AdX Final Project Write Up: An Attempt to Find True Values in a Multi-Objective Environment

Nikhil Das [1] **Camilo Becerra** [1]

Abstract

In the Ad Exchange game, agents participate in daily auctions for impression opportunities and for campaign procurement. Although truthful bidding is DSIC in a second-price auction, determining an agent's *true value* is complicated by the multi-objective nature of the game: agents try to jointly maximize profit and maintain a quality score that depends on an effective reach function. In this paper we analyze only the impression opportunity auction. We construct a simplified version of the game that is analytically tractable, model the behavior of other agents, propose an interpretable true-value function, and derive the corresponding equilibrium bidding strategy.

1. Introduction

The AdX game consists of repeated interaction between ad networks attempting to both earn profit and to maintain a high enough quality score to secure future campaigns. One of the two core mechanisms in this environment is the **Impression Opportunities Auction**: a single shot second-price auction for an impression.

From class, we know truthful bidding is a DSIC strategy in second-price auctions. Therefore, the main theoretical challenge in this isolated component of AdX is *not* how to bid strategically, but how to calculate a meaningful *true value* for an impression when both short-run profit and long-term quality score matter. Because each impression simultaneously contributes in a nonlinear way to both objectives, its value is inherently state dependent.

This raises our guiding question:

> *How should an ad network compute its true value for impressions in a multi-objective environment where both profit and quality score matter?*

We attempt to answer this question by constructing a simplified model of the impression auction, proposing a value function, analyzing the behaviors under which such function must exist, and evaluating the limitations of our approach.

2. A Simplified Game & Competitor Model

To make this question more analytically tractable, we introduce the following simplifications while trying to preserve the strategic essence of the actual game.

2.1. Simplifying Assumptions

1. **One-Day, One-Campaign Environment.** The agent manages only a single, one-day campaign with target reach R and current effective reach $\rho(C)$.
2. **Quality Score Exists.** Even though it is a one-day campaign, we assume there already exists a Q score.
3. **i.i.d. Impression Opportunities.** Eligible impressions arrive independently with probability q.
4. **Competitor Behavior.** Competing bidders draw from a fixed price distribution:
$$b_{-i} \sim F(p).$$
5. **Continuous Bid Timing** Rather than a bid being sent at the beginning of the day, assume bids are sent continuously throughout the day.

This simplified environment attempts to capture opponent uncertainty, state dependence in valuation, and the second-price mechanism while allowing us to derive a closed-form equilibrium strategy.

3. Game Structure

For each campaign C with budget B and target reach R, the game computes an effective reach score:

$$\rho(C) = \frac{2}{a}\left(\arctan\left(a\left(\frac{x}{R}\right) - b\right) - \arctan(-b)\right)$$

where x is the number of delivered impressions, $a = 4.08577$, and $b = 3.08577$. Profit is

$$\text{Profit} = \rho(C)B - K$$

where K is total spend.

Because each impression affects the sigmoidal effective reach function, its marginal contribution changes throughout the campaign life cycle. Therefore, its true value must reflect nonlinear and state-dependent marginal benefits.

4. A Proposed True Value Function

Let d denote an impression's demographic group and let C be a campaign. We define the true value of an impression as

$$v_i(d) = \mathbb{1}_{\{d \in C\}} \left[\Phi(\rho(C)) \frac{B}{R} \right]. \tag{1}$$

Here:

- $\frac{B}{R}$ is the natural base value per target impression.
- $\Phi(\rho(C))$ scales this base value according to where the campaign lies on the effective reach curve.

Intuitively, $\Phi(\rho)$ captures the "importance weight" of impressions as reach accumulates.

4.1. Behavioral Requirements for Φ

Rather than committing to a particular definition of Φ, we propose three key behavioral requirements that meaningfully restrict the function space which we hope will provide a more contained path for future exploration.

1. Budget Feasibility. Expected total value assigned to impressions cannot exceed the campaign budget:

$$\mathbb{E}[\text{Total Spend}] = \mathbb{E}\big[\Phi(\rho(C))\big] \cdot \frac{B}{R} \cdot R \leq B.$$

This ensures that truthful bidding in expectation does not exceed the available budget. Although this behavior may seem trivial and straightforward, it plays a critical role in ensuring our proposed value function is compatible with the incentive structure of the game.

Because truthful bidding is only optimal when the agent can afford the bid implied by its true value, this condition effectively constrains the scale of Φ and prevents unreasonable valuations that would exhaust the budget.

This inequality further narrows our search for Φ by providing an upper bound on its integral.

