Real-Time Bidding rules of thumb: analytically optimizing the programmatic buying of ad-inventory

ARTICLE · AUGUST 2015

1 AUTHOR:

Joaquin FT
Tradelab

Real-Time Bidding rules of thumb: analytically optimizing the programmatic buying of ad-inventory

Joaquin Fernandez-Tapia*

July 16, 2015

Abstract

This study focuses on the optimization aspects of ad-inventory buying through real-time bidding, by taking as starting point the fundamental mathematical relationships between the different variables involved. We take as point of view an algorithm aiming to purchase a maximum amount of ad-inventory under a budget constraint and restricted to the situation where user-segments and contexts are homogeneous (in terms of its conversion rate). In particular we focus on three tactical issues: (1) comparative statics relating the different macroscopic variables important for the operation of an advertising campaign (2) optimal bidding across different aggregates of inventory and (3) optimal budget pacing. Based on an statistical model of the interaction with ad-exchanges in the case of second-price auctions (and its variations), which lead to a straightforward derivation of the relationship with macroscopic performance indicators, we show that the optimization problem is analytically tractable just by considering a first principles analysis, without the need of blindly rushing into extensive data mining. This provides a baseline framework permitting to obtain explicit mathematical relations between the different macroscopic variables, allowing a deeper understanding of the solutions and a straightforward extension as we weaken the homogeneity hypothesis and pinpoint where data can leverage value, which set a clearer ground for further data mining methods.

Keywords: real-time bidding, constrained optimization, second-price auctions, budget pacing, variational calculus

1 Introduction

1.1 Industrial context

Real-Time Bidding (or RTB) is a new paradigm in the way online advertising is purchased: media-trading-desks (in behalf of advertisers or their advertising agency) can buy online inventory (display a banner) through real-time auctions, allowing to target individual users in a per-access basis by taking as decision variables: data about the user, features of the website, time of the day, budget constraints, goals of the campaign etc. Today RTB represents around a 30% of the market share for display advertising [7].

*R&D Engineer and Data Scientist at Tradelab.
73 Rue d'Anjou, 75008, Paris, France.
E-mail: jf@tradelab.fr

In a nutshell, each time a user have access to a website, publishers connect with a virtual marketplace, called *ad-exchange*, in order to trigger an auction for each available slot that can be allocated to advertising. As a counterpart, media trading desks receive auction requests (plus some information about the user and the website) in order to choose the bid level that best suits their strategy. The highest bid wins the auction and the price paid depends on the type of auction. This entire process, from the user entering to the website to the display of the banner, takes less that 100 milliseconds [12].

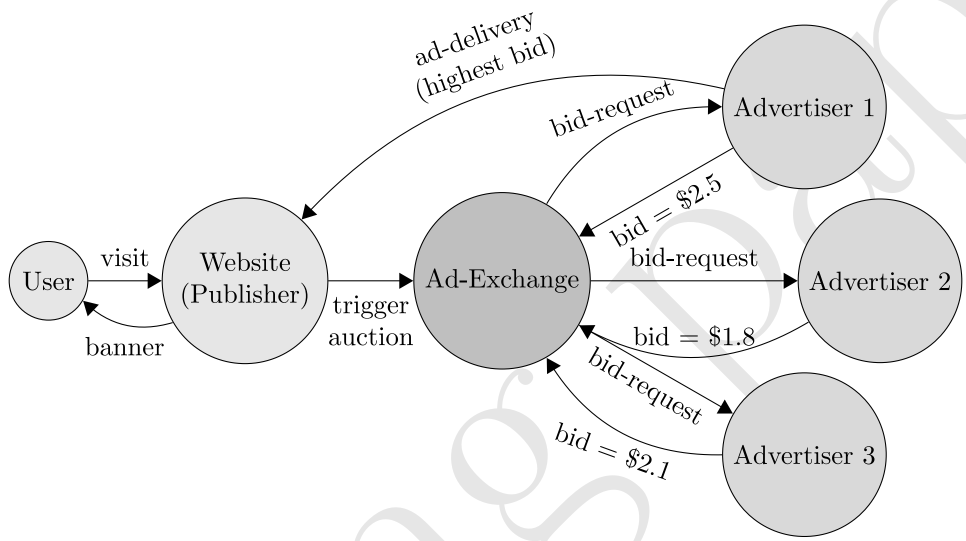

Figure 1: Diagram of how a Real-Time Bidding process works

From an advertiser perspective, the bottom line of this technology is to leverage its ROI (return on investment), this can be achieved via branding campaigns or by turning users, potentially in affinity with a campaign, into final conversions. In that view, media-trading desks track different metrics, expected to be lead indicators of the success of a campaign. Among these metrics, we can mention the CPM (i.e. *cost-per-mille*: price paid per thousand impressions), the CTR (i.e. *click-through-rate*: ratio clicks/impressions) and the CPA (i.e. *cost-per-acquisition*: media cost per unit of conversion).

1.2 Modeling across scales

From an quantitative standpoint, the problem can bee seen from three different perspectives. First, the optimization of the bidding process on a per-auction basis (i.e. how to interact with the exchanges taking in account constraints, campaign objectives and the auction mechanism). Second, the strategic issues related to budget allocation across different contexts in terms of the campaign's performance (considering the interaction with exchanges as a black-box). Third, the analysis of the marketing funnel on a per-user basis (i.e. the journey from the first banner to the final conversion for a given user).

