection onto $\mathcal{B}$ of the trigger sets of the first samples has one point), we have that $\pi_{\mathcal{B}}\left(\Pi\left((T_m)_{m=1}^{|\mathcal{B}|}, (S_m)_{m=1}^{|\mathcal{B}|}\right)\right) = \{z'_{\mathcal{B}}\}$ for some point $z'_{\mathcal{B}} \in \mathbb{R}^{|\mathcal{B}|}$. Therefore, the point $z'_{\mathcal{B}}$ is pinned down for the coordinates in $\mathcal{B}$ by the first samples $\{T_m\}_{m=1}^{|\mathcal{B}|}$ and $\Pi(\vec{T}, \vec{S})$ is non-empty iff $z'_{\mathcal{B}} \in \pi_{\mathcal{B}}\left(\zeta(T_{|\mathcal{B}|+1}, S_{|\mathcal{B}|+1})\right)$. For $S_{|\mathcal{B}|+1}$ we have that $\mathbb{P}\{z \in \zeta(T_{|\mathcal{B}|+1}, S_{|\mathcal{B}|+1})\} = 0$ for all $z \in \mathbb{R}^A$ because

$S_{|\mathcal{B}|+1} \in \mathcal{T}_a$. Because $S_{|\mathcal{B}|+1} \subseteq \mathcal{B}$ the coordinates not in $\mathcal{B}$ are only bounded from above, so by setting $z_a = \infty$ for $a \notin \mathcal{B}$ we get that $\mathbb{P}\{z_{\mathcal{B}} \in \pi_{\mathcal{B}}\left(\zeta(T_{|\mathcal{B}|+1}, S_{|\mathcal{B}|+1})\right)\} = 0$ for all $z_{\mathcal{B}} \in \mathbb{R}^{|\mathcal{B}|}$. Hence, by conditioning on the value of $z'_{\mathcal{B}}$ and using that samples are independent we get that the second term is zero.

When the event C_∞ is true (i.e., projection onto $\mathcal{B}$ of the trigger sets of the first samples has multiple points), for every sample there is some coordinate $a' \in \mathcal{B}$ such that $a' \notin S_m$ for all $m = 1, \ldots, |\mathcal{B}|$ because for the coordinates in S_m the values are pinned down by sample T_m. Consider the subset $\mathcal{B}' = \mathcal{B} \setminus \{a'\}$ that satisfies that $S_m \subseteq \mathcal{B}'$ for all $m = 1, \ldots, |\mathcal{B}|$. We have that $\Pi\left((T_m)_{m=1}^{|\mathcal{B}|}, (S_m)_{m=1}^{|\mathcal{B}|}\right) \subseteq \bigcap_{m=1}^{|\mathcal{B}'|+1} \bigcup_{S \in \mathcal{T}, S \subseteq \mathcal{B}'} \zeta(T_m, S)$ because $|\mathcal{B}'| = |\mathcal{B}| + 1$ and $S_m \subseteq \mathcal{B}'$. By the induction hypothesis we have that with probability one the latter is empty, implying that C_∞ is false with probability one. □

EC.6. Incorrect Assignments in the User Type Model

In §5 we introduced a user-type model with good-will penalties to accommodate the fact that advertisers have specific targeting criteria. If the contracts are feasible, that is, there is enough inventory to satisfy the targeting criteria; one would expect our policy to assign only impressions within the criteria. In this section we formalize the concept of a feasible operation, and give sufficient conditions under which the stochastic control policy does not assign any impressions outside of the targeting criteria.

