TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

D I P L O M A R B E I T

Optimal Bidding in the Sponsored Search

Ausgeführt am Institut für

Wirtschaftsmathematik

der Technischen Universität Wien

unter der Anleitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Gernot Tragler

durch

Sophie Grünbacher

Robert Stolz-Gasse 9

2344 Maria Enzersdorf

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Ort, Datum	Sophie Grünbacher

Danksagung

Ich möchte mich an dieser Stelle bei all jenen bedanken, die mich bei der Anfertigung meiner Diplomarbeit und während meines gesamten Studiums so tatkräftig unterstützt und motiviert haben.

Mein Dank gilt Herrn Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Gernot Tragler, der mit sehr viel Engagement und unermüdlichem Einsatz meine Arbeit betreut hat.

Bedanken möchte ich mich auch bei dem Kartenbüro (in dieser Diplomarbeit *ticket agency* genannt), das mir ermöglicht hat, die zugrundeliegende Theorie dieser Arbeit in die Praxis umzusetzen, zu testen und zu evaluieren.

Mein besonderer Dank gilt meiner Familie, insbesondere meinen Eltern, die mir mein Studium ermöglicht haben, auch emotional immer für mich da waren und mich vor allem in all meinen Entscheidungen unterstützt haben.

Herzlich bedanken möchte ich mich auch bei meinem Verlobten, der mich immer wieder ermutigte und mich mit vielen nützlichen Ideen und Ratschlägen tatkräftig unterstützt hat.

Großer Dank gebührt auch meinen Studienkollegen, die mir eine ganz besondere Studienzeit mit vielen wunderschönen Momenten bereitet haben.

Nicht zuletzt gebührt Mag. Karl Tutschek Dank, der durch seinen sehr guten und interessanten Unterricht im Kollegium Kalksburg meine Leidenschaft an der Mathematik geweckt und mich bereits beim Schreiben meiner Fachbereichsarbeit betreut hat.

Abstract

In this thesis, I describe how search engine advertising works in the sponsored search, the underlying Generalized Second-Price (GSP) Auction, which is the mechanism used by search engines to sell online advertising, and the difficulties to find an optimal bid. For that purpose, I present a model to optimize the bids in an advertiser's campaign and describe the Generalized Method of Moments (GMM) estimator needed to estimate necessary parameters for the bid optimization. Moreover, I analyze whether or not it is possible to improve the performance of a ticket agency's search engine advertising by using the bidding policy and by automizing the bid optimization. To validate the effectiveness of that model, I use a data set from the Google Adwords campaign of that ticket agency, compute the optimal bids, and implement them into their campaign. It appears that for the ticket agency the proposed bidding technique is not as effective as previously imagined. On the contrary, the return on investment of the ticket agency's advertising campaign based on a Difference-in-Differences (DiD) approach decreases by 500%. That result shows that it would be necessary to provide much more information to the model than in this thesis, to be able to improve this ticket agency's advertising campaign.

Contents

1 Introduction

Advertising has changed a lot in the past few years. People have altered their behavior: they are buying less in retail stores and more on the web. Even if they do not buy a certain product on the internet they often use search engines in advance to receive informations about brands, prices, and stores. Hence, online advertising has become more and more important not only for online shops but also for retail stores. It is even possible to reach a target group at the perfect time. By making use of the internet, companies don't have to pay a lot for a TV spot or a huge poster on the road and hope for the intended audience to see their advertisement, but they can reach people exactly at that moment they are looking for the product or the service the company is selling. They can achieve that by using search engine advertising, where advertisers submit bids for a variety of search terms.

In my diploma thesis I will analyze whether or not it is possible to improve the performance of a ticket agency's search engine advertising by using the bidding policy and by automizing the bid optimization as proposed by [Abhishek and Hosanagar, 2013].

For that purpose, I had to convince the ticket agency to allow me to use their data of search engine advertising in Google Adwords. Having that permission I implemented a daily automated download of a summary of the data from the keywords of the preceding day on Google Adwords. While collecting a lot of data I started to write a program, which uses the ticket agency's search engine advertising data as the input and the optimal bid – according to the bidding policy of [Abhishek and Hosanagar, 2013] – as the output. After four months of data collection I was able to implement the computed optimal bid into the

Google Adwords account of the ticket agency and observe the change in the performance, which I will describe in Section 4.5.

My diploma thesis is structured as follows: In Chapter 2, I describe how the underlying mechanism of search engine advertising works and how difficult it is to find an optimal bid. In Chapter 3, I formulate the analytical model given by [Abhishek and Hosanagar, 2013] and introduce the Generalized Method of Moments (GMM) estimator, which is used to estimate the parameters of our model required to solve the underlying bid optimization problem. In Chapter 4, I first present the data set of the ticket agency before applying the methods of Chapter 3 to the data, compute the optimal bids, present the implementation of them into Google Adwords, compare the results of the implemented bidding policy with the advertiser's previous policy, and finally analyze the results of that comparison. Chapter 5 concludes with a summary and discussion of the obtained results.

2 Online Advertising

There are two different kinds of search results: the *organic* and the *sponsored* search results, which can be seen in the following example. If someone is looking for shoes on Google (s)he will receive a search result similar to the one depicted in Figure 2.1. The goal of Search Engine Marketing (SEM) for a company is to be listed at the highest possible position both on the organic and on the sponsored search results. Hence SEM consists of two different sections: Search Engine Optimization (SEO) to appear in the organic results and Search Engine Advertising (SEA) to be listed in the sponsored search. There are a lot of possibilities to improve the listing of your Website with SEO but in this thesis I will concentrate on the second mentioned subgroup. For that purpose, I will give next an overview of how sponsored search advertising works.

2.1 Sponsored Search Advertising

To be listed in the paid search, each advertiser has to specify search terms, which will be called 'keywords' in what follows, that are related to the products (s)he wants to sell, create text ads, and indicate how much (s)he is willing to pay if a user clicks on the corresponding ad after searching for a certain keyword. If a user enters a search query, not only the usual list of links appears but also the ads of some advertisers that are bidding on that keyword or on a very similar keyword.

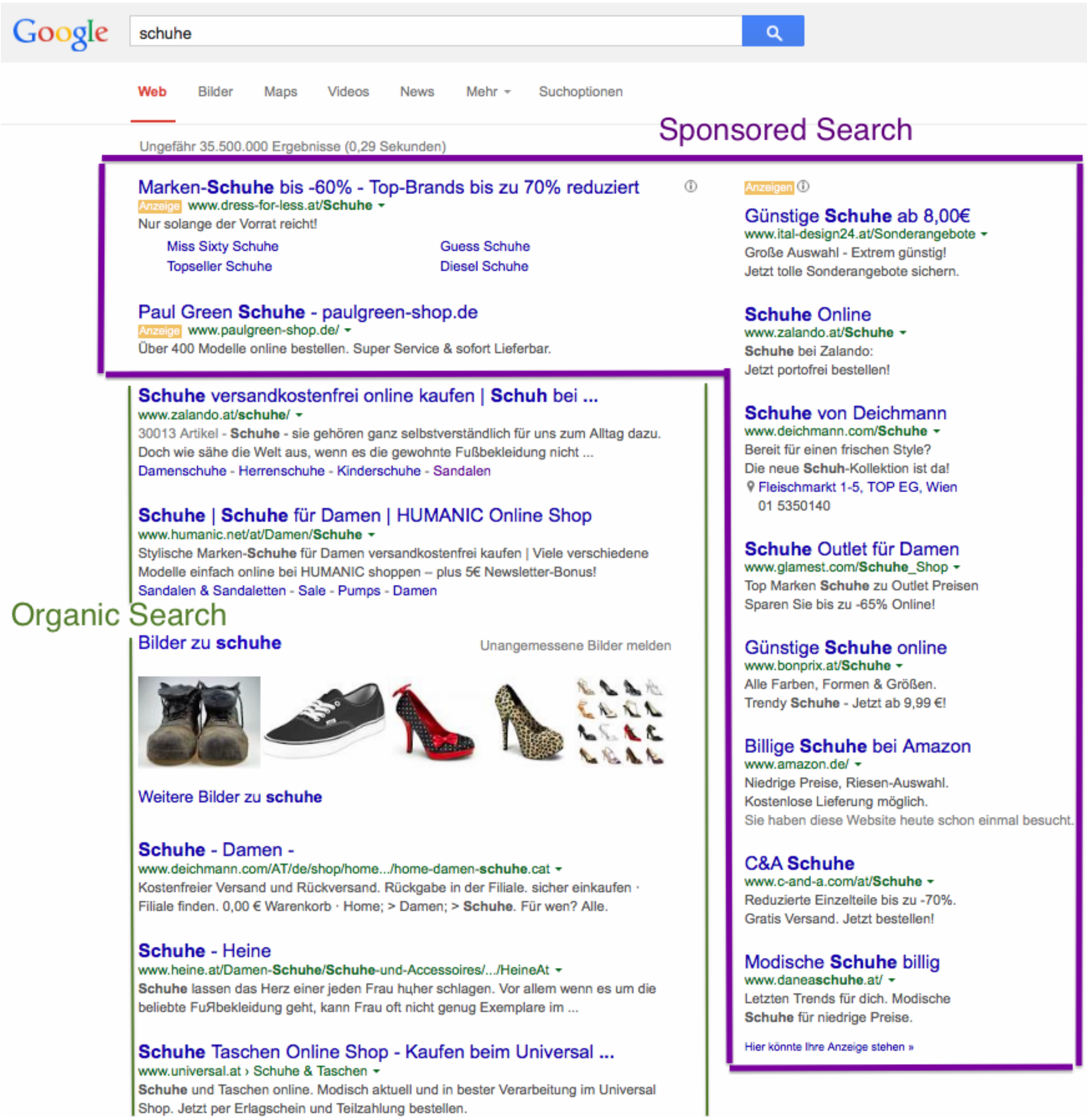

Figure 2.1: Search results on Google Adwords; see [Google, 2014c].

The relative positions of the ads are determined by a function of their bids in an auction mechanism called Generalized Second-Price (GSP) auction. In contrast to traditional advertising the advertiser only has to pay if a user clicks on one of the advertiser's ads, and the Cost-Per-Click (CPC) will be not more than the bid (s)he has stated for that keyword. The position of the ad on the search result page is very important for advertisers because each location has a different Click-Through Rate (CTR), which means that on each position the likeliness of a user to click is different. The ads placed at the top of the page are much more likely to be clicked than the ads appearing further down. Every advertiser wants to be at the top of the page but has a certain maximum willingness to pay for a click on her/his ad. To determine which text ad appears at which position, the search engines use an auction mechanism called GSP. At some search engines the ad position depends only on the bid, while on others such as Google, the quality of the ads and of the associated web page also plays a significant role. In this Chapter I will focus on those search engines, which do not consider the quality but only rank the ads by bids, hence giving the advertiser with the highest bid the position at the top of the page. In this case, the CTR and the CPC are positively correlated with the position of the text ad, thus the companies face a trade-off between the amount of clicks and the cost of each click. Each advertiser has to figure out what is the value of each click for her/him, and at which bid (s)he has the highest revenue. This is difficult because GSP is not incentive-compatible, thus it is not optimal to bid truthfully and the advertiser never knows how many clicks (s)he would have got more or less if her/his bid had been slightly higher or lower. To be able to concentrate on the bid optimization problem I will first explain how the underlying auction mechanism works.

2.2 GSP: Generalized Second-Price Auctions

To explain how a generalized second-price auction works, I will use in this Section the results of [Edelman et al., 2007], [Varian, 2007], and [Varian, 2009]. Every advertiser specifies how much (s)he is willing to pay per click on her/his ad, where each ad is connected with a set of keywords. The search engines rank the advertisers from the highest to the lowest bid, such that the one with the highest bid receives the first position, the one with the second-highest bid receives the second position, and so on. It is called *second-price* because each advertiser does not have to pay her/his own bid but only the bid of the company listed in the next-lowest position of her/him plus an increment, e.g., 0.01$. This type of auction is less susceptible to gaming than the generalized *first-price* auction, where anyone has to pay exactly her/his bid. Suppose that there are four bidders competing for three slots. The values per click of advertisers A, B, C, D are 1.50$, 1.00$, 0.75$, 0.50$, respectively. Advertiser C would start with a bid of 0.51$ to ensure being at the third position, advertiser B with a bid of 0.52$ to be at the second position, and advertiser A with 0.53$ to be at the top of the page. Advertiser D would not be able to participate successfully in that auction because (s)he would spend more than her/his value per click. A and B would start to bid always 0.01$ more than the other until the bid reaches 1.00$ and B could not raise her/his bid anymore. After that, B would lower her/his bid again to 0.52$ to ensure being on the second position. Now advertiser C would see that (s)he has a chance to obtain the second position and (s)he would raise her/his bid to 0.53$ to be listed at the second position. Like A and B before, B and C would now outbid each other until the bid reaches the maximum of advertiser B.

Figure 2.2 shows this behavior, which is called 'sawtooth' pattern, on the search engine Overture. Obviously there is no pure strategy equilibrium in this version of the game and the advertisers would raise or lower their bids as often as possible.