Hence, it is important to understand that different scales are important to take in account when tackling the optimization of the bidding process, i.e. even though the bidding process arises on a *microscopic* scale (low-latency/per-auction basis, thousands of time a day), the performances, on the other hand, are measured at larger scales, i.e. an intraday scale (*mesoscopic*), taking in account the aggregate behavior of the system over few minutes/hours, to a *macroscopic* scale, taking in account the overall performance of the advertising campaign over several days, weeks or even months.

Macroscopic Scale (Campaign)	Mesoscopic Scale (Minutes, Hours)	Microscopic Scale (Auction)
Spent	Average Bid	Bid (Per-auction)
CPM	Average Price	User Data
CPA	Winrate	Context Data

Figure 2: Key variables depending on the time-horizon we consider.

1.3 The *homogeneous inventory* hypothesis

In order to solve the optimal bidding problem in a per-auction basis taking in account macroscopic performance goals, a powerful approach is to focus on the situation where we display advertising on a homogeneous segment of users/contexts[1]. By homogeneous, we refer to the fact that the probability to turn an impression into a conversion is the same across the different auction-request. Although this hypothesis seems an oversimplification, it is in fact consistent with real applications, considering that (with the exception of *retargeting* campaigns, where users are on an evolved stage of the marketing funnel and so, there is a strong signal of engagement between the user and the campaign) most of the time there is little information justifying to price a user differently from another just by considering real-time data. The same argument counts for the contexts (websites, domains, placements) we consider: the fact of considering a homogeneous conversion rate can be interpreted as having a external algorithm focusing on the optimal budget allocation (for example a bandit algorithm taking in account conversion rates) then consider our approach here as the tactical problem of how to spend that budget optimally on each one of these aggregates of inventory. From a practical perspective, our approach is specially consistent where the focus of the campaign is in prospecting new users from a broad homogeneous population, as well as on the optimization of branding campaigns. Moreover, it allows to naturally untangle *strategic* issues (budget allocation across contexts with different conversion rates, segmentation of users etc) with *tactical* issues (interaction with ad-exchanges, i.e. the bidding process on homogeneous segment/context).

1.4 Optimization problems

In a homogeneous setting, three tactical problems arise when dealing with inventory buying through real-time bidding.

1. Comparative statics: How to quantify the effect on performance indicators when varying global parameters as the total budget or the conversion rate.

[1]For example, by focusing on each of several user-segments, previously classified by data mining techniques on historical data, by using human expertise, or by buying information from a 3rd party.

2. Inventory pricing: Given a fixed budget, how to choose the optimal bid, at a given instant, for different aggregates of inventory (each one with different statistical properties in terms of the way their auctioning dynamics).

3. Budget pacing: Given a fixed budget and restricted to a given aggregate of inventory, how to optimally spent the budget throughout the day.

The goal of this study is to tackle the aforementioned problem from a mathematical perspective, focusing on the fundamental relation between the different variables. This contrast the mainstream quantitative literature on the topic of advertising, in which the main focus has been the application of machine-learning techniques to real-time problems [10]. Even though the latter work is interesting when focusing in a user-based perspective, it misleads into thinking that the problem is not tractable from an analytical perspective and rushing into brute force methods is always necessary. We show in this study that, on the contrary, an analysis of the problem by using elementary identities (without making strong hypothesis) lead to analytical results that helps to completely understand the problem on an idealized setting (homogeneous inventory) which can serve as starting point to pinpoint the type of variables that create real leverage when wanting to apply data mining techniques.

1.5 Outline and contributions

In the light of the previous discussion, we start our work by providing in Chapter 2 a statistical model that allows to understand the important quantities when dealing with a large number of second-price auctions, mostly based on the mathematical results on [4]. We also study comparative statics relating the impact of macroscopic variables on performance indicator. On Chapter 3 we focus on the problem of inventory pricing by showing that the choice of the optimal bid under constant budget can be formalized as a convex optimization problem. On Chapter 4 we focus on the dynamic optimization (i.e. Budget pacing) by using tools from the theory of variational calculus which allows to find the optimal rate at which spend the budget in several context, the most interesting result being the case of frequency capping (when we limit the maximum number of impressions per user). A conclusion and a bibliography follow.

2 Modeling and comparative statics

2.1 Introduction

We start by considering an agent continuously bidding in a large number of auctions aiming to purchase ad-inventory (e.g. a trading-desk acting in behalf of an advertiser). We consider that these auctions can arise in different *contexts* (this can refer to different ad-exchanges, publishers or any other suitable aggregate of inventory), however, for the sake of this section we focus on an agent bidding on only one context.

Because the bidding process happens thousands of times per minute, from a mesoscopic scale (minutes, hours), we can consider these contexts as black-boxes where, for each level of bid, there is a probability to win the auction and a random-variable defining the price we will pay if we win[2] (for an in-depth analysis, see [4]).

[2]We only focus on the statistical features that emerges from the interaction of different participants

The goal of this section is to study the fundamental relationships between the different mesoscopic variables, taking as starting point the auction mechanisms. We also provide comparative statics relating macroscopic variables as CPA, CPM and total budget. The latter is of particular interest from an industrial standpoint, as it provides a rational understanding of how parameters are related and thus, managing expectations of practitioners and clients with a weaker quantitative background, which do not necessarily grasp the natural limitations of a programmatic advertising campaign.