It is straightforward to state the problem of determining whether contracts can be satisfied or not as a feasible flow problem on a bipartite graph. The problem can be formulated on a graph with one node for each user type T with a supply of $\pi(T)$, on the left side; one node for each advertisers $a \in \mathcal{A}_0$ with a demand ρ_a, on the right side; and one arc joining user type T with advertisers a if and only if $a \in T \cup \{0\}$. Then, we say that the operation is *feasible* if the user type-advertiser graph admits a feasible flow. That is, there exists flows $y_{T,a} \geq 0$ for $a \in \mathcal{A}_0$ and $T \in \mathcal{T}$ satisfying

$$\sum_{a \in T \cup \{0\}} y_{T,a} = \pi(T), \quad \forall T \in \mathcal{T}, \tag{EC.8}$$
$$\sum_{T \in \mathcal{T} : a \in T \cup \{0\}} y_{T,a} = \rho_a, \quad \forall a \in \mathcal{A}_0 .$$

The feasibility of the operation, albeit necessary, does not suffice to guarantee that no impressions outside the targeting criteria are assigned to the advertisers. When advertisers compete for the same type, and one of them obtains a potentially unbounded reward for that type; it may be optimal to allow the latter advertiser to cannibalize the user type, and force the others advertisers to take types outside of their criteria. This may occur, surprisingly, for all conceivable penalties. However, if qualities are bounded, and penalties are set high enough, then the optimal policy would not recommend the assignment of impressions outside the targeting criteria. Even in this case some impressions may be incorrectly assigned in the left-over regime, but the probability of this event decays exponentially fast. We formalize this discussion in the following proposition. We prove the result under the additional assumption that qualities are discrete and with finite support, though we conjecture the result holds for arbitrary distributions.

Proposition EC.6. *Assume that the revenue function is Lipschitz continuous with constant L, placement qualities are bounded by $\max_a Q_a \le \bar{Q}$ almost surely, and that penalties satisfy $\tau_a > A\bar{Q} + 2L/\gamma$ for all $a \in \mathcal{A}$. Assume, additionally, that qualities are discrete and with finite support. If the user type-advertiser graph admits a feasible flow, then the stochastic control policy does not assign any impressions outside of the targeting criteria, except perhaps for the left-over regime.*

Proof of Proposition EC.6. Given an optimal solution that assigns types outside of advertisers' targeting criteria, we will construct a deviation that achieves a strictly higher yield, thus contradicting the optimality of the solution. Some definitions are in order. Let $\Omega_T \subseteq \Omega$ be the outcome space associated with type $T \in \mathcal{T}$, which we assume to be a finite subset of $\mathbb{R}^A$. The probability of observing a placement quality $q = \{q_a\}_{a=1}^A \in \Omega_T$ given that the user type is T is denoted by $g_T(q)$, that is, $g_T(q) = \mathbb{P}\{Q = q \mid T\}$.

In this context a solution to the DAP can be written as a vector of functions $z_T : \Omega_T \to [0,1]$ and $y_{T,a} : \Omega_T \to [0,1]$ for $a \in \mathcal{A}_0$, where $z_T(q)$ gives the total probability that the impression is accepted by AdX and $y_{T,a}(q)$ the total probability that impression is assigned to contract a when the observed placement quality is q. The conditional probability that the impression is accepted in AdX given that the type is T and the quality is q is $z_T(q)/(\pi(T)g_T(q))$, and the expected AdX revenue given total probability z is

$$r_{T,q}(z) = \pi(T)g_T(q)r\left(\frac{z}{\pi(T)g_T(q)}\right).$$

The DAP can be written as the following concave maximum flow problem

$$\begin{aligned}
\max \sum_{T\in\mathcal{T}} \sum_{q\in\Omega_T} & r_{T,q}\left(z_T(q)\right) + \gamma \sum_{a\in T} q_a y_{T,a}(q) - \gamma \sum_{a\notin T} \tau_a y_{T,a}(q) \\
& z_T(q) + \sum_{a\in\mathcal{A}_0} y_{T,a}(q) = \pi(T)g_T(q), \quad \forall q \in \Omega_T, T \in \mathcal{T}, \\
& \sum_{T\in\mathcal{T}} \sum_{q\in\Omega_T} y_{T,a}(q) = \rho_a, \quad \forall a \in \mathcal{A}, \\
& \sum_{T\in\mathcal{T}} \sum_{q\in\Omega_T} z_T(q) + y_{T,0}(q) = \rho_0, \\
& z_T(q) \ge 0, y_{T,a}(q) \ge 0, \quad \forall a \in \mathcal{A}, q \in \Omega_T, T \in \mathcal{T}.
\end{aligned} \tag{EC.9}$$

The first term in the objective accounts for AdX's revenue while the second for the contracts' quality. The first constraints state that the flow of impressions originating from a given realization of quality is equal to the probability of that realization occurring, the second guarantee that contracts are fulfilled, and the third state flow conversation for impressions accepted by AdX together with those discarded.