'Under the generalized first-price auction, the bidder who could react to its competitors' moves fastest had a substantial advantage. The mechanism therefore encouraged inefficient investments in gaming the system. It also created volatile prices that in turn caused allocative inefficiencies. Google addressed these problems when it introduced its own pay-per-click system, AdWords Select, in February 2002. Google also recognized that a bidder in position i will never want to pay more than one bid increment above the bid of the advertiser in position $(i + 1)$, and Google adopted this principle in its newly-

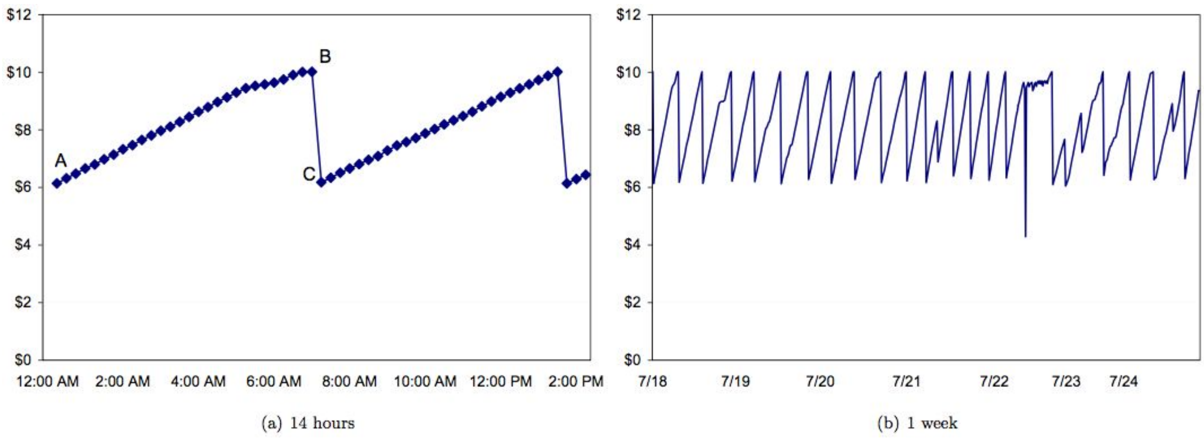

Figure 2.2: The 'sawtooth' pattern; see [Edelman et al., 2007], page 21.

designed generalized second price auction mechanism. In the simplest GSP auction, an advertiser in position i pays a price per click equal to the bid of an advertiser in position $(i + 1)$ plus a minimum increment (typically \$0.01). This second-price structure makes the market more user friendly and less susceptible to gaming.'[1]

Another important fact is that truth-telling is not a dominant strategy under GSP. This property of GSP can be seen in the following example of [Edelman et al., 2007]:
Suppose that there are 'three bidders, with values per click of \$10, \$4, and \$2, and two positions. However, the click-through rates of these positions are now almost the same: the first position receives 200 clicks per hour, and the second one gets 199. If all players bid truthfully, then bidder 1's payoff is equal to $(\$10 - \$4) * 200 = \$1200$. If, instead, (s)he shades her/his bid and bids only \$3 per click, (s)he will get the second position, and her/his payoff will be equal to $(\$10 - \$2) * 199 = \$1592 > \$1200$.'

Now we know how GSP works, the reason why Google does not use generalized first-price auctions, and that finding an optimal bid is not trivial because of the fact that truthful bidding is not a dominant strategy under GSP.

[1][Edelman et al., 2007], page 6.

3 The Model

Each advertiser has a lot of keywords relevant to the products (s)he want to sell. For all keywords (s)he has to decide on the ad text, the budget, the destination URL to which the ad is linked, and the bid with which (s)he want to take part in the auctions. All these factors have to be updated regularly, but I will focus on automizing the bid optimization.

3.1 Notation and Setup

Before I formulate the model I will explain the notation, which will be used for further analysis as used in [Abhishek and Hosanagar, 2013].

Each advertiser has a set of keywords $\mathcal{K} = \{1, 2, \ldots, K\}$ where b_k denotes the bid for keyword $k \in \mathcal{K}$ on a specific day with the assumption that the advertiser does not change the bid during the day. The impressions, i.e. the number of searches for keyword k where the advertiser's ad has been shown, are denoted by the random variable S_k, and the expected number of impressions is $\mu_k = \mathbb{E}[S_k]$.

'Every time the key phrase is searched, the advertiser's ad is placed at some position in the list of all sponsored results. Let $pos_k^{(s)}$ be the position at which the ad was shown in the sth search of the day, with the topmost position denoted position 0. Let $\delta_k^{(s)}$ be an indicator of whether a person who was searching for the keyword clicked on the advertiser's link, or not: $\delta_k^{(s)} = I(click_k^{(s)})$. The advertiser's value from a click is denoted by an independent random variable w_k. We assume that the precise value from a click is not known a priori but that its expected value $\mathbb{E}[w_k]$ is known and equals Ew_k[1]. [...] Let $v_k^{(s)}$ denote the advertiser's value from the sth impression. Furthermore, $v_k^{(s)} = \delta_k^{(s)} w_k$, i.e., it equals w_k if the user clicks on the ad or 0 otherwise. Let $\underline{b}_k^{(s)}$ be the advertiser's cost-per-click, i.e., the bid of the advertiser at the next position $pos_k^{(s)} + 1$.[2] The cost associated with impression s may then be expressed as $c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$. Because consumers do not know the bids placed by advertisers, it seems reasonable to assume that given an ad's position in the list, the probability that a person clicks on the ad does not depend on the bid of the next advertiser. That is, conditional on the position $pos_k^{(i)}$, the vector $(\underline{b}_k^{(i)}, \delta_k^{(i)})$ has independent components. We also assume that S_k is independent of other variables.'[3]

N_k denotes the number of other advertisers who are bidding for the keyword k. In our model it is assumed that N_k is known to the advertiser but unfortunately it can not be observed directly. In Google Adwords you can see a list of advertisers competing for a specific keyword. When I asked my contact person at Google, he told me that the list is not complete but certainly offers an insight into how many advertisers have the same keyword in their portfolio.[4]

'We note that the number of competitors may in reality vary a bit from one impression to another due to advertiser budget constraints of the advertisers, but we do not observe significant variation in this to warrant a random treatment for N_k.

The bids of the competitors cannot be directly observed because the auction is a sealed-bid auction. The key assumption we make is that the competitors place their bids according to

[1] The expected value Ew_k is estimated from historical data.

[2] In Google Adwords the advertiser's cost-per-click depends also on her/his quality score, which depends on different factors, e.g., the landing page experience and the ad relevance. To simplify the model the quality score of each advertiser will not be regarded and I will refer to this simplification later in Section 4.5.

[3] [Abhishek and Hosanagar, 2013], page 858.

[4] [Abhishek and Hosanagar, 2013] propose to observe the number of competitors by submitting sample queries to the search engine and observing the number of ads displayed. In my opinion this is not a good idea because in reality there are sometimes more than 20 competitors but there are not more than 11 ads shown in the Google search.

some distribution $F_k(\cdot)$, and this does not change during the estimation period. The bids of competing advertisers are based on two factors – their intrinsic valuations for a click and their competitive responses in the GSP auction. We assume that there is an underlying valuation distribution (for clicks), which when combined with the advertisers' bidding strategies gives rise to the bid distribution $F_k(\cdot)$. Finally, D denotes the advertiser's budget in a given time period of interest. Table 3.1 summarizes our notation.'[5]

k	Variable that indexes keywords
S_k	Random variable denoting number of searches for keyword k
μ_k	Expected number of search for keyword k ($\mathbb{E}[S_k]$)
(s)	Superscript to denote sth search
b_k	Bid for keyword k
$pos_k^{(s)}$	Position for keyword k in sth search. $pos_k^{(s)} = 0$ denotes the top position
$\delta_k^{(s)}$	Indicator variable for click on sth search
w_k	Random variable indicating value of a click
Ew_k	Expected value of a click on keyword k ($\mathbb{E}[w_k]$)
$v_k^{(s)}$	Value of the sth search ($v_k^{(s)} = \delta_k^{(s)} w_k$)
$\underline{b}_k^{(s)}$	The bid of the next advertiser
$c_k^{(s)}$	The cost of the sth search ($c_k^{(s)} = \delta_k^{(s)} \underline{b}_k^{(s)}$)
N_k	Number of competitors
$F_k(\cdot)$	Distribution of bids of competitors
D	Advertiser's budget

Table 3.1: Summary of notation; see [Abhishek and Hosanagar, 2013], page 859.

3.2 Model Formulation

Our goal is to choose the bid b_k for each keyword in order to maximize the expected value we gain from users clicking on the ads and buying something. Our unique constraint is the daily budget D such that the expected costs of the clicks on the ads must not exceed this given value by the advertiser. Thus our constrained optimization problem is the

[5] [Abhishek and Hosanagar, 2013], page 858.

following:

$$\max_b \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} \middle| b\right], \quad \text{s.t. } \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b\right] - D \leq 0, \tag{3.1}$$

where $b = (b_1, \dots, b_k)$ and $b_k \geq 0, \forall k \in \mathcal{K}$. We will search for a solution of this problem by using the Karush-Kuhn-Tucker (KKT) conditions, which are first order necessary conditions for a solution of an optimization problem. In addition there are some regularity conditions that have to be met. According to [Tragler, 2012], our problem is of the form

$$\begin{aligned} &\max_b f(b) \\ \text{s.t. } &g(b) \leq 0 \\ &b \geq 0 \end{aligned}$$

with $b \in \mathbb{R}^K, f : \mathbb{R}^K \to \mathbb{R}, b \mapsto \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} v_k^{(s)} \middle| b\right]$ and $g : \mathbb{R}^K \to \mathbb{R}$, $b \mapsto \mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b\right] - D$. The condition $f, g \in C^1([0, D]^K)$ is met because of the continuity of the expectation. Let b^* be a solution of our optimization problem and assume the budget constraint is binding ($g(b^*) = 0$). Because of $\dim(g(b^*)) = 1$ the regularity condition is

$$\text{Rk}(\nabla g(b^*)) = \text{Rk}\left(\frac{d\left(\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b^*\right] - D\right)}{db_1}, \dots, \frac{d\left(\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b^*\right] - D\right)}{db_K}\right) = 1,$$

which is also met. From this it follows that there is a Lagrange multiplier λ^* satisfying the KKT conditions:

$$\begin{aligned} &\exists \lambda^* \in \mathbb{R} : \\ &\nabla f(b^*) - \lambda^* \nabla g(b^*) = 0, \\ &\lambda^* g(b^*) = 0, \\ &\lambda^* \geq 0. \end{aligned}$$

[Abhishek and Hosanagar, 2013] show that the optimization problem always has a solution and we get the optimality condition

$$\forall k : \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} v_k^{(s)} \middle| b_k\right] = \lambda \frac{d}{db_k}\mathbb{E}\left[\sum_{s=1}^{S_k} c_k^{(s)} \middle| b_k\right],$$

which can be simplified further by assuming that not only the advertiser's bidding behavior but also the consumer's click behavior is i.i.d. across the searches $s = 1, \ldots, S_k$, i.e., $v_k^{(s)}$ and $c_k^{(s)}$ are i.i.d., thus our optimality condition is

$$\forall k : \frac{d}{db_k}\mathbb{E}[v_k|b_k] = \lambda \frac{d}{db_k}\mathbb{E}[c_k|b_k]. \tag{3.2}$$

Because of the difficulty of computing $\mathbb{E}[v_k|b_k]$ and $\mathbb{E}[c_k|b_k]$ I will next show how to express that condition in terms of estimable parameters. First I will focus on the expected value of a search for a keyword subject to the bid on that keyword ($\mathbb{E}[v_k|b_k]$). To be able to know the inter-relation between the bid and the value of a search, the first thing we need to know is how the probability of a click depends on the bid. This is still too general because the probability of a click depends in turn on the position of the ad, which depends on the bids of the other advertisers. Therefore the position is dependent on the advertiser's bid b_k and the bids of the other competitors. Assuming that the bids of the competitors have the distribution function $F_k(\cdot)$, we are able to determine the probability of being at position i conditional on the bid b_k:

$$P\{pos_k = i|b_k\} = \binom{N_k}{i}(1 - F_k(b_k))^i F_k(b_k)^{N_k - i}, \tag{3.3}$$

where N_k is the number of competitors. As mentioned above we consider 0 as the best position because there are 0 competitors on a higher position than our advertiser. The probability of someone bidding higher and therefore being on a better position is $1 - F_k(b_k)$. To be on the position i there have to be i competitors bidding higher. Therefore the position conditional on the bid is determined by a Bernoulli process. Now that we know the probability of being on a certain position we can go one step further and determine the probability of a click – the Click-Through Rate CTR – conditional on the position i:

$$P\{\delta_k = 1|pos_k = i\} = \frac{\alpha_k}{(\gamma_k)^i}, \tag{3.4}$$

where α_k denotes the CTR at the best position $i = 0$. α_k varies from keyword to keyword because it represents the overall attractiveness of the ad to the consumers that are searching for that keyword. Each keyword has also a specific rate of decay γ_k at which the probability of a click decays with the position. It is possible that the ad is so attractive that the decay rate is very low because a lot of consumers are attracted by that ad although it is further down than the ad of a competitor. Of course there are other factors influencing the CTR of the keyword, such as the presence of well-known competitors or whether the advertiser appears in the organic results or not. In consistency with [Katona and Sarvary, 2010] and [Ghose and Yang, 2009] α_k captures a lot of important factors and holding it constant, γ_k just incorporates the change in the CTR while changing the position.[6] As [Abhishek and Hosanagar, 2013], [Agarwal et al., 2011], [Yang and Ghose, 2010], and [Ghose and Yang, 2009] I also assume that consumer behavior is i.i.d. because it is not possible to receive user-level data on impressions and clicks from the search engine.