2.2 Auction mechanisms

In the RTB auctions, inventory is attributed to the agent posting the highest bid in a context where a participant cannot know the bid of the others. The only information an agent gets it is if he wins the auction or not. In the latter case he knows the price he ends up paying. The price paid by the winner depends on the type of auction:

- **Second-price sealed-bid auctions** (or Vickrey auctions) in which, in case of winning the auction, the price paid corresponds to the bid proposed by the second best participant. **This is the standard type of auction in RTB.**

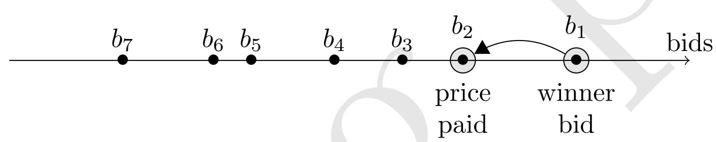

Figure 3: Mechanisms for the second-price auction.

- **First-price sealed-bid auctions** (or FPSB), in which, in case of winning the auction the price paid equals the bid proposed.

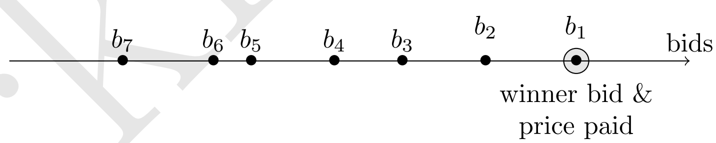

Figure 4: Mechanisms for the first-price auction.

2.3 Preliminary definitions

We consider an homogeneous context in which we participate to R auction-requests. We assume this quantity R to be fixed. The length of the time window is irrelevant.

For a given bid-level b with which we participate to the auctions, we define the *winrate* $w(b)$ as the function giving the probability to win a bid-request for that level of bid b.

in the RTB environment, not on their strategic behavior. This hypothesis is reasonable in our context where user segments and contexts are treated as homogeneous. In context where the user is interesting for several advertisers (e.g. retargeting) this hypothesis is no longer valid.

Similarly, we note the *average price* paid at that bid-level, by $p(b)$. Notice that $p(b)$ is a proxy of the CPM, hence we will use both names interchangeably.

We also introduce the macroscopic variables S (total budget spent), I (total inventory purchased) and C the total number of conversions. The ratio between the conversions and the inventory purchased is called *conversion rate* and denoted by γ. The hypothesis of homogeneous inventory means in particular that γ is constant.

Even though S, C and I are random variables, throughout this study we consider that the number of bid-request is large enough so that the ratios such as the CPM or the CPA are deterministic quantities (due to the law of large numbers).

2.4 Fundamental identities

The following differential identities are the basis of all that follows (they are derived in a similar way that fundamental relationships in thermodynamics [1]).

- If we are in the context of a **second-price auction**, increasing the winrate by an infinitesimal term dw makes the algorithm to purchase Rdw more inventory. This inventory was not possible to earn for a bid lesser than b, so, considering that the increase in bid db is also infinitesimal, we have that the new inventory is purchased at a price close to b. Hence the increase in spending satisfies

$$dS = R \cdot b \cdot dw. \tag{1}$$

- In the case of a **first-price auction**, an increase in the winrate by dw also means an increase in bid of db, however, now all the inventory is purchased at $b + db$, not only the new inventory. Hence, the variation in spent satisfies

$$dS = \underbrace{R \cdot b \cdot dw}_{\text{cost of new inventory}} + \underbrace{R \cdot w \cdot db}_{\text{old inventory becomes more expensive}}. \tag{2}$$

On the other hand, the spent satisfies $S = R \cdot w \cdot p$. Which, if we divide the precedent equations by R, leads to the following theorem

Theorem 1. *The relationship between price, winrate and bid is given by the following identities:*

- *For a second-price auction:*

$$d(p \cdot w) = b \cdot dw \tag{3}$$

- *For a first-price auction:*

$$d(p \cdot w) = b \cdot dw + w \cdot db \tag{4}$$

Proof. It follows immediately from equations (1) and (2). ☐

2.5 Relation between macroscopic variables

In order to measure the effects of the different variables on the performance in terms of conversions, we introduce a parameter γ measuring the probability to turn an impression into a conversion. Thus, the number of conversions is given by

$$C = \gamma \cdot I = \gamma \cdot R \cdot w$$

In this way we define the CPA as the ratio between the total spent and the number of conversions

$$\text{CPA} = \frac{S}{C} = \frac{p \cdot w \cdot R}{\gamma \cdot w \cdot R} = \gamma^{-1} \cdot p.$$

In particular, we have that, on a given aggregate of inventory, under the hypothesis of homogeneous user segments and constant conversion rate, **a relative change in CPA yields the same relative change in CPM** (their ratio being constant).