We construct a solution (z^0, y^0) to (EC.9) based on a feasible flow $y^0_{a,T}$ from the user type-advertiser feasibility problem (EC.8) as follows. Put $z^0_T(q) = 0$ so that no impression is sold in the exchange, and put

$$y^0_{T,a}(q) = \begin{cases} g_T(q) y^0_{a,T}\,, & a \in T \cup \{0\}\,, \\ 0\,, & a \notin T \cup \{0\} \end{cases}$$

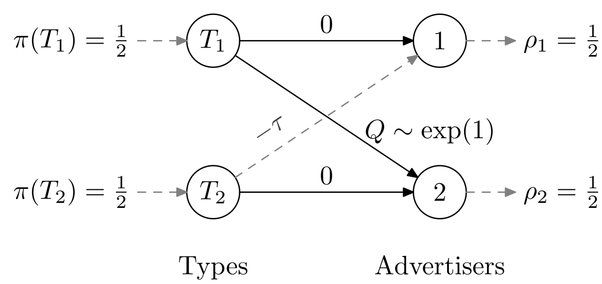


Figure EC.5 Example with two user types, and two advertisers.

for all type $T \in \mathcal{T}$ and advertiser $a \in \mathcal{A}_0$. It is straightforward to show that the latter solution is feasible for the discrete DAP in (EC.9).

Let (z, y) be an optimal solution of the discrete DAP in (EC.9) that assigns types outside of advertisers' targeting criteria. Take the difference $(\Delta z, \Delta y)$ with $\Delta z_T(q) = z^0_T(q) - z_T(q)$ and $\Delta y_{T,a}(q) = y^0_{T,a}(q) - y_{T,a}(q)$ which is a circulation in DAP. This circulation may have components of mixed signs. Because (z^0, i^0) has no incorrect assignments, if advertiser a is assigned a type $T \not\ni a$ not in her criteria, then the circulation verifies that $\Delta y_{T,a}(q) = -y_{T,a}(q)$. Hence, the arcs with incorrect assignments have negative flow.

By the Flow Decomposition Theorem the circulation $(\Delta z, \Delta y)$ can be decomposed into at most $(A + 2)\sum_{T\in\mathcal{T}} |\Omega_T|$ cycles with positive flow. Let $(\delta z, \delta y)$ be a cycle containing an arc assigning an impression outside the targeting criteria with flow $\epsilon > 0$. Because the right-hand side of the DAP graph has $A + 1$ nodes this cycle has at most $A + 1$ positive edges, at most $A + 1$ negative edges, and at most 2 AdX edges with opposite sign. Let $\delta y_{a_1^+, T_1^+}(q_1^+), \ldots, \delta y_{a_K^+, T_K^+}(q_K^+), \delta z_{T^+}(q^+) = \epsilon$ be the positive edges and $\delta y_{a_1^-, T_1^-}(q_1^-), \ldots, \delta y_{a_K^-, T_K^-}(q_K^-), \delta z_{T^-}(q^-) = -\epsilon$ be the negative edges. Without loss of generality we let $y_{a_1^-, T_1^-}(q_1^-)$ be an arc with incorrect assignment, that is, $a_1^- \notin T_1^-$.