Now that we know the probability of being at a position conditional on the bid and the probability of a click conditional on the position, we are able to state the probability of a click conditional on the bid b_k:

$$\begin{aligned} P\{\delta_k = 1|b_k\} &= \sum_i P\{\delta_k = 1|pos_k = i\}P\{pos_k = i|b_k\} \\ &= \sum_i \frac{\alpha_k}{(\gamma_k)^i} \begin{pmatrix} N_k \\ i \end{pmatrix} (1 - F_k(b_k))^i F_k(b_k)^{N_k - i} \\ &= \alpha_k \gamma_k^{-N_k} (1 + (\gamma_k - 1)F_k(b_k))^{N_k}. \end{aligned} \tag{3.5}$$

Using Equation (3.5) the expected value of an impression conditional on the bid, $\mathbb{E}[v_k|b_k]$,

[6] A very important fact that has to be mentioned is that the first three positions are on the top of the organic results and the other positions are on the right side. Therefore consumers are much more likely to click on the ads on the top than on the ones on the side. In addition my experience tells me that if the ad is at the top, the ad text is much more relevant to the probability of clicking on the ad than the exact position (0, 1, or 2). A possible solution would be to treat the first three positions equally, e.g., $P\{\delta_k = 1|pos_k = i\} = \alpha_k(\gamma_k)^{-\max\{0,i-3\}}$.

is given by

$$\begin{aligned}\mathbb{E}[v_k|b_k] &= \mathbb{E}[\delta_k w_k|b_k] = P\{\delta_k = 1|b_k\}\mathbb{E}[w_k] \\ &= \alpha_k \gamma_k^{-N}(1 + (\gamma_k - 1)F_k(b_k))^{N_k} E w_k. \end{aligned} \tag{3.6}$$

The derivate of that expected value with respect to b_k can be written as

$$\frac{d}{db_k}\mathbb{E}[v_k|b_k] = \alpha_k N_k \gamma_k^{-N_k}(\gamma_k - 1) f_k(b_k)(1 + (\gamma_k - 1)F_k(b_k))^{N_k - 1} E w_k, \tag{3.7}$$

where $f_k(\cdot)$ is the density function of the probability with the distribution function $F_k(\cdot)$.

Now that we are able to express the expected value v_k of an impression conditional on the bid b_k in terms of estimable parameters, we will focus on deriving such an expression for the expected cost of an impression conditional on the bid, $\mathbb{E}[c_k|b_k]$. The cost of an impression depends on the probability of someone clicking on the ad and on the bid of the next advertiser. As we already know $P\{\delta_k = 1|b_k\}$ stated in Equation (3.5), our next goal is to know the distribution of the bid of the next advertiser, again first conditional on the bid and the position so then we can characterize the distribution function conditional on the bid and the indicator variable for click δ_k. The first distribution function we want to express in terms of estimable parameters is

$$\begin{aligned} F_k(\underline{b}_k = x|b_k, pos_k = i) &= P\{\underline{b}_k < x|b_k, pos_k = i\} \\ &= \frac{P\{\underline{b}_k < x, pos_k = i|b_k\}}{P\{pos_k = i|b_k\}}. \end{aligned} \tag{3.8}$$

Obviously the bid of the next advertiser is lower than b_k because otherwise (s)he would not be further down than the advertiser with bid b_k, thus $x < b_k$. As stated in Equation (3.8), the probability of the next bid being less than $x < b_k$ conditional on the bid b_k and the position i is equal to the probability of i competitors bidding higher than b_k and $N_k - i$ of them bidding less than $x < b_k$ divided by the probability of being at position

i, which we already know from Equation (3.3), thus

$$\begin{aligned}
&\frac{P\{\underline{b}_k < x, pos_k = i | b_k\}}{P\{pos_k = i | b_k\}} = \\
&= \frac{\begin{pmatrix} N_k \\ i \end{pmatrix} (1 - F_k(b_k))^i F_k(x)^{N_k - i}}{\begin{pmatrix} N_k \\ i \end{pmatrix} (1 - F_k(b_k))^i F_k(b_k)^{N_k - i}} \\
&= \frac{F_k(x)^{N_k - i}}{F_k(b_k)^{N_k - i}} \\
&\Rightarrow F_k(\underline{b}_k | b_k, pos_k = i) = \left(\frac{F_k(\underline{b}_k)}{F_k(b_k)} \right)^{N_k - i}.
\end{aligned} \tag{3.9}$$

As mentioned above, the bid of the next advertiser is always less than b_k, and no matter which value b_k is, by definition the next advertiser will always be one position further down than the advertiser bidding b_k, because otherwise (s)he would not be the next advertiser. Moreover, the bid $\underline{b}_k$ is always the bid of the next advertiser no matter who it is. The consumer only cares about the position i because (s)he is not able to know how much each advertiser is bidding and therefore her/his click can not be dependent on $\underline{b}_k$. Considering that fact, the distribution of the bid of the next advertiser conditional on the bid b_k and the fact that the ad was clicked ($\delta_k = 1$) is

$$\begin{aligned}
F_k(\underline{b}_k = x|b_k, \delta_k = 1) &= P\{\underline{b}_k < x|b_k, \delta_k = 1)\\
&= \sum_{i=0}^{N_k} P\{\underline{b}_k < x|b_k, \delta_k = 1, pos_k = i\} \cdot P\{pos_k = i|b_k, \delta_k = 1\}\\
&\overset{\delta_k=1 \text{ independent of } \underline{b}_k}{=} \sum_{i=0}^{N_k} F_k(x|b_k, pos_k = i) \cdot \frac{P\{pos_k = i, \delta_k = 1|b_k\}}{P\{\delta_k = 1|b_k\}}\\
&\overset{(3.9)}{=} \sum_{i=0}^{N_k} \left(\frac{F_k(x)}{F_k(b_k)}\right)^{N_k-i} \cdot \frac{P\{\delta_k = 1|b_k, pos_k = i\}P\{pos_k = i|b_k\}}{P\{\delta_k = 1|b_k\}}\\
&\overset{(3.3),(3.4),(3.5)}{=} \sum_{i=0}^{N_k} \left(\frac{F_k(x)}{F_k(b_k)}\right)^{N_k-i} \cdot \frac{\frac{\alpha_k}{(\gamma_k)^i}\begin{pmatrix} N_k \\ i \end{pmatrix}(1-F_k(b_k))^i F_k(b_k)^{N_k-i}}{\alpha_k\gamma_k^{-N_k}(1+(\gamma_k-1)F_k(b_k))^{N_k}}\\
&= \sum_{i=0}^{N_k} \begin{pmatrix} N_k \\ i \end{pmatrix} (\gamma_k F_k(x))^{N_k-i} \frac{1-F_k(b_k))^i}{(1+(\gamma_k-1)F_k(b_k))^{N_k}}\\
&= \frac{(1-F_k(b_k)+\gamma_k F_k(x))^{N_k}}{(1+(\gamma_k-1)F_k(b_k))^{N_k}}. \qquad (3.10)
\end{aligned}$$

Now we are ready to express the expected cost c_k of an impression conditional on the bid b_k with the help of Equation (3.5) and Equation (3.10):

$$\begin{aligned}
\mathbb{E}[c_k|b_k] &= \mathbb{E}[\delta_k \underline{b}_k|b_k]\\
&= \mathbb{E}[\underline{b}_k|b_k, \delta_k = 1]P\{\delta_k = 1|b_k\}\\
&= \alpha_k\gamma_k^{-N_k}(1+(\gamma_k-1)F_k(b_k))^{N_k} \int_0^{b_k} \underline{b}_k d\left(\frac{1-F_k(b_k)+\gamma_k F_k(\underline{b}_k)}{1+(\gamma_k-1)F_k(b_k)}\right)^{N_k}\\
&= \alpha_k\gamma_k^{-N_k} \int_0^{b_k} \underline{b}_k d(1-F_k(b_k)+\gamma_k F_k(\underline{b}_k))^{N_k}\\
&\overset{\text{integration by parts}}{=} \alpha_k\gamma_k^{-N_k}\Bigg(b_k(1+(\gamma_k-1)F_k(b_k))^{N_k}\\
&\qquad - \int_0^{b_k}(1-F_k(b_k)+\gamma_k F_k(\underline{b}_k))^{N_k} d\underline{b}_k\Bigg). \qquad (3.11)
\end{aligned}$$

To express Equation (3.2) we derive (3.11) and get

$$\begin{aligned}
&\frac{d}{db_k}\mathbb{E}[c_k|b_k] = \\
&= \alpha_k\gamma_k^{-N_k}\Bigg((1+(\gamma_k-1)F_k(b_k))^{N_k} + N_k(\gamma_k-1)b_kf_k(b_k)(1+(\gamma_k-1)F_k(b_k))^{N_k-1} \\
&\quad - (1-F_k(b_k)+\gamma_kF_k(b_k))^{N_k}\cdot 1 + N_kf_k(b_k)\int_0^{b_k}(1-F_k(b_k)+\gamma_kF_k(\underline{b}_k))^{N_k-1}d\underline{b}_k\Bigg) \\
&= \alpha_kN_k\gamma_k^{-N_k}f_k(b_k)\Bigg((\gamma_k-1)b_k(1+(\gamma_k-1)F_k(b_k))^{N_k-1} \\
&\quad + \int_0^{b_k}(1-F_k(b_k)+\gamma_kF_k(\underline{b}_k))^{N_k-1}d\underline{b}_k\Bigg).
\end{aligned} \tag{3.12}$$

Finally we are able to express the optimality condition (3.2) in terms of estimable parameters using Equations (3.7) and (3.12):

$$\forall k: \frac{d}{db_k}\mathbb{E}[v_k|b_k] = \lambda\frac{d}{db_k}\mathbb{E}[c_k|b_k]$$

$$\Leftrightarrow$$

$$\forall k: const = \frac{1}{\lambda} = \frac{1}{Ew_k}\left(b_k + \frac{\int_0^{b_k}(1-F_k(b_k)+\gamma_kF_k(\underline{b}_k))^{N_k-1}d\underline{b}_k}{(\gamma_k-1)(1+(\gamma_k-1)F_k(b_k))^{N_k-1}}\right), \tag{3.13}$$

where *const* is a constant, which we will alter while computing the optimal bid.

We want a unique bid b_k^* to satisfy the optimality condition in Equation (3.13) for each keyword k. A sufficient condition to have a unique bid b_k^* is that $\Psi(b_k) = b_k + \Big(\int_0^{b_k}(1-F_k(b_k)+\gamma_kF_k(\underline{b}_k))^{N_k-1}d\underline{b}_k\Big)/\Big((\gamma_k-1)(1+(\gamma_k-1)F_k(b_k))^{N_k-1}\Big)$ is monotonically increasing. To simplify the illustration of $\Psi'(b_k)$ we will use $h_{N_k}(b_k) = \int_0^{b_k}(1-F_k(b_k)+\gamma_kF_k(x))^{N_k}dx$ and $g_{N_k}(b_k) = (1+(\gamma_k-1)F_k(b_k))^{N_k}$. Thus,

$$\begin{aligned}
\Psi(b_k) &= b_k + \frac{h_{N_k-1}(b_k)}{(\gamma_k-1)g_{N_k-1}(b_k)} \\
\Psi'(b_k) &= 1 + \frac{h'_{N_k-1}(b_k)}{(\gamma_k-1)g_{N_k-1}(b_k)} - \frac{h_{N_k-1}(b_k)g'_{N_k-1}(b_k)}{(\gamma_k-1)g^2_{N_k-1}(b_k)} \\
&= \frac{g_{N_k-1}(b_k)((\gamma_k-1)g_{N_k-1}(b_k)+h'_{N_k-1}(b_k)) - h_{N_k-1}(b_k)g'_{N_k-1}(b_k)}{(\gamma_k-1)g^2_{N_k-1}(b_k)}.
\end{aligned}$$

For $(\gamma_k-1)>0$ we get $\Psi'(b_k)>0$, if

$$g_{N_k-1}(b_k)((\gamma_k-1)g_{N_k-1}(b_k)+h'_{N_k-1}(b_k)) - h_{N_k-1}(b_k)g'_{N_k-1}(b_k) > 0. \tag{3.14}$$

We can reformulate Equation (3.14) getting ratios of the form of $h_{N_k}(b_k)/g_{N_k}(b_k)$, which is by intuition decreasing in N_k for all $N_k \geq 2$. This intuition is reasonable because h_{N_k} is an integral over $[0, b_k]$ of the function $(1 - F_k(b_k) + \gamma_k F_k(x))^{N_k}$ that is monotonically increasing in x, is equal to g_{N_k} at $x = b_k$ and is less than g_{N_k} at $x \in [0, b_k)$, thus if N_k increases, g_{N_k} will grow faster than h_{N_k}. Taking a sample Weibull distribution[7] $F(x; \theta, \lambda) = 1 - \exp\left\{-(x/\lambda)^\theta\right\}$ with $\lambda = 1.59, \theta = 1.37$, and $\gamma = 1.42$, we see in Figure 3.1 that the ratio $h_{N_k}(b_k)/g_{N_k}(b_k)$ decreases as N_k increases. To reformulate Equation

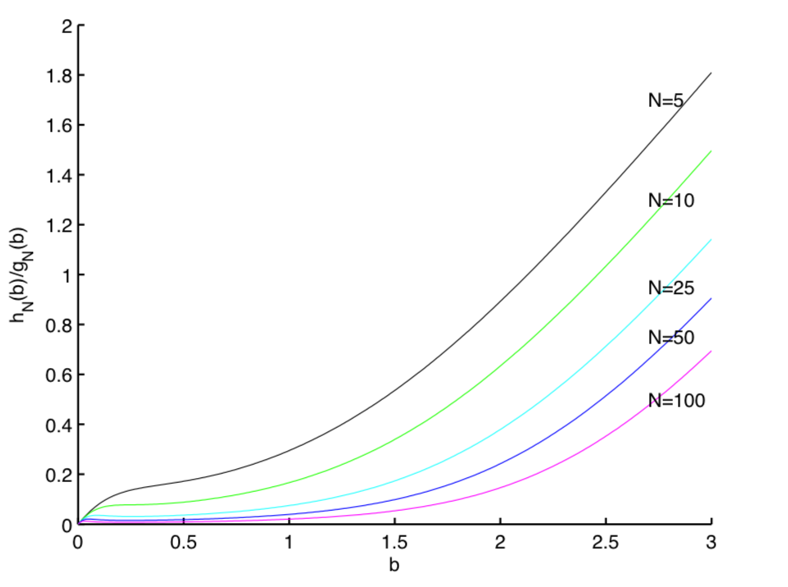

Figure 3.1: Illustration of the fact that h_{N_k}/g_{N_k} decreases as N_k increases; see [Abhishek and Hosanagar, 2013], Appendix, page 38.