In what follows, we want to measure the effects on the performance (i.e. CPA) when we increase the campaign's budget. This is interesting because practitioners use the CPA as a benchmark for advertising, regardless of the total budget of the campaign. However, it is easy to argue that this is misleading as budget increases can erode the CPA due to the fact that the algorithm is forced to buy impressions at higher prices so, if the conversion rate remains the same, the CPA will become invariably worst. Otherwise said, arguing that the CPA is a good benchmark, independent of the budget, is the equivalent of saying that more expensive inventory has a higher probability of conversion (i.e. a sort of fair pricing hypothesis, similar to the efficient markets hypothesis in finance).

In that view, first, we will consider γ constant and measure how much the CPA is eroded by an increase in budget. In a second time, we will suppose that buying more expensive inventory yields to a higher probability of conversion (i.e. the so called 'premium inventory' is really premium – i.e. higher γ) and in that case we will quantify how much the conversion rate needs to improve in order to justify paying a higher CPM.

2.6 Effects of the total budget on the CPA

First of all, we have

$$d\text{CPA} = \gamma^{-1} dp.$$

Following Theorem 1 we obtain

$$d\text{CPA} = \gamma^{-1}(b-p)\frac{dw}{w}.$$

Which is equivalent to

$$\frac{d\text{CPA}}{\text{CPA}} = \frac{b-p}{b} \cdot \frac{dS}{S}$$

Otherwise said, if the conversion rate is constant on a particular pool of inventory which homogeneous user-segments, then the relative variation of CPA can be computed as the relative variation of budget S, multiplied by a term $(b-p)b^{-1}$, which is positive.

2.7 When *premium inventory* can be justified?

If we believe in a sort of market efficiency (the more expensive impressions of a given context are higher-quality), a natural question is how to to offset the CPA erosion due the increase of budget with the purchasing of the more 'quality' inventory we are purchasing a higher prices. This is an interesting question as in the latter years several sellers claim that the fact their inventory is more expensive (regardless of the user) is because of its 'premium' quality.

If we suppose that the conversion rate γ depends on the average price paid, we have

$$d\log(\text{CPA}) = d\log(p) - d\log(\gamma).$$

Otherwise said

$$\frac{d\text{CPA}}{\text{CPA}} = \frac{dp}{p} - \frac{d\gamma}{\gamma}.$$

Hence, in order to offset CPA erosion, we need

$$\frac{d\gamma}{\gamma} \geq \frac{b-p}{b} \cdot \frac{dS}{S}.$$

3 Optimal bidding and inventory pricing

3.1 Introduction

In this section we study the problem of choosing the bid for each one of different aggregates of inventory. As in the previous section, the technical details are explored in [4].

We suppose that there are $K \in \mathbb{N}$ contexts in which we can purchase ad-inventory and we would like to choose the bid for each one of these context in order to maximize the number of impressions given a fixed budget $S \in \mathbb{R}_+$ (i.e. CPM-minimization). In other words, our goal is how to correctly *price* different aggregates of inventory, under the hypothesis of an homogeneous user-segment (in terms of probability of conversion).

In a first time, we consider the auctions are *second-price* so the price paid if we win the auction correspond to the second best bid. As we evolve in our discussion we consider the case where the auctions mechanisms are modified by *floor-prices*.

3.2 Optimization problem

Let us note R_k the number of auction-requests on context $k \in \{1, \dots, K\}$ and w_k its winrate. Hence, the total number of purchased impressions is given by:

$$I(w_1, \dots, w_K) = \sum_{k=1}^{K} R_k w_k.$$

Similarly, if we note by $p_k = p_k(w_k)$ the price paid in context k when the winrate is w_k, and $b_k = b_k(w_k)$ the associated bid. Then, the total budget spent is given by

$$S(w_1, \dots, w_K) = \sum_{k=1}^{K} R_k w_k p_k(w_k).$$

In particular, for

$$\mathcal{L}(w_1, \dots, w_K) = I(w_1, \dots, w_K) - \lambda(S(w_1, \dots, w_K) - S), \quad S \in \mathbb{R}_+, \quad \lambda \in \mathbb{R},$$

we have (in general)

$$d\mathcal{L}(w_1, \dots, w_K) = \sum_{k=1}^{K} R_k (dw_k - \lambda d(p_k w_k)) \tag{5}$$

and (in the case of second-price auctions)

$$d\mathcal{L}(w_1, \dots, w_K) = \sum_{k=1}^{K} R_k (1 - \lambda b_k) dw_k$$

Otherwise said, in order to maximize the number of impressions under a budget constraint, then for every $k \in \{1, \ldots, K\}$ we have $b_k = \lambda^{-1}$. Otherwise said, the optimal bid is the same for every context.

Theorem 2. *(Optimal bidding in second-price auctions) Under the hypothesis of homogeneous user-segments and a second-price auction mechanism for every context. Then, the optimal bid is constant across contexts.*

Proof. The quantity $\mathcal{L}(w_1, \ldots, w_K)$ defined above is the Lagrangian of the optimization problem. $\square$

3.3 Dealing with floor-prices

In practice, some sellers of inventory allow modified auction mechanism, which in some cases can be profitable for publishers as they can sell their inventory at a price higher than the profit they would have made in a classical second-price auction setting. The idea is to set *floor-prices* which consists in thresholds such that below the floor-price the auction is no more of the second-price type. The two main settings used in practice are *soft floor-prices* and *hard floor-price* as we explain below[3].