Consider the perturbed solution $(z + \delta z, y + \delta y)$, which is feasible for the DAP. Using the fact that incorrect assignments have negative flow we can lower bound the marginal yield by

$$
\begin{aligned}
J^D(z + \delta z, y + \delta y) - J^D(z, y) &= \left[r_{T^+, q^+}\left(z_{T^+}(q^+) + \epsilon \right) - r_{T^+, q^+}\left(z_{T^+}(q^+) \right) \right] + \gamma \sum_{\substack{\text{positive}\\ \text{edges}}} \epsilon q_{a_k^+} \\
&\quad + \left[r_{T^-, q^-}\left(z_{T^-}(q^-) - \epsilon \right) - r_{T^-, q^-}\left(z_{T^-}(q^-) \right) \right] - \gamma \sum_{\substack{\text{negative}\\ \text{edges}}} \epsilon \begin{cases} q_{a_k^-}, & a_k^- \in T_k^-, \\ -\tau_{a_k^-}, & a_k^- \notin T_k^- \end{cases} \\
&\geq -\epsilon L + 0 - \epsilon L + \gamma \epsilon \tau_{a_1^-} - \gamma A \epsilon \bar{Q} = \epsilon \gamma (\tau_{a_1^-} - A\bar{Q} - 2L/\gamma) > 0,
\end{aligned}
$$

where the first inequality follows from discarding the contract assignments with positive flow that contribute positively to the yield, the Lipschitz continuity of the revenue function, and that qualities are bounded from above by $\bar{Q}$. Thus, the solution $(z + \delta z, y + \delta y)$ achieves higher yield in the DAP, contradicting the optimality of (z, y). $\square$

Note that since the left-over regime is vanishingly small in proportion to the length of the horizon this implies that the number of unassigned impressions is small. Thus in practice, a publisher may set $C'_a = C_a + \epsilon$, discard any impressions assigned by the policy outside the targeting criteria, and ensure that contracts are filled properly.

Next, we prove by example that the requirement that qualities are bounded is necessary for the previous result to hold. Consider a publisher who contracts with two advertisers, and agrees to deliver one half of the arriving impressions to each one of them. Additionally, there are two impression types, denoted by T_1 and T_2, each occurring 50% of the time. The first advertiser only cares about the first type. She obtains a reward of zero for T_1, and the advertisers pays a positive penalty τ each time a T_2 impression is assigned to her. The second advertisers admits both types, but only obtains a positive reward $Q \sim \exp(1)$ for the first type. The setup is shown in Figure EC.5.

A feasible policy could assign all T_1 impressions to the first advertiser, and T_2 impressions to the second advertiser. However, such policy is not optimal. Notice that both advertisers compete for the T_1 impressions, and the first advertiser could extract a potentially high quality from them. It is not hard to see that the optimal dual variables are $v_1 = -\tau$, and $v_2 = 0$; and the optimal objective value is $\frac{1}{2}\mathbb{E}[Q-\tau]^+ = \frac{1}{2}e^{-\tau}$. Hence, it is optimal to assign those T_1 impressions with quality greater than τ to the second advertiser. Thus, no matter the value of the penalty, a fraction $e^{-\tau}$ of the total impression assigned to the first advertiser are undesired.

EC.7. Computation

In this section we describe to compute the optimal policy for our data model. The main problem resides in the computation of the dual objective in (4) and its gradient given a vector of dual variables.

Objective. The first term of the objective can be written as

$$\begin{aligned}
\mathbb{E}R\left(\max_{a\in\mathcal{A}_0}\{Q_a - v_a\}\right) &= \sum_{\forall T}\pi(T)\mathbb{E}\left[R\left(\max_{a\in\mathcal{A}_0}\{Q_a - v_a\}\right) \mid T\right] \\
&= \sum_{\forall T}\pi(T)\sum_{a\in T\cup T^c}\mathbb{E}\left[R\left(Q_a - v_a\right)\mathbf{1}\{Q_a - v_a \geq Q_{a'} - v_{a'}\ \forall a' \neq a\} \mid T\right] \\
&= \sum_{\forall T}\pi(T)\left(I_{T,0}(v) + \sum_{a\in T} I_{T,a}(v)\right)
\end{aligned}$$

where the first equation follows by conditioning on the type, and the second because the events are a partition of the sample space. Next, we show to compute the expectations $I_{T,a}(v)$.