2. Threshold Behavior. There should exist a critical point $x^* \in [0, R]$ at which marginal value of impression peaks. Formally, we would want a function that

$$\Phi'(\rho) = 0 \quad \text{when } x = x^*$$

and

$$\Phi'(\rho) \leq 0 \quad \text{when } x > x^*.$$

This captures that marginal impressions are most valuable near the point where the effective reach function has the greatest slope.

We intentionally do not impose a specific sign on $\Phi'(\rho)$ for $x < x^*$ since multiple reasonable behavior patterns could emerge depending on the agent's strategic priorities:

1. $\Phi'(\rho) > 0$: the agent gradually increases valuation as the campaign gains traction.
2. $\Phi'(\rho) = 0$: the agent treats early impressions uniformly until it hits the critical region.
3. $\Phi'(\rho) < 0$: the agent front loads value to accelerate progress towards x^*.

One potential function that satisfies the first approach is a logistic function centered at x^*.

To find such a x^*, we explore the effective reach function.

We want to find at what percentage of total completion, x/R, the slope is maximized, meaning marginal impression value is greatest. To do this, we examine

$$\rho(C) = \frac{2}{a} \left(\arctan\left(a \left(\frac{x}{R} \right) - b \right) - \arctan(-b) \right).$$

Observe that the term $-\arctan(-b)$ just shifts the graph vertically, so it does not affect the slope.

Only $\frac{2}{a}(\arctan(a(\frac{x}{R}) - b))$ matters for the derivative. Now, take the derivative of the important component to find:

$$\frac{d}{d\frac{x}{R}} \left[\frac{2}{a} \left(\arctan\left(a\Big(\frac{x}{R} \Big) - b \right) \right) \right] = \frac{2}{1 + \big(a(\frac{x}{R}) - b\big)^2}.$$

To maximize this slope, we want to minimize the denominator since the numerator is fixed. This happens when

$$\left(a \left(\frac{x}{R} \right) - b \right)^2 = 0$$

Thus, when $a(x/R) - b = 0$.

Solving this equation we get

$$\frac{x}{R} = \frac{b}{a} = \frac{3.08577}{4.08577} \approx 0.755.$$

Therefore, we find the maximum marginal value occurs at approximately 75.5% of reach. Thus, we know such

$$x^* \approx 0.755 \cdot R$$

3. Dependence on Quality Score. Because the marginal value of effective reach depends on an agent's current quality score Q, we extend Φ to incorporate this state variable:

$$\Phi(\rho(C); Q) = \varphi(Q)\, \Phi(\rho(C)), \qquad \varphi'(Q) < 0.$$

The term $\varphi(Q)$ captures the hidden value of the quality score: when Q is low, additional reach has disproportionately high strategic value because it raises the probability of securing future campaigns. Conversely, when Q is high, the marginal benefit of further improvement diminishes and the agent should shift toward short-run profit maximization.

Therefore, $\varphi(Q)$ aims to quantify the central tradeoff of the game (profit vs future competitiveness) and introduces the necessary state dependence to our true value function. This ensures that otherwise identical impressions may rationally receive different valuations depending on the agent's current market position.

5. Impacts of Campaign Procurement Auction

Although our analysis focuses on determining a true value for a simplified version of impression opportunities, the campaign procurement auction also influences this valuation through its effect on the agent's budget B.

It is important to note that in the complete AdX environment, B is not a completely exogenous factor. For campaigns won through procurement, it reflects how aggressively the agent bid and therefore encodes expectations about future impression prices.

Under our assumption that competitor bids in the impression auction are drawn from a known distribution $b_{-i} \sim F(p)$, the agent can form an estimate of the expected winning price:

$$\mathbb{E}[p \mid \text{win}] = \int_0^{v_i} p f(p)\, dp,$$

In certain parametric cases, such as assuming F is normal, this expectation can be computed explicitly in polynomial time. This connects procurement and impression bidding: a campaign with target reach R implicitly requires a budget roughly consistent with

$$B \approx R \cdot \mathbb{E}[p \mid \text{win}].$$

Thus, we can see how the procurement auction constrains the feasible scale of our proposed true value function by directly having an impact on our presumed base value of B/R.

While for space constraints we do not model the procurement campaign fully, it is important to acknowledge this link to underscore the new complexity that the campaign procurement auction introduces. We see explicitly that impression true values do solely depend on state variables (ρ, Q) at the time of the impression auction, but also on market price expectations formed upstream.