(3.14) in the aforementioned manner we need to express $g'_{N_k}(b_k)$ and $h'_{N_k}(b_k)$:

$$\begin{aligned} g'_{N_k}(b_k) &= \frac{d}{db_k}\left((1 + (\gamma_k - 1)F_k(b_k))^{N_k}\right) \\ &= N_k f_k(b_k)(\gamma_k - 1)(1 + (\gamma_k - 1)F_k(b_k))^{N_k - 1} \\ &= N_k f_k(b_k)(\gamma_k - 1)g_{N_k - 1}(b_k), \end{aligned}$$

[7]We will later assume that $F_k(\cdot)$ is a Weibull distribution for all k.

$$
\begin{aligned}
h'_{N_k}(b_k) &= \frac{d}{db_k}\left(\int_0^{b_k}(1-F_k(b_k)+\gamma_k F_k(x))^{N_k}dx\right)\\
&= (1+(\gamma_k-1)F_k(b_k))^{N_k}\cdot 1 - N_k f_k(b_k)\int_0^{b_k}(1-F_k(b_k)+\gamma_k F_k(x))^{N_k-1}dx\\
&= g_{N_k}(b_k) - N_k f_k(b_k) h_{N_k-1}(b_k).
\end{aligned}
$$

Equation (3.14) can be written as

$$
\begin{aligned}
g_{N_k-1}(b_k)\big((\gamma_k-1)g_{N_k-1}(b_k)+g_{N_k-1}(b_k)-(N_k-1)f_k(b_k)h_{N_k-2}(b_k)\big)&\\
-\,h_{N_k-1}(b_k)(N_k-1)f_k(b_k)(\gamma_k-1)g_{N_k-2}(b_k) &> 0\\
&\Leftrightarrow\\
g_{N_k-1}(b_k)\big(\gamma_k g_{N_k-1}(b_k)-(N_k-1)f_k(b_k)h_{N_k-2}(b_k)\big)&\\
> h_{N_k-1}(b_k)(N_k-1)f_k(b_k)(\gamma_k-1)g_{N_k-2}(b_k)&\\
&\Leftrightarrow\\
\gamma_k g_{N_k-1}(b_k) > \frac{h_{N_k-1}(b_k)(N_k-1)f_k(b_k)(\gamma_k-1)g_{N_k-2}(b_k)}{g_{N_k-1}(b_k)}&\\
+(N_k-1)f_k(b_k)h_{N_k-2}(b_k)&\\
&\Leftrightarrow\\
\frac{\gamma_k g_{N_k-1}(b_k)}{g_{N_k-2}(b_k)} > \frac{h_{N_k-1}(b_k)(N_k-1)f_k(b_k)(\gamma_k-1)g_{N_k-2}(b_k)}{g_{N_k-1}(b_k)g_{N_k-2}(b_k)}&\\
+\frac{(N_k-1)f_k(b_k)h_{N_k-2}(b_k)}{g_{N_k-2}(b_k)}&\\
&\Leftrightarrow\\
\gamma_k(1+(\gamma_k-1)F_k(b_k)) > (N_k-1)f_k(b_k)\left((\gamma_k-1)\frac{h_{N_k-1}(b_k)}{g_{N_k-1}(b_k)}+\frac{h_{N_k-2}(b_k)}{g_{N_k-2}(b_k)}\right).&
\end{aligned}
$$

As mentioned above, the ratio $h_{N_k}(b_k)/g_{N_k}(b_k)$ is decreasing in N_k and therefore $h_{N_k-2}(b_k)/g_{N_k-2}(b_k) \geq h_{N_k-1}(b_k)/g_{N_k-1}(b_k)$ implying $\Psi'(b_k) > 0$ and thus the existence of a unique bid b_k^* satisfying the optimality condition for keyword k if

$$\begin{aligned}\gamma_k(1+(\gamma_k-1)F_k(b_k)) > \gamma_k(N_k-1)f_k(b_k)\frac{h_{N_k-2}(b_k)}{g_{N_k-2}(b_k)}& \\ \Leftrightarrow& \\ \gamma_k > 1+\frac{1}{F_k(b_k)}\left(f_k(b_k)(N_k-1)\frac{h_{N_k-2}(b_k)}{g_{N_k-2}(b_k)}-1\right).& \qquad (3.15)\end{aligned}$$

Equation (3.15) shows that for the existence of a unique bid b_k^* satisfying the optimality condition for keyword k the parameter γ_k, which is the rate of decay of the Click-Through Rate CTR with respect to the position pos_k, has to be high enough. For important distributions like the Weibull, Gamma, and Log-Normal [Abhishek and Hosanagar, 2013] numerically find that $\Psi(b_k)$ is monotonically increasing in b_k and that there always exists a unique bid, as demonstrated for some sample parameters in Figure 3.2.

Having a parametrical form of the optimality condition we can use Equation (3.13) to compute the optimal bids. We assume that for our distribution Funktion $F_k(\cdot)$ Equation (3.15) holds, because that condition is satisfied for several common distributions and a wide range of parameters.

To compute the optimal bids we do not only need the optimality condition but also the budget constraint $\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b\right] = D$. We assumed that the consumer click and the competitor bidding behavior is i.i.d. across ad impression and thus the budget constraint is given by

$$\begin{aligned}\mathbb{E}\left[\sum_k \sum_{s=1}^{S_k} c_k^{(s)} \middle| b\right] &= D \\ &\overset{\mu_k=\mathbb{E}[S_k]}{\Leftrightarrow} \\ \sum_k \mu_k \mathbb{E}[c_k|b_k] &= D \\ &\Leftrightarrow \\ \sum_k \mu_k\alpha_k\gamma_k^{-N_k}\Bigg(b_k(1+(\gamma_k-1)F_k(b_k))^{N_k}& \\ -\int_0^{b_k}(1-F_k(b_k)+\gamma_k F_k(\underline{b}_k))^{N_k}d\underline{b}_k\Bigg) &= D. \qquad (3.16)\end{aligned}$$

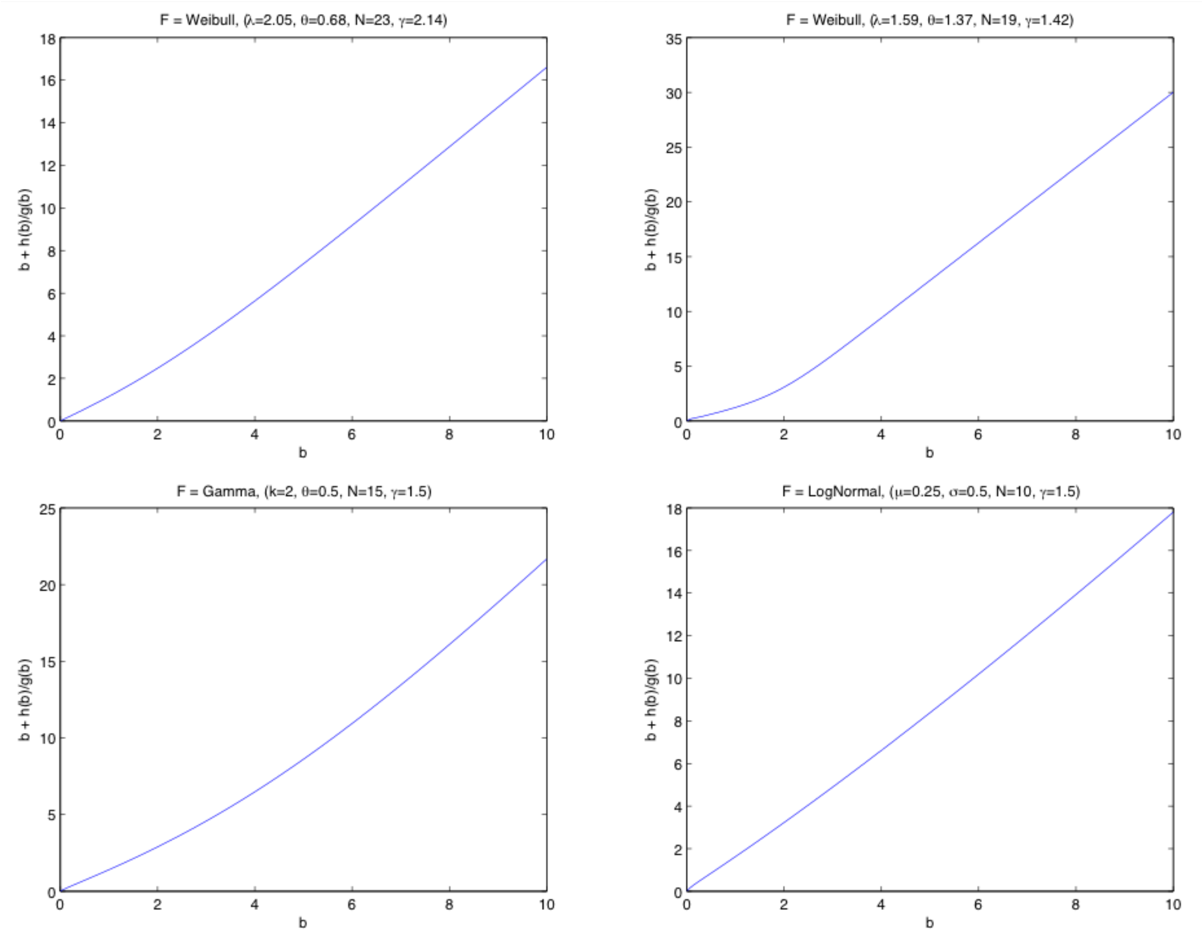

Figure 3.2: $\Psi(b_k)$ for Weibull, Gamma, and Log-Normal distributions; see [Abhishek and Hosanagar, 2013], Appendix, page 39.

The unique bid b has to satisfy not only the optimality condition $const = \frac{1}{Ew_k}\left(b_k + \frac{\int_0^{b_k}(1-F_k(b_k)+\gamma_k F_k(\underline{b}_k))^{N_k-1}d\underline{b}_k}{(\gamma_k-1)(1+(\gamma_k-1F_k(b_k))^{N_k-1}}\right)$ for a given $const$ and for all k but also the budget constraint of Equation (3.16). We can compute the optimal bid by altering $const$ in Equation (3.13). For a chosen value for $const$ we compute the bids b_k for all k satisfying the optimality condition. Using these bids we calculate the left side of Equation (3.16) and get the total daily costs D_b. If $D_b < D$ we have to increase the constant $const$, if $D_b > D$ we decrease the constant in Equation (3.13) and compute a new bid b, which satisfies the optimality condition with the new $const$. If the total costs D_b are sufficiently close to the budget D we stop that process and consider the actual bid as our optimal bid b^*.

Having the procedure to compute the optimal bid b^* we still need the estimates for the Click-Through Rate (CTR) at the top position α_k, the expected daily impressions μ_k, the rate at which CTR decays with position γ_k, the expected value per click Ew_k, the distribution function $F_k(\cdot)$ of the bids of the competitors, and the number of competitors N_k for each keyword k to be able to get the desired result. An estimate for N_k can be seen – as mentioned before – on a list in Google Adwords. For μ_k and Ew_k we have daily aggregates in Google Adwords, thus they are estimated directly by computing a sample mean of our data. In the following Section, I will show how to estimate the parameters α_k, γ_k, and the distribution function $F_k(\cdot)$.

3.3 Generalized Method of Moments (GMM)

The Generalized Method of Moments (GMM) estimator was first introduced by Hansen in 1982 [Hansen, 1982]. Some alternative GMM estimators were introduced 1996 by Hansen, Heaton, and Yaron in [Hansen et al., 1996]. It is used to estimate parameters – often of econometric models – by making the use of orthogonality conditions 'in which expected cross products of unobservable disturbances and functions of observable variables are equated to zero. Heuristically, identification requires at least as many orthogonality conditions as there are coordinates in the parameter vector to be estimated. The unobservable disturbances in the orthogonality conditions can be replaced by an equivalent expression involving the true parameter vector and the observed variables. Using the method of moments, sample estimates of the expected cross products can be computed for any element in an admissible parameter space. A GMM estimator of the true parameter vector is obtained by finding the element of the parameter space that sets linear combinations of the sample cross products as close to zero as possible'[8].

[8] [Hansen, 1982], page 1029.

The notation of the above described method to estimate parameters is the following. We are supposed to have a p-dimensional stochastic process and we observe a realization $(x_t, t \in \mathcal{T})$ of that process with $\mathcal{T} = \{1, \dots, T\}$, which means that our sample size is T. The goal is to estimate certain parameters and let β_0 be the q-dimensional vector of these parameters with $\beta_0 \in S \subseteq \mathbb{R}^q$, where S is called the parameter space. Let $\xi_t(\beta) := \xi(x_t, \beta) : \mathbb{R}^p \times S \to \mathbb{R}^u$ be the unobservable disturbances and $z_t := z(x_t) : \mathbb{R}^p \to \mathbb{R}^v$ be the function of observable variables, where x_t are the observable variables of the process and $u+v = r$. The moment conditions, where expected cross products are equated to zero, are needed to estimate the parameters and state that $\forall t \in \mathcal{T} : \mathbb{E}f(x_t, \beta_0) = \mathbb{E}[\xi_t(\beta_0) \otimes z_t] = 0$ with $f : \mathbb{R}^p \times S \to \mathbb{R}^r$ and $r \geq q$ because we need at least as many orthogonality conditions as there are coordinates in the parameter vector. [Hall, 2010] shows that $\mathbb{E}[f(x_t, \beta_0)] = 0$ 'must be a unique property of' β_0, thus at any other value of β the expected cross products must not be zero. The method of moments uses the law of large numbers such that the sample estimate for the expected cross products $\forall t \in \mathcal{T} : \mathbb{E}f(x_t, \beta)$ is given by $\mu_T(\beta) = T^{-1} \sum_{t=1}^{T} f(x_t, \beta)$.