Let $M_T(v) = \max_{a\in\mathcal{A}_0\setminus T}\{-\tau_a - v_a\}$ be the maximum contract adjusted quality of the advertisers (including the outside option) that are not in the type, and $\alpha_T(v)$ the set of advertisers that verify the maximum. Then, we have that

$$\begin{aligned}
I_{T,0}(v) &= R\left(M_T(v)\right)\mathbb{P}\{Q_a - v_a \leq M_T(v)\ \forall a' \in T\} \\
&= R\left(M_T(v)\right)G_T(M_T(v) + v_T),
\end{aligned}$$

where $G_T(\cdot)$ is the c.d.f. of Q_T, and v_T is the vector of dual variables for the advertisers in the type.

For $a \in T$, we compute the expectation by conditioning on the continuous random variable Q_a. Further, suppose that we partition the mean vector and covariance matrix in a corresponding manner. That is, $\mu_T = \left(\begin{smallmatrix}\mu_a \\ \mu_{-a}\end{smallmatrix}\right)$, and $\Sigma_T = \left(\begin{smallmatrix}\Sigma_{a,a} & \Sigma_{a,-a} \\ \Sigma_{-a,a} & \Sigma_{-a,-a}\end{smallmatrix}\right)$. For instance, μ_{-a} gives the means for the variables in $T \setminus \{a\}$, and $\Sigma_{-a,-a}$ gives variances and covariances for the same variables. The matrix $\Sigma_{-a,a}$ gives covariances between variables in $T \setminus \{a\}$ and a (as does matrix $\Sigma_{a,-a}$). Because the marginal distribution of a multivariate normal is an univariate normal, we have that $Q_a \sim \ln\mathcal{N}(\mu_a, \Sigma_{a,a})$. We denote by $g_{T,a}(\cdot)$ the p.d.f. of Q_a. Similarly, let Q_{-a} be the vector of qualities for advertisers in $T \setminus \{a\}$. Conditioning on $Q_a = q_a$, the distribution of Q_{-a} is log-normal with mean vector $\mu_{-a} - \Sigma_{-a,a}(q_a - \mu_a)/(\Sigma_{a,a})$, and covariance matrix $\Sigma_{-a,-a} - (\Sigma_{-a,a}\Sigma_{a,-a})/(\Sigma_{a,a})$. We denote its c.d.f. by $G_{T,-a}(\cdot)$. Putting all together, we have that

$$\begin{aligned} I_{T,a}(v) &= \mathbb{E}\left[R\left(Q_a - v_a\right)\mathbb{P}\{Q_{a'} - v_{a'} \le Q_a - v_a \ \forall a' \neq a \mid Q_a\} \mid T\right] \\ &= \int_{v_a + M_T(v)}^{\infty} R(q_a - v_a) G_{T,-a}(q_a - v_a + v_{-a}) g_{T,a}(q_a) \, \mathrm{d}q_a, \end{aligned}$$

where v_{-a} is the vector of dual variables for advertisers in $T \setminus \{a\}$.

Gradient. The forward derivative of the dual objective can be written as

$$\begin{aligned} \nabla_a \psi(v) &= -\mathbb{P}_R\left\{Q_a - v_a > \max_{a \in \mathcal{A}_0 \setminus a}\{Q_{a'} - v_a'\}\right\} + \rho_a \\ &= -\sum_{\forall T} \pi(T) \mathbb{E}\left[\left(1 - s^*(Q_a - v_a)\right) \mathbf{1}\left\{Q_a - v_a > \max_{a \in \mathcal{A}_0 \setminus a}\{Q_{a'} - v_a'\}\right\} \mid T\right] + \rho_a \\ &= -\sum_{T : a \in T} \pi(T) P_{T,a}(v) - \sum_{\substack{T : a \notin T \\ a \in \alpha_T(v), |\alpha_T(v)| = 1}} \pi(T) P_{T,a}(v) + \rho_a, \end{aligned}$$