Under truthful bidding, our proposed value function implies an expected spend of roughly $\mathbb{E}[\Phi(\rho(C))] \cdot B$. Combined with the procurement auction relation above $B \approx R \cdot \mathbb{E}[p \mid \text{win}]$, this connects our per-impression value to market price expectations and imposes a consistency condition between procurement bids and impression bids. Essentially, the scale of v_i must align with the implicit price forecasts encoded in B, ensuring that procurement and impression bidding form a coherent overall strategy.

6. Insights

We presented a tractable theoretical approach for deriving true values in the impression auction of the AdX game. By simplifying the environment, modeling opponent behavior, and analyzing marginal contributions to both profit and quality score, we derive a clean equilibrium bidding rule:

$$\boxed{b_i^* = v_i = \Phi(\rho; Q)\frac{B}{R}.}$$

Our analysis reveals several important insights about the structure of value in the AdX auction. First, the notion of a *true value* is fundamentally dynamic: because each impression shifts both effective reach and quality score, the agent's valuation is properly understood as a state-dependent function rather than a fixed constant.

Second, the effective reach function embeds a universal turning point at approximately 75.5% campaign completion, meaning agents should concentrate their budget around this peak marginal value region rather than pacing spending uniformly.

Third, quality score creates an intertemporal externality, making some impressions valuable not for immediate profit but for improving future competitiveness in procurement auctions.

Finally, by simplifying the environment, we have identified a small set of sufficient statistics - ρ, Q, B, R - that, in our simplified setting, fully determine optimal bids. This highlights which components of the full game may be strategically essential.

These insights collectively clarify how multi-objective constraints can shape our equilibrium bidding behavior in both this theoretical simplified version, and potentially in the actual game.

7. Limitations

7.1. Theoretical Limitations

Although Equation (1) provides a structured approach, several theoretical challenges remain:

- The model does not capture multi-day dynamics or compounding effects. One example of a compounding effect not explored is the probability of seeing a par-

ticular demographic impression given what you have seen previously in the day.

- In a multi-day, multi-campaign setting, the auction environment is no longer DSIC. This is because multi-day, multi-campaign, auctions are not independent as B and other factors carry over.
- The interaction between campaign procurement and impression bidding forms a feedback loop that requires further theoretical investigation. While we briefly discussed how campaign procurement impacts impression opportunities, the same is true the other way as impression opportunities set historical precedence for what such a distribution may be.
- Distributions $F(p)$ are likely unknown or constantly changing. This is more probable if agent's are not aware of the other campaigns players have as some demographics may be more competitive than others.

7.2. Application Limitations

- Estimating a sufficiently accurate Φ empirically would require extensive simulation and/or historical data.
- The model includes no explicit uncertainty structure for impression prices or future campaign availability.
- A single scalar Φ may be insufficient to represent complex market conditions, even when augmented with φ.

Despite these limitations, this model provides a useful first step toward a principled theory of true-value reasoning.

Moreover, although truthful bidding is not DSIC in the complete game, accurately characterizing an agent's underlying value remains fundamental, as equilibrium bidding behavior — including optimal shading or inflation — must ultimately be derived from this value.

8. Looking Forward

With more time, we would like to explore the following:

- Develop dynamic programming or more advanced reinforcement learning models to approximate optimal multi-day strategies.
- Learn to tractably compute Φ from empirical data or simulation-based inference.
- Introduce explicit stochastic models for impression prices and market competitor behavior.
- Extend the value function to multi-day, multi-campaign optimization with budget constraints.

Although the modeling challenges are quite substantial, our framework offers a principled starting point for analyzing truthful bidding strategies in this complex, multi-objective ad marketplace.

9. Our Model

9.1. Ideas Explored

Across the past month, we implemented and tested two core approaches to deriving Φ: a non-RL hand-tuned heuristic model and an Deep Recurrent Q-Network (DRQN) approach.

9.1.1. Non-RL Agent, Hand Tuned

Our first implementation operationalized the conceptual role of Φ by mapping key campaign state variables - effective reach, remaining budget, days left, demographic overlap - into simple bidding heuristics that approximate marginal value.

Impression Level Bidding. For each active campaign, the agent computed: current completion x/R, effective reach $\rho(C)$, remaining budget $B - K$, and remaining days. We then partitioned the space into three phases and assigned a multiplicative factor to the base value B/R. These multipliers served a rough approximations of Φ, increasing bids when impressions were most valuable and reducing them as the campaign passes x^*.