If we had as many orthogonality conditions as there are coordinates in the parameter vector $(r = q)$ we could solve $\mu_T(\beta) = 0$ for β and would have the estimate of the parameter. But in our case we will have more conditions than coordinates $(r > q)$ and therefore we will set linear combinations of the sample cross products as close to zero as possible, thus the GMM estimator is given by

$$\hat{\beta} = \arg\min_{\beta \in S} \mu_T(\beta)' W \mu_T(\beta),$$

where W is a weighting matrix, the choice of which is a crucial factor for the asymptotic properties of the GMM estimator. As suggested in [Hall, 2010] and in [Hansen et al., 1996], we will use the inverse of a consistent estimator of the covariance matrix $S(\beta_0) = \lim_{T\to\infty} \mathbb{V}[\sqrt{T} \cdot \mu_T(\beta_0)]$ for the computation of the weighting matrix W. [Hall, 2010] states that the 'variance can be estimated by a member of the class of *heteroscedasticity autocorrelation covariance (HAC)* estimators defined as

$$\hat{S}_{HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega(i; b_T)(\hat{\Gamma}_i + \hat{\Gamma}_i') \tag{3.17}$$

where $\hat{\Gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{f}_t \hat{f}'_{t-j}$, ' $\hat{f}_t = f(x_t, \hat{\beta}_T)$, '$\omega(\cdot)$ is known as the *kernel*, and b_T is known as the *bandwidth*. The kernel and bandwidth must satisfy certain restrictions to ensure $\hat{S}_{HAC}$ is both consistent and positive semi-definite. As an illustration, Newey and West' [Newey and West, 1987] 'propose the use of the kernel $\omega(i, b_t) = \{1 - i/(b_T + 1)\}\mathcal{I}\{i \leq b_T\}$ where $\mathcal{I}\{i \leq b_T\}$ is an indicator variable that takes the value of one if $i \leq b_T$ and zero otherwise.' [Newey and West, 1987] suggest the bandwidth b_T to be a function of the sample size T with $\lim_{T\to\infty} b_T = +\infty$ and $b_T = o(T^{1/4})$.[9] As we can see, the consistent estimator $\hat{S}_{HAC}$ of Equation (3.17) is dependent on an estimator $\hat{\beta}$ of β_0, which has to be consistent. Thus, to be able to estimate the covariance matrix, which we need to estimate β_0, we need a consistent estimate for β_0. As is apparent we will use a multi-step procedure to estimate the vector of parameters and let $V(\hat{\beta}_T)$ denote the consistent estimator $\hat{S}_{HAC}$ obtained using the estimator $\hat{\beta}_T$.

Summarizing the aforementioned results, the consistent estimator $V(\hat{\beta}_T^j)$ of the covariance – where j denotes the step of the estimation procedure – is defined as:

$$\begin{aligned} V(\hat{\beta}_T^j) &= \hat{\Gamma}_0 + \sum_{i=1}^{b_T} \left(1 - \frac{i}{b_T + 1}\right) \left(\hat{\Gamma}_i^{j-1} + (\hat{\Gamma}_i^{j-1})'\right), \\ \text{with} \quad \hat{\Gamma}_i^j &= \frac{1}{T} \sum_{t=i+1}^{T} f(x_t, \hat{\beta}_T^j)' f(x_{t-i}, \hat{\beta}_T^j) \text{ and } b_T = \text{floor}(T^{2/9}). \end{aligned} \tag{3.18}$$

[Hansen et al., 1996] compare three alternative GMM estimators: the two-step, the iterative, and the continuous-updating estimator. In the first two variants the identity matrix is used to weight the moment conditions in the first step, such that

$$\hat{\beta}_T^1 = \arg\min_{\beta \in S} \mu_T(\beta)' I_r \mu_T(\beta).$$

The two-step estimator is given by $\hat{\beta}_T^2$, which uses the weighting matrix $V(\hat{\beta}_T^1)$, so that $\hat{\beta}_T^2 = \arg\min_{\beta \in S} \mu_T(\beta)' V(\hat{\beta}_T^1) \mu_T(\beta)$.

To compute the iterative estimator (which will be denoted by $\hat{\beta}_T^\infty$), $V(\beta)$ is reestimated again and again to compute every time a new estimator $\hat{\beta}_T^j$ using the weighting matrix

[9] In my calculations I chose $b_T = \text{floor}(T^{2/9})$, which have the required properties.

$V(\hat{\beta}_T^{j-1})$. Following [Hansen et al., 1996] this procedure is repeated until there exists an $\epsilon > 0$, which is small enough and $\hat{\beta}_T^{j} - \hat{\beta}_T^{j-1} < \epsilon$, or until the number of iterations reaches a certain predefined value. [Hansen et al., 1996] describe the continous-updating estimator as follows: 'Instead of taking the weighting matrix as given in each step of the GMM estimation, we also consider an estimator in which the covariance matrix is continuously altered as' β 'is changed in the minimization'. As suggested by [Abhishek and Hosanagar, 2013], I will use the iterative GMM estimator because it derives better results than the two-step estimator and is not as difficult to implement as the continuous-updating estimator.

4 Testing with Real Data

I will test the bid optimization model proposed by [Abhishek and Hosanagar, 2013] using a data set from a ticket agency, which has a store located in the first district of Vienna and sells tickets for more than 25,000 events (concerts, operas, musicals, . . .) worldwide – thanks to the internet and online advertising. The ticket agency does not only sell tickets for events worldwide but also serves costumers worldwide. Of course one can book tickets by phone or by going to the office, but the majority buys their tickets online, many of them after searching for a certain event like 'John Legend tickets' or for general terms like 'concerts in vienna' on Google. The set of keywords of the ticket agency has to be always up to date so that the potential customer sees the ad of the ticket agency on Google, clicks on it, finds what (s)he was searching for, and buys the desired tickets. Thus, for each event the ticket agency creates a so-called ad group with an ad, a destination URL to which the ad is linked, a set of keywords, and for each keyword the bid with which it wants to take part in the auctions.

For an imaginary event called 'Optimal Bidding', which would have taken place in the Vienna State Opera in May 2014, typical keywords would be 'Optimal Bidding Vienna', 'Optimal Bidding May 2014' or 'Optimal Bidding Vienna State Opera'. A possible ad for that event is shown in Figure 4.1.

For each of these keywords the ticket agency has to decide on the bid and adapt it regularly, which is quite time-consuming regarding the quantity of events on their homepage and thus the resulting quantity of keywords they have to manage. To support the advertisers with this challenge, Google Adwords offers different bidding strategies besides

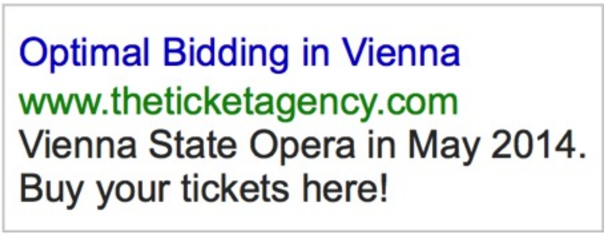

Figure 4.1: Example for an ad of the imaginary event 'Optimal Bidding'

the common cost-per-click bidding where the user determines for each keyword the maximum Cost-Per-Click (max. CPC)[1]: Enhanced Cost-Per-Click (ECPC), automatic CPC bidding, and many more. Google describes ECPC as 'a bidding feature that raises your bid for clicks that seem more likely to lead to a sale or conversion[2] on your website. That helps you get more value from your ad budget'[3]. Google also states that 'Automatic bidding allows you to put your bidding on autopilot with the goal of getting the most possible clicks within your budget. [...] Automatic bidding is ideal for advertisers who don't want to spend a lot of time setting bids, but would like to get the most clicks possible for their ads within their budget.'[4] Because of Google being a profit oriented company it is difficult to know how well they really optimize the bids if an advertiser chooses automatic bidding or ECPC. That's why the ticket agency handles as much keywords as possible on manual cost-per-click basis and lets Google support them only on the remaining keywords using different bidding strategies.

4.1 Data Set

In this Chapter, I will test the bid optimization on a data set containing 255 keywords that the ticket agency uses to promote their tickets on Google. This data set contains

[1] In Google Adwords the bid on a keyword is called maximum Cost-Per-Click (max. CPC).

[2] A conversion is a desired action that the visitor takes like signing up for a newsletter or filling in a registration form.

[3] https://support.google.com/adwords/answer/2464964?hl=en (Visited on August 6, 2014)

[4] https://support.google.com/adwords/answer/2470106?hl=en (Visited on August 6, 2014)

only keywords, which are related to event venues like 'vienna golden hall' or 'staatsoper budapest', to genres combined with a city like 'musical wien' or 'oper in prag', or to events, which take place regularly like 'spanish riding school', 'vienna boys choir' or 'palau de la musica catalana'. The bidding for keywords of events which take place once, like the concert of John Legend as a part of her/his 'All of me'-Tour in Vienna, has to be treated differently because the demand for the tickets changes a lot from the start of the presale to the day of the event. Thus an automatic bid optimization model like the one of [Abhishek and Hosanagar, 2013] cannot be used, because it needs previous data collection to estimate the bids, which obviously cannot be collected before the start of the presale and changes a lot during the sale of the tickets (between the start of the presale and the day of the event).

Google Adwords submits a lot of data regarding each keyword in the form of daily summary measures, but does not provide the advertisers with the information about clicks, position, or CPC for each individual ad impression. The daily data for each keyword, which we will use to optimize the bids, consists of the following fields:

$$(id, t, avgpos, avgcpc, i, cl, ctr, b, w, c, n),$$

where

- id identifies the keyword,
- t is the variable that indexes the date,
- $avgpos$ shows the average position during the day,
- $avgcpc$ is the average cost-per-click on the day,
- i is the number of impressions during the day,
- cl is the number of clicks during the day,
- ctr $(= cl/i)$ is the click-through rate on the day,

- b is the bid, which the advertiser has chosen for that day,
- w gives the average value per click on the day,
- c are the total costs for the clicks cl, and
- n is the number of competitors bidding for keyword id.

The 255 keywords belong to 35 product categories like event venues, genres (combined with a city), and regular events. To test the analytical model of Chapter 3, I randomly divided the 35 product categories into two distinct groups: a control group and an experimental group. The bids of the control group are computed the same way as before and this group is 'used to account for any time trends that might enter the analysis due to seasonality in retail, search engine design changes, and other such factors.'[5] The bids of the experimental group are computed by the bidding policy of Chapter 3.

We further divide our data set into three periods: the 'before', the estimation, and the 'after' period. Our 'before' period runs from February 2, 2014 until May 16, 2014. During this period the ticket agency computed their bids as usual. We can see in Table 4.1 that the data of the two randomly divided groups are quite similar during this period.[6]

	Both Groups	Control Group	Experimental Group
Impressions	i_{c+e}	$50.1\% \cdot i_{c+e}$	$49.9\% \cdot i_{c+e}$
Clicks	cl_{c+e}	$38.2\% \cdot cl_{c+e}$	$61.8\% \cdot cl_{c+e}$
CTR	ctr_{c+e}	$ctr_{c+e} - 1.63pp$	$ctr_{c+e} + 1.63pp$
Avg. CPC	$avgcpc_{c+e}$	$85.7\% \cdot avgcpc_{c+e}$	$114.3\% \cdot avgcpc_{c+e}$
Avg. Position	$avgpos_{c+e}$	$avgpos_{c+e} - 0.5$	$avgpos_{c+e} + 0.5$
Avg. RPC[7]	w_{c+e}	$110.7\% \cdot w_{c+e}$	$89.3\% \cdot w_{c+e}$

Table 4.1: Summary for the different groups in the 'before' period

[5] [Abhishek and Hosanagar, 2013], page 861.

[6] Differences between two percentages are denoted by pp (percentage point).

[7] Let Avg. RPC be the Average Revenue Per Click, thus the revenue that the firm yields in average

The data of the 'before' period are used 'to compute the expected value *per-click* (Ew) and the expected daily impressions (μ) for each keyword'[8] by taking the daily aggregates and computing the sample mean over this period of 90 days.

During the estimation period [Abhishek and Hosanagar, 2013] suppose to 'submit random bids for the keywords. [...] The bids are uniformly drawn from \$0.10 $\times$ $[1, 30]$.', where this period runs from May 18, 2014 until June 30, 2014. In the case of the ticket agency this is not useful because the company is satisfied with their bids, clicks, and average revenue per click, so they comprehensibly do not want to risk a decrease in the profitability of their Google Adwords campaign. However, the ticket agency agreed to use exclusively the manual CPC strategy and to change the bids weekly during this period of time to help us with the identification of the parameters with GMM. As one can see in Table 4.2, the variation in the position, avg. CTR, and avg. CPC did not change a lot (and even decreased for the average position), but I think that for data collection it is more important to observe the variation using various but realistic bids than using randomly drawn bids of such a great range from \$0.10 to \$3.00.

	'before' period	estimation period
S.D.[9] of *pos*	0.89	0.80
S.D. of CTR	0.11	0.13
S.D. of CPC	0.19	0.20

Table 4.2: Variation for experimental group in 'before' and estimation periods

The 'after' period runs from July 7, 2014 until August 4, 2014. In this last period, the bids are computed by estimating the parameters with the generalized method of moments and optimizing the bids using Equations (3.13) and (3.16) of Chapter 3. The data from this period are used to determine whether the bidding policies proposed by

when someone clicks on their ad.

[8][Abhishek and Hosanagar, 2013], page 862.

[9]Let S.D. denote the Standard Deviation of the sample given by $\sqrt{(T-1)^{-1}\sum_{t=1}^{T}(x_t - \overline{x})}$, where x_t is the sample data from which we want to compute the standard deviation.

[Abhishek and Hosanagar, 2013] are effective or not.