where the contributing types for the forward derivative are those where a is in, and those where a is not in but verifies exclusively the maximum of the types not in $(M_T(v))$. If two or more advertisers verify the maximum $M_T(v)$, then increasing v_a does not have an impact of the type's contribution to the objective. When $a \notin T$, the expectation is given by

$$P_{T,a}(v) = \big(1 - s^*(M_T(v))\big) G_T(M_T(v) + v_T).$$

Similarly to the objective, when $a \in T$ we have that

$$P_{T,a}(v) = \int_{v_a + M_T(v)}^{\infty} \big(1 - s^*(q_a - v_a)\big) G_{T,-a}(q_a - v_a + v_{-a}) g_{T,a}(q_a) \, \mathrm{d}q_a.$$

The backward derivative is computed in a similar fashion. The only exception is that, when $a \notin T$, and a verifies the maximum $M_T(v)$,the advertiser always contributes to the derivative regardless of the number of advertisers that attain the maximum. Hence,

$$\nabla_{-a} \psi(v) = \sum_{T : a \in T} \pi(T) P_{T,a}(v) + \sum_{\substack{T : a \notin T \\ a \in \alpha_T(v)}} \pi(T) P_{T,a}(v) - \rho_a.$$

Ties. For the following, we assume that the instance is not *degenerate*, that is, the variances within the types are positive, and no two advertisers are perfectly correlated. Then, within each type, non-trivial ties can only occur between the advertisers that are not in the type (we refer to the non-trivial ties as those in which multiple advertisers attain the same contract adjusted quality). Moreover, there can be at most one tie within each type, and this happens when the maximum $M_T(v)$ is verified by many advertisers, that is $|\alpha_T(v)| > 1$. With some abuse of notation, the probability of such a tie is given by $\pi(T)P_{T,\alpha_T(v)}$ and it should be split among the advertisers $\alpha_T(v)$. Note that the number of non-trivial ties is $O(T)$, and the tie-breaking rule can be computed efficiently by solving a feasible flow problem.

References

Agrawal, S., Z. Wang, Y. Ye. 2009. A dynamic near-optimal algorithm for online linear programming. Working paper posted at http://www.stanford.edu/ yyye/.

Ahuja, Ravindra K., Thomas L. Magnanti, James B. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc.

Aven, Terje. 1985. Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables. *Journal of Applied Probability* **22**(3) 723–728.

Besbes, Omar, Assaf Zeevi. 2009. Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research* **57**(6) 1407–1420.

Broder, Josef, Paat Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* **60**(4) 965–980.

Devanur, Nikhil R., Thomas P. Hayes. 2009. The adwords problem: online keyword matching with budgeted bidders under random permutations. *Proceedings of the 10th ACM conference on Electronic commerce*. EC '09, ACM, 71–78.

Feldman, Jon, Monika Henzinger, Nitish Korula, Vahab S. Mirrokni, Cliff Stein. 2010. Online stochastic packing applied to display ad allocation. *Proceedings of the 18th annual European conference on Algorithms: Part I*. ESA'10, Springer-Verlag, 182–194.

Laffont, Jean-Jacques, Eric Maskin. 1980. Optimal reservation price in the vickrey auction. *Economics Letters* **6**(4) 309 – 313.

Mendelson, Shahar. 2003. A few notes on statistical learning theory. Shahar Mendelson, Alexander J. Smola, eds., *Advanced Lectures on Machine Learning*, *Lecture Notes in Computer Science*, vol. 2600. Springer Berlin Heidelberg, 1–40.

Myerson, R. 1981. Optimal auction design. *Mathematics of Operations Research* **6**(1) 58–73.

Shapiro, A., D. Dentcheva, A. Ruszczynski. 2009. *Lectures on Stochastic Programming: Modeling and Theory.*. MOS-SIAM Series on Optimization 9, Society for Industrial and Applied Mathematics (SIAM).

Vee, Erik, Sergei Vassilvitskii, Jayavel Shanmugasundaram. 2010. Optimal online assignment with forecasts. *Proceedings of the 11th ACM conference on Electronic commerce*. EC '10, ACM, 109–118.