Definition 1. *(Soft-floor) Let us consider a second-price auction where M participants post bids $\{b_1, \ldots, b_M\}$. We said that a price-level $s \in \mathbb{R}^+$ is a soft-floor if, a new participant offering a bid b such as $b \geq b^* := \max\{b_1, \ldots, b_M\}$, will win the auction paying b if $b \leq s$ and $\max\{b^*, s\}$ if $b \geq s$. Otherwise said, if all the bids yield below the soft floor-price, the auction mechanics becomes of the first-price type.*

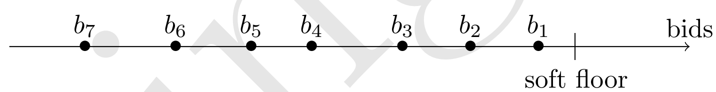

Figure 5: If the soft floor-price is higher than all the bids in the auction we are in the context of a first-price auction. i.e. the best bid b^* (in this case b_1) is at the same time the price which is paid by the winner of the auction.

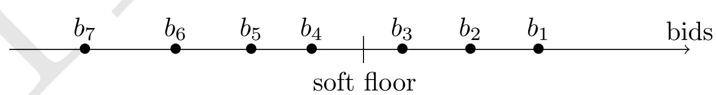

Figure 6: If the soft floor-price is lower than the best bid in the auction, the auction remains of the second-price type.

[3] It is commonly assumed that floor-price are always advantageous for the publisher, however, this is not an accurate statement. Otherwise said, even though, for a particular auction, the floor-price can increase the profits for the publisher, in the long-run, higher prices for the advertiser on a subset of the inventory, can represent lower performance indicators which motivates the advertiser to decrease the budget allocated to inventory subject to floor-prices. The only case when floor-prices are advantageous for a publisher is when the advertiser has no choice but to win the auction (e.g. a retargeting campaign where we have an strong signal that a particular user is interested in the campaign).

Definition 2. *(Hard-floor) Let us consider second-price auction where M participants post bids $\{b_1, \ldots, b_M\}$. We said that a price-level $h \in \mathbb{R}^+$ is a hard-floor if, in order to win the auction a new participant should post a bid higher than h and, if he wins the auction, the price paid is* $\max\{h, b_1, \ldots, b_M\}$*. Otherwise said, a hard-floor can be interpreted as if there where a 'ghost' participant always bidding at a level h.*

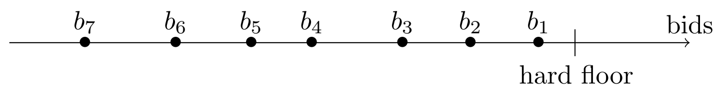

Figure 7: If the hard-floor (also know as *reserve price*) is higher than all the bids in the auction, the inventory is not attributed among the auction participants.

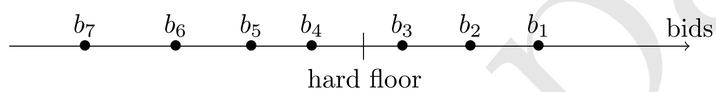

Figure 8: If the hard-floor is lower than the best-bid the winner of the auction will pay the maximum between the second-best bid (in this case b_2) and the hard-floor.

3.3.1 Optimization under soft floor-prices

The case of soft floor-prices is similar to considering that some aggregates of inventory function on a logic of first-price auction. Equations (5) and (2) lead to the condition

$$\frac{b_k}{b^*} = 1 - w_k b'_k(w_k),$$

where b^* is the optimal bid for the inventory purchased on second-price auctions.

Otherwise said, the optimal bid for inventory purchased on first-price auctions should be lesser than the bid for second-price auctions, by a factor of $1 - w_k b'_k(w_k)$ (i.e. depending on the statistical features of the auction – for more details, see [4]).

3.3.2 Optimization under hard floor-prices

In the case of hard-floor prices the auction remains of the second-price type. In can be shown that in this case our main result (uniform bid across inventory purchased through second-price auctions) remains valid. A subtle point needs to be made as throughout all this study we have assumed that infinitesimal changes of bid lead to infinitesimal changes of winrate, which is not true in the case of hard floor-prices. However, the main consequences of Theorem 2 still remain true in this case (see [4] for details).

4 Budget pacing and frequency capping

4.1 Introduction

As discussed in one of the previous sections, under the hypothesis of second-price auction and homogeneous inventory, the optimal strategy is to bid uniformly across the different aggregates of inventory. Notice that this result remains true regardless the laps of time we consider, so, in particular, the result can be interpreted as: at any given instant, the bids are uniform across the different aggregates of inventory. Moreover, we do not yet explicitly define the value of this optimal bid, and how it evolves. As we will see in this section, the factor that will (dynamically) define the bid value at any instant is the *budget pacing* (i.e. the optimal rate to spent the budget throughout the day).

Here, we consider the problem of how to optimally smooth the spent throughout the trading session, in the sense of maximizing the number of impressions (i.e. minimizing the CPM). We keep considering the inventory homogeneous in terms of conversion rates (in particular, considering a time window where the statistical properties of the system are stationary). We introduce an optimization framework based on variational calculus which permits to obtain closed formulas. Our main result is that the optimal spending is not linear in time, but proportional to the number of auction-requests we participate in. We also show that the same framework can be applied in the case of frequency-capping, i.e. when the maximum number of banners exposed to each user is capped.