Campaign Procurement. For procurement auctions, the agent generated bids based on campaign duration, existing workload, and demographic overlap with active campaigns. Bids were scaled down when the agent was capacity-constrained or when a new campaign competed heavily with current demographic segments.

Faults & Learnings. While interpretable and easy to tune, these state-based multipliers and spending rules were ultimately a hand-crafted stand-in for Φ. This approach helped us to test the qualitative predictions of our theory but could not capture the full multi-day, multi-campaign, dynamic value structure. These insights motivated us to move toward a learned approximation of Φ in our next implementation.

9.1.2. DRQN Agent

Inspiration. We considered a DRQN model as a second approach, inspired by the papers introducing the subject to the context of RL agents. We first considered what was described in Playing Atari with Deep Reinforcement Learning by Mnih et al. However, we soon transitioned past this to apply the modifications described in Deep Recurrent Q-Learning for Partially Observable MDPs by Hausknecht and

Stone, and considering the implementation by Bhowmik. Considering the relevance of previous days on future betting, we thought that implementing a Recurrent Neural Network, specifically a Long Short-Term Memory (LSTM) would be logical. Further, we removed the Convolutional Neural Network layers, as spatial information would be irrelevant for our problem. This led us to a model that included a LSTM, followed by a sequential, Multi-Layer Perceptron (MLP).

Impression Opportunities. We framed the problem as asking the model to calculate the Q-values for some range of actions, where each would be a percentage of the budget which to spend bidding on all subsets of a campaign (including itself). The limits were hard-coded to be the full budget divided over the days the campaign lasted and the number of segments bid on. For Q-learning, the states were defined as

$$\begin{bmatrix} \text{campaign progress} \\ \text{effective reach} \\ \text{campaign reach} \\ \text{campaign budget} \\ \text{accumulated cost} \\ \text{demographic} \\ \text{length of campaign} \end{bmatrix}$$

With this definition, when the call to get_ad_bids() happened, the model would:

1. Update values in memory from past days
2. Retrain NN (if in training mode) to incorporate latest data
3. Use RNN+MLP to calculate q-values based on the state per campaign
4. Choose actions with an ϵ-greedy algorithm, exploring new actions with ϵ probability

For training, the model kept a memory buffer that stored campaigns, and would sample mini-batches randomly from this buffer. Once the buffer reached capacity, the oldest campaigns would be removed.
Finally, the definition of reward was set as a product of the change in effective reach multiplied by the percentage of budget left. That way, the reward represents the goal of the agent: to complete the campaign as cheaply as possible.

Campaign Procurement. For campaign procurement, the agent would calculate a multiplier for each campaign for auction based on if it overlaps with any owned campaign, its length, and how specific of a segment it demands for. This multiplier begins at 1 and is modified either positively (if something is not desirable), or negatively (if something is desirable). This value is then multiplied by the reach of the campaign, which is submitted as the bid.

9.2. Parameters

To support the algorithm, a variety of parameters were given, most notably:

1. Neural network parameters: # of layers, # neurons per layer, # of RNN layers, # of possible actions.
2. Training parameters: Number of cycles per training step, batch size, impression memory size, ϵ, γ.

9.3. Limitations

This agent was limited by the common limitations of deep learning models. The lack of computing power of running the model locally caused it to be less complex and train for less iterations than a more robust model that could be trained on greater compute. Also, the lack of data caused the model to learn less efficiently. As campaigns would usually not last very long, the effectiveness of learning on sequential data would be limited by the minimal length of the sequences.

9.4. Assumptions & Relaxing Potential

The greatest assumptions of this agent are the abstractions made to define state, reward, and action, whose details were explained previously. There are also simplifications made to the campaign procurement algorithm, where the value of future campaigns is determined by the current price of owned campaigns, limiting in both the amount of information about historical pricing, and biased towards only the agent's past beliefs about campaign pricing.

9.5. Correctness Argument

This agent is based on an assumption on how rewards should be modeled. However, once that is defined, the agent takes an RL/DL approach to maximize this value, training and learning how to do so. Therefore, as the agent begins to win after training, it can be said that it has effectively completed this goal of correctly approaching the game, up to the limitation of its worldview about rewards. Further, both state and reward representations could be modified, either making them more or less complex. By relaxing these assumptions (namely simplifying these representations), we could have a model that is more effective in reasoning about the game's noisy environment, but possibly less correct in its calculations of q-values.

9.6. Complexity Analysis

With regards to space complexity, this agent takes constant space. With regards to time complexity, this agent takes $O(n + c)$ where n is the number of training steps, and c is the number of owned campaigns.