4.2 Applying GMM to the Model

The central goal of this section is to calculate the GMM estimator as described in Section 3.3 for α_k, γ_k, and the distribution function $F_k(\cdot)$ for an individual keyword k. For ease of exposition, I will omit the subscript k in this section.

The observable variables of the process $(x_t, t \in \mathcal{T})$ are the average position $avgpos_t$, the average cost-per-click $avgcpc_t$, the click-through rate ctr_t, the vector of bids for each keyword b_t, and the number of competitors N_t, such that $x_t = (avgpos_t, avgcpc_t, ctr_t, b_t, N_t)$ has the dimension $p = 5$. To get equivalent expressions involving the true parameter vector and the observed variables for the unobservable disturbances [Abhishek and Hosanagar, 2013] suggest: 'Following the idea of the method of moments, we derive analytical expressions for the moments we observe empirically, namely, the expected position ($avgpos_t$), cost-per-click ($avgcpc_t$), and click-through rate ($ctr_t = cl_t/i_t$), given the bid for each keyword. These moments are as follows:

$$\begin{aligned}\mathbb{E}[pos_t|b_t] &= N_t(1 - F(b_t)),\\ \mathbb{E}[\underline{b}_t|b_t, \delta_t = 1] &= \int_{x<b_t} xd\left(\frac{1 - F(b_t) + \gamma F(x)}{1 - (1-\gamma)F(b_t)}\right)^{N_t},\\ \mathbb{E}[\delta_t|b_t] &= \alpha\gamma^{-N_t}(1 - (1-\gamma)F(b_t))^{N_t}.\end{aligned} \tag{4.1}$$

The observed moments can be expressed in terms of the analytical moments as follows:

$$\begin{aligned}avgpos_t &= \mathbb{E}[pos_t|b_t] + \xi_{1t},\\ avgcpc_t &= \mathbb{E}[\underline{b}_t|b_t, \delta_t = 1] + \xi_{2t},\\ ctr_t &= \mathbb{E}[\delta_t|b_t] + \xi_{3t},\end{aligned} \tag{4.2}$$

where $\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})'$ are the random shocks.' To be able to express the unobservable disturbances as a function of the parameters we need to know the vector of parameters we want to estimate. Beside of the click-through rate at the top position α and the rate γ

at which CTR decays with position, another unknown factor is the distribution function of the bids of the competitors $F(\cdot)$.

'Because the data set contains only daily aggregates, we cannot directly estimate the distribution function $F(\cdot)$ using non-parametric approaches because we have very few bids for each keyword. We therefore use a parametric form for $F(\cdot)$, and estimate its parameters using the first moments associated with the position, cost-per-click, and click-through rate. For the parametric form of the distribution $F(\cdot)$, we choose the Weibull distribution. This choice is based on two factors. Firstly, the Weibull distribution can take on diverse shapes and offers a great deal of flexibility. Secondly, an analysis of a secondary data set of bids submitted to a search engine for several keywords in the insurance sector [Abhishek et al., 2012] shows that the Weibull distribution is reasonably good for modeling the bids. Note that we are not assuming that the distribution of bids for keywords is the same across the two data sets, rather the bids are from the same family (Weibull) and the parameters can vary across keywords. The Weibull distribution has the following cumulative distribution function

$$F(x; \theta, \lambda) = 1 - \exp\left\{-\left(\frac{x}{\lambda}\right)^{\theta}\right\}. \tag{4.3}$$

It is defined by two parameters θ and λ.'[10]

Therefore, additionally to the two parameters α and γ, we also want to estimate the parameters θ and λ of the distribution function $F(\cdot)$. The parameter vector β is in the parameter space $S \subseteq \mathbb{R}^4$, which can be defined as $S = \{(\alpha, \gamma, \lambda, \theta) | \alpha, \lambda, \theta \geq 0 \wedge \gamma \in [0, 1]\}$. Thus the unobservable disturbances can be denoted as the function of the parameters

$$\xi_t(\beta_0) = \begin{pmatrix} avgpos_t - \mathbb{E}[pos_t | b_t] \\ avgcpc_t - \mathbb{E}[\underline{b}_t | b_t, \delta_t = 1] \\ ctr_t - \mathbb{E}[\delta_t | b_t] \end{pmatrix}, \tag{4.4}$$

with $\xi_t(\beta) : \mathbb{R}^5 \times S \to \mathbb{R}^3$.

[10] [Abhishek and Hosanagar, 2013], page 863.

[Hansen, 1982] supposes: 'A common way to obtain orthogonality conditions is to exploit the assumption that disturbances in an econometric model are orthogonal to functions of a set of variables that the econometrician observes.' In our model we can exploit the assumption that the random shocks are orthogonal to the bids $\mathbb{E}[b\xi] = 0$ to find a function z_t such that $\mathbb{E}[\xi_t(\beta_0) \otimes z_t] = 0$, which is stated by the moment conditions. Let $\tilde{b}$ be a vector, which contains each bid of $\{b_t | t \in \mathcal{T}\}$ only once such that the dimension $\tilde{T}$ of $\tilde{b}$ equals the cardinality of $\{b_t | t \in \mathcal{T}\}$. I choose the function $z_t : \mathbb{R}^5 \to \mathbb{R}^{\tilde{T}}$ to be of the form

$$z_t = (b_t \cdot \delta(b_t, \tilde{b}_i))_{i \in \{1,\dots,\tilde{T}\}},$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta[11], because the expectation $\mathbb{E}[b_i \xi_t]$ obviously only equals zero if b_i was the bid set at the day t. Otherwise we just have the trivial orthogonality condition $\mathbb{E}[0 \cdot \xi_t] = 0$.

Therefore we can define the function $f : \mathbb{R}^5 \times S \to \mathbb{R}^{3 \cdot \tilde{T}}$ as

$$f(x_t, \beta) = \xi_t(\beta) \otimes z_t, \qquad \text{with } \mathbb{E} f(x_t, \beta_0) = 0,$$

having three moment conditions for each unique bid. To satisfy the necessary condition $r \geq q$ – where $r = 3\tilde{T}$ is the dimension of $f(x_t, \beta)$ and $q = 4$ is the dimension of the vector of parameters – we need the minimum of two unique bids per keyword ($\tilde{T} = 2$) to have six moment conditions, which is enough to estimate the parameter vector $\beta_0 = (\alpha_0, \gamma_0, \lambda_0, \theta_0)$. As stated in Section 4.1, the ticket agency changes their bids weekly during the 4-week estimation period. Thus we have the required minimum of two unique bids per keyword and are ready to estimate our parameters with the generalized method of moments.

4.3 Computing the Optimal Bids

I implemented a daily automated download of a summary of the aggregated data from the keywords of the experimental group with the data of the preceding day on Google

[11] The Kronecker delta is defined as $\delta(t,i) = \begin{cases} 1 & \text{if } i = t \\ 0 & \text{otherwise} \end{cases}$.

Adwords and an extract of such a summary is shown in Figure 4.2. As one can see there are a lot more informations than I actually need, and for each day there is a different file. I used a pivot table in Excel to get the data in a way I could easily use in MATLAB, which can be seen in Table 4.3.[12] Using the data from the 'before' period I compute the sample means μ_k and Ew_k, the daily budget D_k based on the mean daily spent, and I got the data for the number of competitors N_k for each keyword, which I also included in my pivot table. Having the data this way, they can be loaded easily into

	Avg. position	Avg. CPC	CTR	bid	mu	Ew	D	N
musical wien					50	1.5	0.80	14
18.05.14	2.8	0.50	0.02	0.55				
19.05.14	2.5	0.55	0.03	0.60				
20.05.14	2.6	0.53	0.03	0.60				
⋮	⋮	⋮	⋮	⋮				
spanische hofreitschule					100	3.0	1.50	20
18.05.14	3.8	1.50	0.12	1.55				
19.05.14	3.5	1.55	0.13	1.60				
20.05.14	3.6	1.53	0.13	1.60				
⋮	⋮	⋮	⋮	⋮				

Table 4.3: Table with daily summary measures

MATLAB. To compute the optimal bid I wrote a function named `computeBid`, which needs as an input the name of the excel file `filename`, the number of estimation days the data contains `numDays`, the number of keywords `numKws`, and the row `row1` where the table with the data begins. The output of `computeBid` are the optimal bids `Bid` and the estimated parameters `Beta`. I upload the file with the function `xlsread`, define the vectors μ, Ew, D, and N and set the matrix `Data` including the columns 'Avg. position', 'Avg. CPC', 'CTR', and 'bid' of the pivot table.

[12]The data in the table are only for the purpose of illustration and not the real data of the ticket agency.

	A	B	C	D	E	F	G	H	I	J	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB
1	myopic Bericht (Jul 27, 2014)																										
2	Parameters:InMicros=FALSE;EntityType=ADGROUPCRITERION;																										
3	Action	Keyword state	Keyword	Match type	Campaign	Ad group	Ad group ID	Status	Keyword max CPC	Ad group max CPC	Avg. position	Avg. CPC	CTR	Quality score	First page CPC	Top of page CPC	Clicks	Impressions	Cost	Conversions	Cost / conv.	Conv. rate	Total conv. value	Conv. value / cost	Conv. value / click	Search Impr. share	Keyword ID
4		enabled	oper wien	Broad	Deutsch	Wien C	19649	below	0.04	0.04	3.1	0.04	0.00	10	0.01	0.03	100	8	4.00	0	0.00	0.00%	0.0	0.0	0.0	17.02%	1058907
5		enabled	wiener op	Broad	Deutsch	Wien C	19649	below	0.04	0.04	3.0	0.04	0.00	10	0.01	0.03	100	32	4.00	0	0.00	0.00%	0.0	0.0	0.0	21.33%	1092853
6		enabled	Karten O	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	3.5	0.04	0.00	10	0.01	0.03	100	2	4.00	0	0.00	0.00%	0.0	0.0	0.0	100.00	1234833
7		enabled	opernhau	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	3.4	0.04	6.45	10	0.01	0.03	100	31	4.00	0	0.00	0.00%	0.0	0.0	0.0	26.96%	3315529
8		enabled	oper in w	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	3.2	0.04	9.52	10	0.01	0.03	100	21	4.00	0	0.00	0.00%	0.0	0.0	0.0	46.67%	3315553
9		enabled	oper wien	Exact	Deutsch	Wien C	19649	below	0.04	0.04	3.3	0.04	4.26	10	0.01	0.03	100	47	4.00	0	0.00	0.00%	0.0	0.0	0.0	40.52%	3924369
10		enabled	oper kart	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	4.0	0.04	0.00	10	0.01	0.03	100	3	4.00	0	0.00	0.00%	0.0	0.0	0.0	100.00	4152290
11		enabled	wien oper	Broad	Deutsch	Wien C	19649	below	0.04	0.04	3.7	0.04	0.00	10	0.01	0.03	100	17	4.00	0	0.00	0.00%	0.0	0.0	0.0	56.67%	4799957
12		enabled	wien oper	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	2.4	0.04	3.05	10	0.01	0.03	100	262	4.00	0	0.00	0.00%	0.0	0.0	0.0	29.21%	4995459
13		enabled	wien oper	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	4.2	0.04	20.0	10	0.01	0.03	100	5	4.00	0	0.00	0.00%	0.0	0.0	0.0	100.00	5362642
14		enabled	wien oper	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	3.3	0.04	0.00	10	0.01	0.03	100	35	4.00	0	0.00	0.00%	0.0	0.0	0.0	50.72%	5362648
15		enabled	opernpro	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	2.3	0.04	14.2	10	0.01	0.03	100	7	4.00	0	0.00	0.00%	0.0	0.0	0.0	87.50%	2383906
16		enabled	oper wien	Broad	Deutsch	Wien C	19649	eligible	0.04	0.04	2.3	0.04	0.00	10	0.01	0.03	100	23	4.00	0	0.00	0.00%	0.0	0.0	0.0	46.00%	5467142
17		enabled	musicalka	Broad	Deutsch	Wien n	19699	eligible	0.04	0.04	3.3	0.04	33.3	10	0.01	0.03	100	3	4.00	0	0.00	0.00%	0.0	0.0	0.0	60.00%	2547755
18		enabled	musicals	Broad	Deutsch	Wien n	19699	eligible	0.04	0.04	2.5	0.04	0.00	10	0.01	0.03	100	21	4.00	0	0.00	0.00%	0.0	0.0	0.0	87.50%	4996416
19		enabled	musical w	Broad	Deutsch	Wien n	19699	eligible	0.04	0.04	1.8	0.04	15.1	10	0.01	0.03	100	844	4.00	0	0.00	0.00%	0.0	0.0	0.0	68.40%	4996417
20		enabled	wien mus	Broad	Deutsch	Wien n	19699	eligible	0.04	0.04	1.9	0.04	15.3	10	0.01	0.03	100	39	4.00	0	0.00	0.00%	0.0	0.0	0.0	73.58%	4996421
21		enabled	wien mus	Broad	Deutsch	Wien n	19699	eligible	0.04	0.04	2.0	0.04	18.0	10	0.01	0.03	100	122	4.00	0	0.00	0.00%	0.0	0.0	0.0	92.42%	4996421

Figure 4.2: Part of daily summary of aggregated data

To estimate the parameter vector β I use the function `estimate_beta`, which needs the starting value `Beta0` and the maximum number of iterations of the GMM estimation `maxConvNumb`. Having the estimated parameters, I finally use the function `optimal_bid` to compute the bid using the bidding policy of Chapter 3. This is my main function and of course I will also explain my functions `estimate_beta` and `optimal_bid`.[13]