4.2 Model setup

Let us consider an algorithm participating in a sequence of Real-Time Bidding auctions (e.g. a media trading-desk or a DSP in behalf of an advertiser) aiming to spend a fixed budget $S \in \mathbb{R}_+$, over a given period of time $[0, T]$ (this can represent few hours) and on a pre-defined audience segment (e.g. users, domains). Because of the large number of auctions algorithms face in practice, we consider the number of auction-requests per unit of time the algorithm receives as a continuous positive function $\mu : t \mapsto \mu(t)$. Hence, the total auction-requests received, up to time $\tau \in [0, T]$, is given by

$$\nu(\tau) = \int_0^\tau \mu(t)dt. \tag{6}$$

At each instant, the algorithm chooses the bid-level $b(t)$ to participate in the auctions, aiming to maximize the inventory purchased under the budget constraints.

At this point, two quantities are (re-)introduced:

- The *winrate*: An increasing function $w : \mathbb{R}_+ \to [0, 1]$ such as $w(b)$ represents the ratio of auctions won when the bid level is equal to $b > 0$.
- The *CPM (cost-per-mille)*: An increasing function $p : \mathbb{R}_+ \to \mathbb{R}_+$ such as $p(b)$ represents the price paid per one thousand units of inventory when the bid level is equal to $b > 0$. Throughout the calculations we will use the *cost-per-unit* instead of CPM, in order to avoid carrying the *per-mille* factor.

In a first time we will assume both functions strictly increasing and differentiable (we discuss about this hypothesis on subsection IV).

With this in mind, for a bidding strategy $(b(t))_{t\in[0,T]}$, we can define now the evolution of the *running purchased-inventory* and the *running spent* as follows.

- Running purchased-inventory $I : t \mapsto I(t)$ which satisfies the equation

$$I'(t) = \underbrace{\mu(t)}_{\text{nb. of requests}} \times \underbrace{w(b(t))}_{\text{winrate}}, \qquad I(0) = 0. \tag{7}$$

- Running spent $S : t \mapsto S(t)$ which satisfies the equation.

$$S'(t) = \underbrace{I'(t)}_{\text{new inventory}} \times \underbrace{p(b(t))}_{\text{inventory cost}}, \qquad S(0) = 0. \tag{8}$$

Two optimizations problems can be defined at this stage

Definition 3. *(Direct optimization problem) Maximizing the purchased inventory $I(T)$ as a functional of the winrate $(w_t)_{t\in[0,T]}$ (controlled by the algorithm through the bid level, i.e. $w_t = w(b(t))$) under the budget constraint $S(T) = S \in \mathbb{R}_+$.*

Definition 4. *(Dual optimization problem) Minimizing the total budget spent $S(T)$ as a functional of the winrate $(w_t)_{t\in[0,T]}$ (controlled by the algorithm through the bid level, i.e. $w_t = w(b(t))$) under the inventory constraint $I(T) = I \in \mathbb{R}_+$.*

Even though these problems are different, from a dynamic standpoint they lead to the same solution i.e. *how the winrate should be controlled in order to buy inventory at the lower cost*. The following theorem states that by choosing to solve the dual optimization problem we obtain a dynamic relation stating that *the speed at which the budget is spent should be proportional to the number of incoming auction-requests*[4].

Theorem 3. *The optimal budget-pacing strategy yields the dynamic relation*

$$\forall t \in [0, T], \quad S'(t) = C\mu(t), \quad C \in \mathbb{R}.$$

i.e. the optimal pacing strategy consists on spending proportional to the number of auction-requests.

Proof. We outline the proof in several steps as it will be the same way of reasoning than in the case with frequency-capping (Chapter 4).

Step 1: Change of time

Let us consider the process $\nu(t)$ as in equation (6). Because ν is, by definition, increasing, we can introduce the two following processes

$$\begin{aligned} \widetilde{I} &: \mathbb{R}_+ \to \mathbb{R}_+, \quad \text{such as} \quad I(t) = \widetilde{I}(\nu(t)), \\ \widetilde{S} &: \mathbb{R}_+ \to \mathbb{R}_+, \quad \text{such as} \quad S(t) = \widetilde{S}(\nu(t)). \end{aligned}$$

i.e. time is now counted in *number of auction-requests*, not physical time.

[4]This means, from a practical standpoint, that the optimal pacing strategy can be mechanically obtained by a feedback-control process matching the instantaneous spent to the number of auction-requests. Otherwise said, there is no need to have an analytical form for the bid.