Listing 4.1: computeBid.m

```
function [ Bid, Beta ] = computeBid( ...
    filename,numDays,numKws,row1 )

Range = ...
    ['B',num2str(row1),':I',num2str((numDays+1)*numKws+row1-1)];
Keywords=xlsread(filename, 1, Range);

Beta=zeros(numKws,4);
mu=zeros(1,numKws);
Ew=zeros(1,numKws);
D=0;
N=zeros(1,numKws);

Beta0 = [0.03,1.3,0.07,1.2];
maxConvNumb=15;

for k=1:numKws
    mu(k)=Keywords((k-1)*(numDays+1)+1,5);
    Ew(k)=Keywords((k-1)*(numDays+1)+1,6);
    D=D + Keywords((k-1)*(numDays+1)+1,7);
    N(k)=Keywords((k-1)*(numDays+1)+1,8);
    Data_k=Keywords((k-1)*(numDays+1)+2:k*(numDays+1),1:4);
    [Beta(k,:)]=estimate_beta(Data_k,N(k),Beta0,maxConvNumb);
end

Bid=optimal_bid( Beta,Ew,mu,N,D );
```

[13] I used the MATLAB code gmmestimation.m by Zhiguang Cao, which can be downloaded from http://www.mathworks.com/matlabcentral/fileexchange/12114-gmm/content/gmmestimation.m (Visited on August 7, 2014). To write my functions and to present the MATLAB codes I used the M-Code LaTeX Package by Florian Knorn on http://www.mathworks.com/matlabcentral/fileexchange/8015-m-code-latex-package (Visited on August 1, 2014).

```
end
```

First of all, to estimate the parameter vector β_0 for each keyword we have to compute the unobservable disturbances as a function of the parameters $\xi_t(\beta_0)$, which is shown in Equation (4.4). For the purpose of using this equation in MATLAB we have to write the integral $\mathbb{E}[\underline{b}_t|b_t, \delta_t = 1] = \int_{x<b_t} xd\left(\frac{1-F(b_t)+\gamma F(x)}{1-(1-\gamma)F(b_t)}\right)^{N_t}$ differently, namely as a integral over x, which I achieve by using integration by parts:

$$\int_{x<b_t} xd\left(\frac{1-F(b_t)+\gamma F(x)}{1-(1-\gamma)F(b_t)}\right)^{N_t} = b_t - \int_{x<b_t}\left(\frac{1-F(b_t)+\gamma F(x)}{1-(1-\gamma)F(b_t)}\right)^{N_t} dx.$$

I will use $\overline{N}$ with the data of the 'before' period instead of the daily data N_t because it is not possible to obtain the exact information of everyday's competitors for each keyword on Google Adwords. In `estimate_beta` I will use the function called `moment` with the input of `Data`, `N`, the function `z` which is orthogonal to ξ_t, `Beta`, the weighting matrix `W`, and `type` defining whether we want $f(x_t, \beta)$ or the objective function $\mu_T(\beta)'W\mu_T(\beta)$ as the output `f_or_obj`.

Listing 4.2: moment.m

```
function f_or_obj = moment ( Data,N,z,Beta,W,type )

% check if the input parameters are all >= 0 and if gamma
% is >1 and not to high.
if ((sum(Beta <= 1e-2) > 0)||Beta(2)<=1.0001||Beta(2)>=5)
    value = 1000; %is multiplicated with the objective ... function if necessary
else
    value=1;
end

T = size(Data,1);
alpha=Beta(1);
gamma=Beta(2);

b=Data(:,4);

%computation of xi_t
xi=zeros(T,3);
```

```
xi(:,1)=Data(:,1) - N.*(1-F(b,Beta));

for t=1:T
    Ecpc =...
        b(t) - integral(@(x) (1-F(b(t),Beta)+gamma*F(x,Beta))/...
          (1-(1-gamma)*F(b(t),Beta)).^N,0,b(t));
    xi(t,2) = Data(t,2) - Ecpc;
end

xi(:,3)=Data(:,3)-alpha*gamma.^(-N).*(1-(1-gamma).*F(b,Beta)).^N;

%computation of the kronecker product f = xi x z
f=zeros(T,size(z,2)*3);
for t=1:T
    f(t,:)=kron(xi(t,:),z(t,:));
end
mu=mean(f,1)';
obj=mu'*W*mu;

if type=='obj'
    f_or_obj=value*obj;%to avoid getting unwanted parameters
elseif type=='f'
    f_or_obj=f;
end

end
```

The distribution function of Equation (4.3) is given by `F.m`.

Listing 4.3: F.m

```
function Fx = F ( x,Beta )
lambda = Beta(3);
theta = Beta(4);

Fx = 1-exp(-(x./lambda).^theta);

end
```

The computation of the weighting matrix as defined in Equation (3.18) is carried out in `WeightingMatrix.m`.

Listing 4.4: WeightingMatrix.m

```
function W = WeightingMatrix( f )
T=size(f,1); %number of days

b_T = floor(T^(2/9)); %'function of T that grows slowly enough ...
    with T' (Newey and West)

V=f'*f*1/T;
w=zeros(b_T);
for i=1:b_T
    Gamma = f((i+1):T,:)'*f(1:(T-i),:)*1/T;
    w=1-i/(b_T+1);
    V=V + w*(Gamma + Gamma');
end

W=pinv(V);
end
```

Having specified these functions, it is finally possible to present you my MATLAB code `estimate_beta.m` for the estimation of β_0.

Listing 4.5: estimate_beta.m

```
function [ Beta_dach ] = estimate_beta( ...
    Data,N,Beta0,maxConvNumb )

Data(:,1)=Data(:,1)-1; %the top position is defined by 0 in ...
    our model

%excluding some irrelevant data only caused by too many clicks
Data=Data(find(Data(:,1)>=0),:);
Data=Data(find(Data(:,2)>0),:);
Data=Data(find(Data(:,3)>0),:);
Data=Data(find(Data(:,3)<1),:);
T = size(Data,1); %number of days

% define the function z, which is orthogonal to f
b=Data(:,4);%bids
b_unique = unique(b);
num_bids = length(b_unique);%number of unique bids
z=zeros(T,num_bids);
for i=1:T
```

```
        for j=1:num_bids
            if b(i) == b_unique(j)
                z(i,j) = b(i);
            end
        end
    end

%iterated GMM
num_mc=3*num_bids;%number of moment conditions
W(:,:,1) = eye(num_mc);
moment_obj = @(Beta) moment(Data,N,z,Beta,W(:,:,1),'obj');
[Beta(:,1),fminvalue(1)] = fminsearch(moment_obj,Beta0,...
            optimset('MaxIter',5000,'MaxFunEvals',5000));

for i=2:maxConvNumb
    f = moment(Data,N,z,Beta(:,i-1),W(:,:,i-1),'f');
    %compute weighting matrix and new estimators of Beta
    W(:,:,i) = WeightingMatrix(f);
    moment_obj = @(Beta) moment(Data,N,z,Beta,W(:,:,i),'obj');
    [Beta(:,i),fminvalue(i)] = fminsearch(moment_obj,...
            Beta(:,i-1),optimset('MaxIter',5000,'MaxFunEvals',5000));

    %termination conditions (also periodical repetitions)

    if abs(fminvalue(i)-fminvalue(i-1))<1e-6||...
            fminvalue(i)<=1e-10 || ...
            (i>2 && abs(fminvalue(i)-fminvalue(i-2))<1e-6 && ...
               fminvalue(i)<fminvalue(i-1))||...
            (i>3 && abs(fminvalue(i)-fminvalue(i-3))<1e-6 && ...
               fminvalue(i)<fminvalue(i-1) && ...
               fminvalue(i)<fminvalue(i-2))||...
            (i>4 && abs(fminvalue(i)-fminvalue(i-4))<1e-6 && ...
               fminvalue(i)<fminvalue(i-1) && ...
               fminvalue(i)<fminvalue(i-2) && ...
               fminvalue(i)<fminvalue(i-3))||...
            (i>5 && abs(fminvalue(i)-fminvalue(i-5))<1e-6 && ...
               fminvalue(i)<fminvalue(i-1) && ...
               fminvalue(i)<fminvalue(i-2) && ...
               fminvalue(i)<fminvalue(i-3) && ...
               fminvalue(i)<fminvalue(i-4))
        break
    end
end

if i == maxConvNumb
```

```
        minvalue=min(fminvalue);
        i_minvalue=find(fminvalue==minvalue);
        Beta_dach = Beta(:,i_minvalue);
        display('The number of iterations equals maxConvNumb');
    else
        Beta_dach = Beta(:,i);
    end

    end
```

To be able to compute the optimal bids we still need the function `optimal_bid`. As stated in Section 3.2, we get the optimal bids by computing the bids with Equation (3.13) for a chosen value of *const* and afterwards checking with Equation (3.16) if we are already sufficiently close to the daily budget D. My MATLAB function `bid` computes the bids $\forall k : b_k$ for a given `const` and a starting value `b0` using Equation (3.13).

Listing 4.6: bid.m

```
function b = bid ( const,b0,Beta,N,Ew )
K = length(N); %number of keywords
b = zeros(K,1);
for k=1:K
        frac1=integral(@(x) ...
            (1-F(b,Beta(k,:))+Beta(k,2)*F(x,Beta(k,:))).^(N(k)-1),0,b);
        frac2=((Beta(k,2)-1)*(1+(Beta(k,2)-1)*...
                F(b,Beta(k,:))).^(N(k)-1));
        b(k) = fsolve(@(b) const-1/Ew(k)*(b+frac1/frac2),b0(k),...
            optimset('MaxFunEvals',5000,'MaxIter',5000));
end
end
```

Let D_b be the left-hand side of Equation (3.16). My MATLAB function `budget_diff` then computes the difference $D - D_b$.

Listing 4.7: budget_diff

```
function diff = budget_diff( b,Beta,mu,N,D ) %budget_difference
K = length(N); %number of keywords
```

```
diff = D;
for k=1:K
    int=integral(@(x) (1-F(b(k),Beta(k,:))+Beta(k,2)*...
           F(x,Beta(k,:))).^N(k),0,b(k));
    diff=diff-mu(k)*Beta(k,1)*Beta(k,2)^(-N(k)) *...
          (b(k) *(1+(Beta(k,2)-1)*F(b(k),Beta(k,:)))^N(k) - int);
end
end
```

In my function `optimal_bid` I use a bisection method to alter `const`. I bisect intervals in the form of [`const_l`, `const_r`] such that the budget difference of `const_l` is always greater than zero and the one of `const_r` is always less than zero, until one of the budget differences equals zero or a termination condition is met.[14]

Listing 4.8: optimal_bid.m

```
function [ b ] = optimal_bid( Beta,Ew,mu,N,D )
%bisection method to find suitable constant
%for the optimality condition

const_l = 0; %at const_l the expected cost equals 0 < D <=> ...
    diff_l=D>0
const_r = 5;

K = length(N); %number of keywords
b_l = zeros(K,1);
diff_l = D;
b0=ones(K,1);
b_r=bid(const_r,b0,Beta,N,Ew);

diff_r = budget_diff(b_r,Beta,mu,N,D);

while (diff_r >=0) %find const_r with diff_r <0
    const_r = 2*const_r;
    b_r = bid(const_r,b0,Beta,N,Ew);
    diff_r = budget_diff(b_r,Beta,mu,N,D);
end
%now at const_r the expected cost is higher than D <=> diff_r <0

```

[14] By definition a budget difference, which is greater than zero, means that the expected cost of the corresponding *const* is less than the daily budget D.

```
while (const_r-const_l>1e-5 && diff_l>1e-5 &&...
        abs(diff_r-diff_l)/abs(diff_l)>1e-4)

    const_m = (const_l+const_r)/2;
    b_m = bid(const_m,b_l,Beta,N,Ew);
    diff_m = budget_diff(b_m,Beta,mu,N,D);
%     display(b_m);
    display(const_r-const_l);
%     display(diff_l);
%     display(diff_r);

    if (diff_l*diff_m<0)
        diff_r=diff_m;
        b_r=b_m;
        const_r=const_m;
    elseif (diff_l*diff_m>0)
        diff_l=diff_m;
        b_l=b_m;
        const_l=const_m;
    else
        b=b_m;
        break;
    end

end

b=b_l; %expected cost <= D
end
```

Now that we have specified all functions, we are able to run `computeBid` of Listing 4.1. The parameters $\alpha, \gamma, \lambda, \theta$, and the optimal bids b_k are estimated for all keywords. During the estimation period I recomputed the bids each week[15] using the newly available data to reestimate the parameters. I figured out that the majority of the bids do not change a lot. In Figure 4.3 histograms of the estimated parameters are plotted for all keywords of the experimental group. In Table 4.4 the real value of the estimated parameters, the average number of competitors N_k, the height of the expected value of a click Ew_k, and the height of the computed bid b_k are shown exemplarily for seven keywords of the experimental group.

[15] Because of the short duration of that period I only changed the bid once.