In the new variables we have:

$$
\begin{aligned}
I'(t) &= \widetilde{I}'(\nu(t))\mu(t) \\
S'(t) &= \widetilde{S}'(\nu(t))\mu(t)
\end{aligned}
$$

Step 2: New dynamic equations

Replacing the relations above on equations (7) and (8) yields

$$
\begin{aligned}
\widetilde{I}'(\nu) &= w(b(\nu)) \\
\widetilde{S}'(\nu) &= \widetilde{I}'(\nu)p(b(\nu))
\end{aligned}
$$

As said before, we assume $w(\cdot)$ to be strictly increasing, thus it admits an inverse $w^{-1}(\cdot)$. This leads to the following equation; central in the proof of the theorem

$$
\widetilde{S}'(\nu) = \widetilde{I}'(\nu)p\left(w^{-1}\left(\widetilde{I}'(\nu)\right)\right). \tag{9}
$$

Step 3: Variational problem

We focus on the dual problem, which consists to minimize the budget spent $\widetilde{S}(\nu_T)$ in order to achieve a given purchased inventory $I > 0$. Otherwise said, we want to solve

$$
\min_{\widetilde{I}\in\mathcal{I}} \int_0^{\nu_T} \widetilde{I}'(\nu)p\left(w^{-1}\left(\widetilde{I}'(\nu)\right)\right) d\nu
$$

under the constraint

$$
I(0) = 0 \quad \text{and} \quad I(\nu_T) = I.
$$

Here $\mathcal{I}$ represents the set of admissible dynamics for the evolution of the purchased inventory. We do not make special assumptions other than the regularity of $\widetilde{I}(\cdot)$.

The problem defined above can be solved by means of the Euler-Lagrange equation, which in this case reads[5]:

$$
\left(\frac{\partial}{\partial I} - \frac{d}{d\tau}\frac{\partial}{\partial \dot{I}}\right)[\dot{I}(p\circ w^{-1})(\dot{I})] = 0.
$$

In this case, this means, it exists C, constant, such as

$$
\frac{\partial}{\partial \dot{I}}[\dot{I}(p\circ w^{-1})(\dot{I})] = C
$$

Which in particular implies that the quantity $I'(\tau)$ is constant throughout the entire period (this, because there is no dependency on other quantities in the last equation).

Moreover, as the evolution of the budget only depends on $I'(\tau)$, we state our result in the following way: *The optimal budget-pacing strategy satisfies*

$$
S'(t) \propto \mu(t).
$$

$\square$

[5]In variational calculus, $\dot{I}$ is the standard notation for $I'(\tau)$.

Practical remark Our result states that the optimal budget-pacing strategy is such as we should spent proportional to the number of auction-requests we observe (and not necessarily linearly in time). With that in mind, there is no need in practice to define the optimal bid in a explicit way as it can be controlled by a on-line feedback-control system trying to follow, for example, an optimal budget-pacing curve. Moreover, using for example the observed daily patterns of the number of auction-requests (e.g. assuming that is stable across the days), the optimal budget-pacing curve to follow could be defined beforehand. However, an important issue that arises when applying a feedback-control in order to match the realized spent to the target spent is related to floor prices. Along our reasoning we use the hypothesis the winrate is strictly increasing, and differentiable, as a function of the control (i.e. the level of bid at a given instant). This is not always the case as, in practice, some ad-exchanges set floor-prices, which creates discontinuities on the winrate function (until some level of bid the winrate is zero and, suddenly, a infinitesimal increase can make our algorithm purchase a considerable amount of inventory). Moreover, these floor-prices are not known before hand (even though, stable for certain sellers of advertisement). This phenomena does not impact the main result of the previous subsection (i.e. $S'(t) \propto \mu(t)$), however, it creates challenges for media-trading desks and DSP when controlling the bid through a feedback-control system.

4.3 Dealing with frequency-capping

In the set-up of advertising campaigns, practitioners like to define a frequency-capping factor, limiting the exposure (i.e. impressions) each user can see. We assume this parameter is global, that is, not changing from user to user [6].

4.3.1 Effect on the dynamics

We represent the capping effect as a function $\theta : \mathbb{R}_+ \mapsto [0, 1]$, depending on the number of the inventory-purchased. Hence, $\theta(I(\tau))$ represents the number of users are still available to be exposed to advertising, after we have already showed $I(\tau)$ impressions.

The new equations for I and S are:

$$\begin{aligned} I'(\tau) &= w(b(\tau))\theta(I(\tau)) \\ S'(\tau) &= I'(\tau)p(b(\tau)). \end{aligned}$$

Thus,

$$S'(\tau) = I'(\tau)(p \circ w^{-1})\left(\frac{I'(\tau)}{\theta(I(\tau))}\right)$$

4.3.2 A (new) change of time

Proceeding similarly as in subsection II, we define a change of time by defining:

$$\nu = \int_0^\tau \theta(I(u))du$$

[6] From a probabilistic standpoint, the effects on capping are similar than in a urn model with a given quantity of balls (potential users) where, in the beginning, all balls are labeled by 0 (number of exposures) and, as they are drawn randomly from the urn, they are replaced by a ball with a label with a number one unit higher (i.e. $n \mapsto n+1$). If the maximum capping is $K \in \mathbb{N}$, then that means all the balls labeled with a number larger or equal than K go back into the urn without being replaced, and the more the process is repeated, the larger the ratio of balls already labeled by K.

Hence, we obtain

$$I'(\tau) = \tilde{I}'(\nu)\theta(I(\tau))$$

and

$$S'(\tau) = \tilde{S}'(\nu)\theta(I(\tau)).$$

Leading to the dynamics

$$\tilde{S}'(\nu) = \tilde{I}'(\nu)(p \circ w^{-1})(\tilde{I}'(\nu))$$

Where the symbol ˜ denotes the fact that the underlying time is ν.

4.3.3 Solution

By applying the same variational approach than in Theorem 3, we obtain that at the optimum

$$S'(\tau) \propto \theta(I(\tau)),$$

or equivalently:

$$S'(t) \propto \theta(I(t))\mu(t).$$

Otherwise said, the optimal solution is spending proportional to the number of auction-requests that are not 'banned' yet by the capping.

In practice this is slightly more difficult as it demands to track users in order to approximate the effects of the frequency-capping. These can also be obtained by simulating the effects of the capping via the simulation of urn-models.

5 Conclusion

In this article, we proposed a framework to analytically understand different optimization issues arising in Real-Time bidding under an hypothesis of homogeneous inventory (in terms of conversion rates). First, we develop a model and obtain comparative statics between the main variables of interest in the automated operation of an advertising campaign. In a second time, we tackle the problem of bidding on a large set of Vickrey auctions, by buying inventory across different contexts. Our framework allows to obtain the optimal tactics on a straightforward way by solving a constrained optimization problem. We show that in the Vickrey situation the optimal bid has to be uniform across the sellers. We also studied the situation were soft-floor prices and hard-floor prices were presented; in that case our framework allows to easily obtain an analytical form for the optimal bid.

Finally, we solved the problem of optimal budget-pacing in real-time bidding by proposing a variational calculus approach. Our main result is that the optimal way to pace the budget is by spending proportional to the number of auction-requests we observe, and not linearly as practitioners do. We shown that this approach can be also applied in the case advertising campaigns have, in their set-up, frequency-capping factors limiting the exposure to advertising.

From the point of view of automating bidding strategies, the main lines for future research are in the line of (1) the feedback-control problem of trying to follow a pacing curve depending on stochastic processes (the evolving number of bid requests) and issues like hard-floor representing discontinuities for the system, (2) studying the problem not from a continuous perspective but considering the evolution of impressions as a non-homogeneous Poisson process, i.e. the pacing problem becomes a stochastic-control problem and (3) how to include in-homogeneity in the system (adapt the algorithm to signals where the conversion rate can unexpectedly change during the day.

From a modeling perspective, it would be interesting the study of customer journeys and statistics about frequency-capping from the point of view of urn-models. The idea is to understand the probabilities of some users behaving in some given way (i.e. number of users visiting a given pair of contexts, or seeing a given number of impressions) which emerge naturally from random behavior, without representing real correlations.

Another line of research is to explore further the comparative statics and find fundamental relationships in more complex settings, with the aim of quantitatively understand value creation and inventory pricing from a macroscopic perspective. Specially with the aim of quantifying the real value of 'premium inventory' and the real leverage of including floor-prices in the auction mechanics.

References

[1] Cannon, J. W. (2004). Connecting thermodynamics to students calculus. American Journal of Physics, 72(6), 753-757.

[2] Chen, Y., Berkhin, P., Anderson, B., & Devanur, N. R. (2011, August). Real-time bidding algorithms for performance-based display ad allocation. In Proceedings of the 17th ACM SIGKDD

[3] Fernandez-Tapia, J. (2015). Statistical modeling of Vickrey auctions and applications to automated bidding strategies. Working paper.

[4] Fernandez-Tapia, J., & Monzani, C. (2015). Stochastic Multi-Armed Bandit Algorithm for Optimal Budget Allocation in Programmatic Advertising. Available at SSRN 2600473.

[5] Lee, K. C., Jalali, A., & Dasdan, A. (2013, August). Real time bid optimization with smooth budget delivery in online advertising. In Proceedings of the Seventh International Workshop on Data Mining for Online Advertising (p. 1). ACM.

[6] Lee, K. C., Orten, B., Dasdan, A., & Li, W. (2012, August). Estimating conversion rate in display advertising from past performance data. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 768-776). ACM.

[7] Magna Global (2014, October). MAGNA GLOBAL Ad Forecasts: Programmatic Buying Reaching a Tipping Point. http://www.magnaglobal.com/.

[8] Muthukrishnan, S. (2009). Ad exchanges: Research issues. In Internet and Network Economics (pp. 1-12). Springer Berlin Heidelberg.

[9] Perlich, C., Dalessandro, B., Hook, R., Stitelman, O., Raeder, T., & Provost, F. (2012, August). Bid optimizing and inventory scoring in targeted online advertising. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 804-812). ACM.

[10] Richardson, M., Dominowska, E., & Ragno, R. (2007, May). Predicting clicks: estimating the click-through rate for new ads. In Proceedings of the 16th international conference on World Wide Web (pp. 521-530). ACM.

[11] Yuan, Y., Wang, F., Li, J., & Qin, R. (2014, October). A survey on real time bidding advertising. In Service Operations and Logistics, and Informatics (SOLI), 2014 IEEE International Conference on (pp. 418-423). IEEE.

[12] Yuan, S., Wang, J., & Zhao, X. (2013, August). Real-time bidding for online advertising: measurement and analysis. In Proceedings of the Seventh International Workshop on Data Mining for Online Advertising (p. 3). ACM.

[13] Zhang, C. R., & Zhang, E. (2014, August). Optimized bidding algorithm of real time bidding in online ads auction. In Management Science & Engineering (ICMSE), 2014 International Conference on (pp. 33-42). IEEE.

[14] Zhang, W., Yuan, S., & Wang, J. (2014, August). Optimal real-time bidding for display advertising. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1077-1086). ACM.