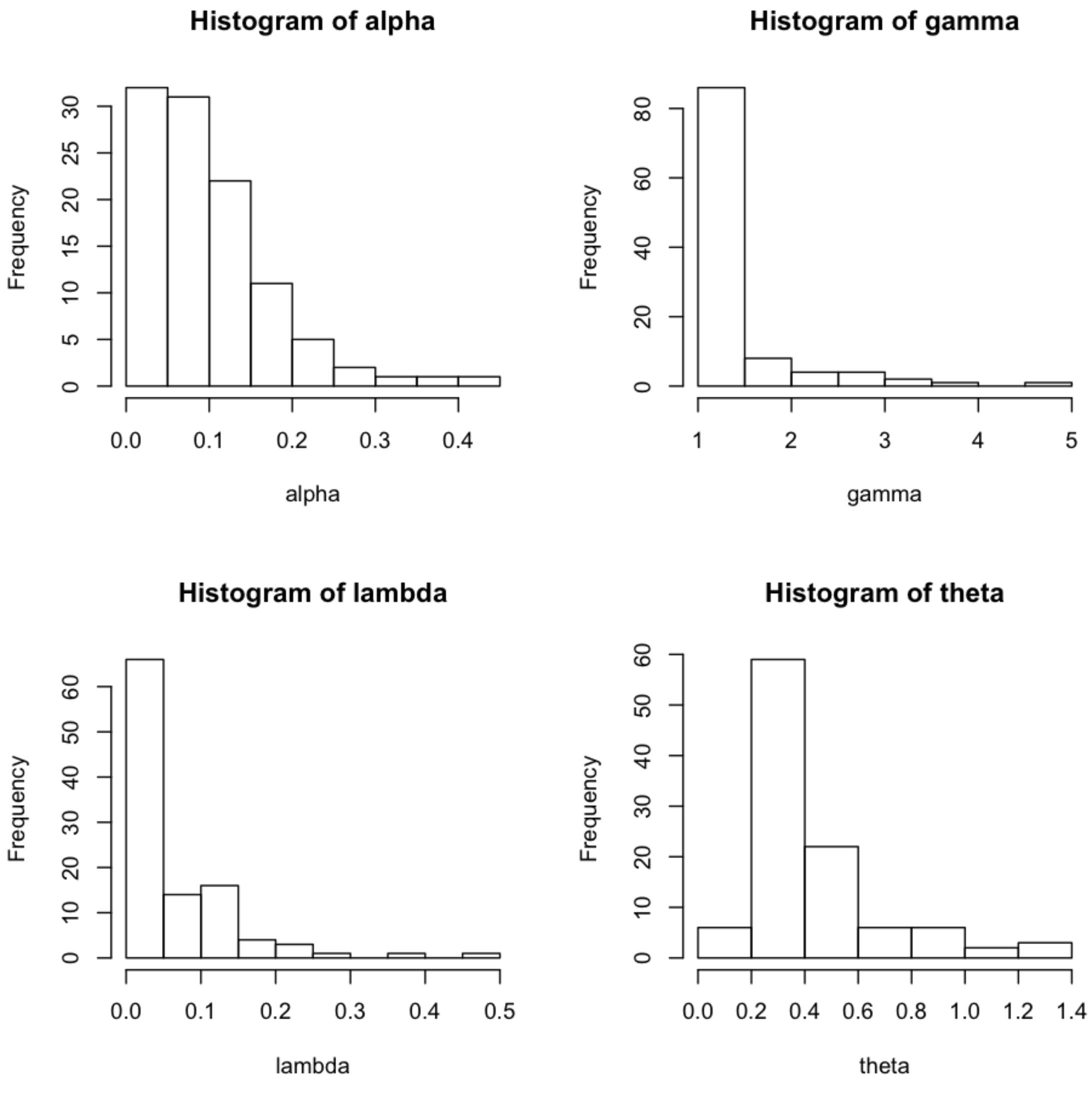

Figure 4.3: Histograms of estimated parameters for the experimental group

Due to the ticket office's company secret, I use expressions such as 'low' and 'high' instead of the real values for Ew_k and b_k comparing the values with their mean value.

Keyword	α	γ	λ	θ	N	Ew	b
vienna golden hall	0.0379	1.0593	0.1234	0.5121	13	very high	very high
staatsoper budapest	0.1319	1.1832	0.1251	0.8698	7	low	low
musical wien	0.1182	4.9994	0.0100	0.2921	14	medium	very high
oper in prag	0.0681	1.5668	0.0100	0.2446	20	low	low
spanish riding school	0.0100	1.0730	0.1331	0.6295	11	high	high
vienna boys' choir	0.0324	1.1973	0.0271	0.2801	10	high	high
palau de la ... [16]	0.0634	1.3685	0.0563	0.5045	4	low	low

Table 4.4: Estimated parameters for keywords of the experimental group

The bids of the keywords 'vienna golden hall', 'musical wien', 'spanish riding school', and 'vienna boys' choir' range between high and very high, and we can see that the expected values per click of these keywords range between medium and very high. For all keywords with a low expected value per click such as 'staatsoper budapest', 'oper in prag', or 'palau de la musica catalana ticket', the computed bids are also represented by a low value. We observe that the height of the expected value per click is a very important value for the decision of the bid, but not the only one as we can see at the keyword 'musical wien', which has only a medium expected value per click, yet its bids are given by a very high value. This makes sense because the estimated parameter γ is very high, thus the number of clicks decreases dramatically at a lower position, and because of the reasonably good expected value per click it is worth to pay a high cost-per-click value to attain the highest possible position.

[16] The entire name of the keyword is 'palau de la musica catalana ticket'.

4.4 Implementation into Google Adwords

After computing the optimal bids as described in the previous Section 4.3 they can be uploaded into Google Adwords after copying the bids into an appropriate excel sheet. I used the Model of Chapter 3 to compute the bids of the ticket agency for two weeks. Now I want to compare if it is better for the company to decide on the bids themselves with the support of google or to compute the bids using the bidding policy suggested by [Abhishek and Hosanagar, 2013]. To compare this I will use a Difference-in-Differences (DiD) approach, where 'the average gain in the second (control) group is substracted from the average gain in the first (treatment) group. This removes biases in second period comparisons between the treatment and control group that could be the result from permanent differences between those groups, as well as biases from comparisons over time in the treatment group that could be the result of trends.'[17]

I use the Return On Investment (ROI), which is computed by dividing the total revenue by the total costs to get the average gain for the two groups. The change in the performance due to the new bidding policy is then given by

$$\begin{aligned}\delta &= \Delta\mathrm{ROI}_{\mathrm{Experimental}} - \Delta\mathrm{ROI}_{\mathrm{Control}} \\ &= -7.4 - (-2.4) \\ &= -5.\end{aligned} \tag{4.5}$$

Contrary to expectations, the return on investment of the ticket agency's advertising campaign decreased by 500% on a DiD basis, indicating that the advertiser's bidding policy outperforms the bidding policy as described in Chapter 3 dramatically. In Section 4.5 I will try to analyze what went wrong and why there was such a performance loss.

Additionally, we see that the ROI of the control group also deteriorated ($\Delta\mathrm{ROI}_{\mathrm{Control}} = -2.4$) in the 'after' period compared to the 'before' period. This is mainly because of the inherent seasonality of ticket sales. The 'after' period runs from July 7 until August 4, which is a typical period for holidays. Although people go to concerts and other events

[17] http://www.nber.org/WNE/lect_10_diffindiffs.pdf (Visited on August 12, 2014)

during their holidays, they often either book their tickets in advance or spontaneously at the event site or at the hotel, but there are also clearly less people booking tickets on the internet during their holidays than at home. What they do, however, is to collect information about where they could go or which concert they would like to see, not only offline but also online. They click on a lot of advertisements to get a good overview of what they could do, but the majority do not actually buy the tickets online, thus the click-through rate increases while the value per click decreases, and both the average position and the cost-per-click do not change. As a result, the total costs increase while the total revenue decreases, which obviously leads to a lower return on investment of the control group.

4.5 Analysis of the Field Experiment

In Section 4.4 we saw that the advertiser's policy performs much better than the bidding policy of Chapter 3. Now there are a few questions to be answered: What went wrong? Which factors lead to the deterioration in the performance of the ticket agency's Google Adwords campaign?

Obviously, the model of Chapter 3 is only a simplification of the reality and I think that some very important aspects of the mechanics of Google Adwords are ignored in that model.

We assumed that the probability of a click conditional on the position *pos* is given by Equation (3.4), which means that the decrease of clicks changes with factor γ from one position to the lower position. As shown in Figure 4.4, there are advertisements at the top of the page, which the customer sees before the organic results, and others on the right side of them. In my opinion this may be an appropriate approach for the advertisements on the right side but definitely not for the ones on the top of the page. But there is additionally the difficulty that there are zero to three advertisements on the top of the page and only Google knows on which factor this depends. It also varies from query to query on the same device from one second to another as one can see on Figure 4.4. It

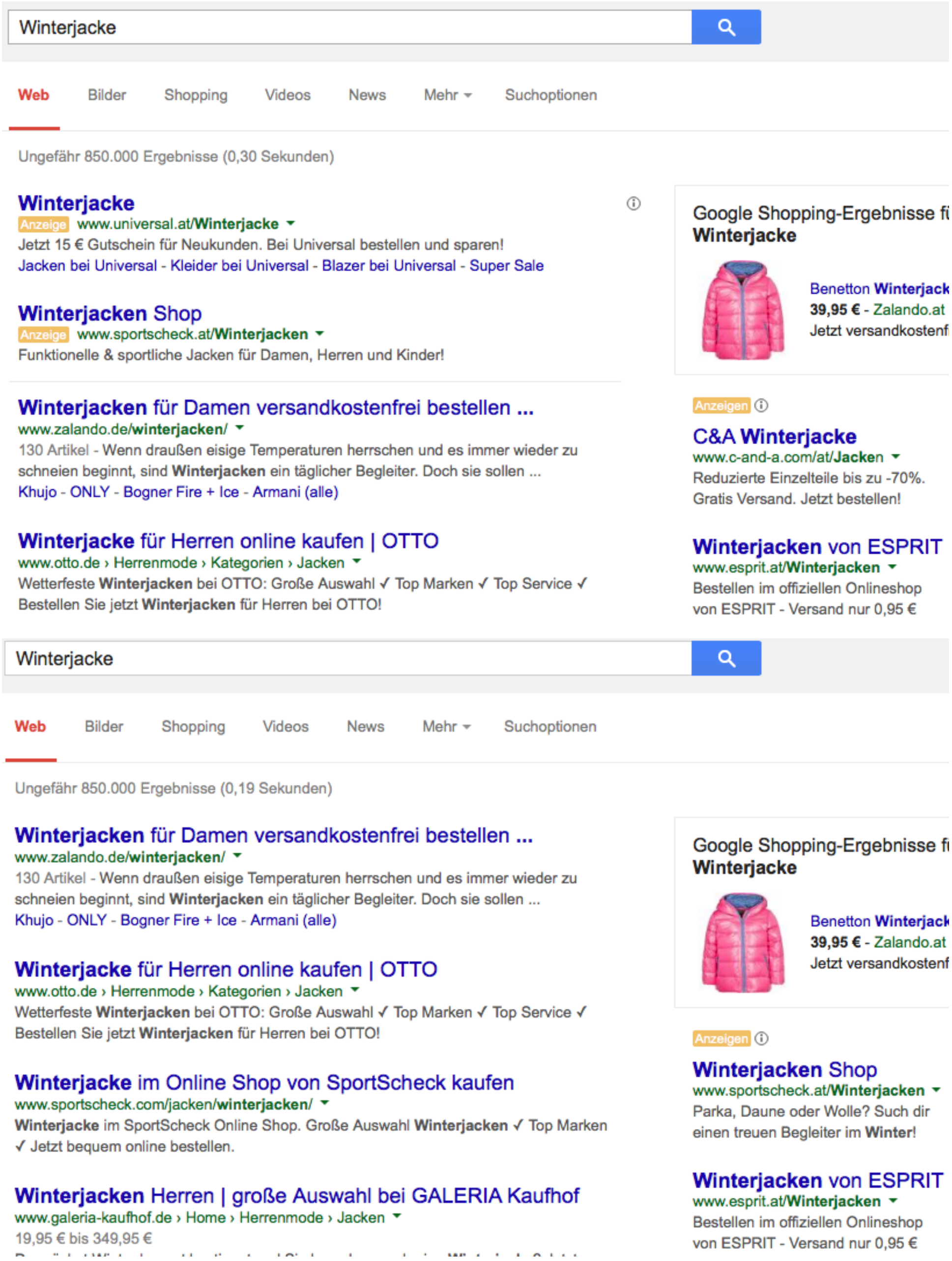

Figure 4.4: Top of Page on Google Adwords; [Google, 2014d].

is basically impossible to model this because of the lack of information, but it would be very important, because the difference between the click-through rate at the top of page and at other positions is enormous. For the keyword 'spanische hofreitschule' the click-through rate at the top of page is 13.84%, while at other positions it is only 1.24%. The keyword 'burgtheater wien' shows a similar result with a click-through rate of 11.55% compared to 0.56%. By means of these two keywords also the variance of the amount of advertisements at the top of page can be shown. The average position for 'spanische hofreitschule' is 1.5 at the top of page and 2.6 at other positions, while for 'burgtheater wien' the average position at the top of page is 1.0 and the one at other positions is 1.4, which shows that for the second keyword Google does not seem to show more than one advertisement at the top of page.

As mentioned in Section 3.1, we simplified our model by ignoring the fact that the quality score[18] plays an important role in the ad rank[19] of each advertiser. We assumed that the only determining factor for the ad rank is the bid of each advertiser, although we know that Google Adwords weights the bids with the quality score so that it is possible that the ad of someone who has a higher bid than all other advertiser is not at the first position. The explanation why the quality score is ignored is the following in [Abhishek and Hosanagar, 2013]:
'The discussion assumes that ads are ordered by bid and that the advertiser pays the bid of the next advertiser. A common practice is to use the product of bid and a quality score to rank order the advertisers, and the payment is the minimum bid needed to secure the position (e.g., the payment per click for an advertiser in position i is $bid(i+1) * \text{Quality}(i+1)/\text{Quality}(i)$. This does not affect our model. If we normalize the bid of all competitors by the ratio of their quality score relative to our advertiser ($NormalizedBid = bid * \text{Quality}_{\text{Competitor}}/\text{Quality}_{\text{Advertiser}}$), our analysis can be interpreted as based on this normalized bid.'

[18] The *quality score* is 'an estimate of the quality of your ads, keywords, and landing page. Higher quality ads can lead to lower prices and better ad positions.' https://support.google.com/adwords/answer/140351?hl=en (Visited on August 19, 2014)

[19] The *ad rank* is 'a value that's used to determine your ad position (where ads are shown on a page) and whether your ads will show at all. Ad Rank is calculated using your bid amount, the components of Quality Score (expected clickthrough rate, ad relevance, and landing page experience), and the expected impact of extensions and other ad formats.' https://support.google.com/adwords/answer/1752122?hl=en (Visited on August 19, 2014)

In my opinion this simplification would only be suitable if the quality of the advertisers did not change, because as soon as the quality of the ticket agency increases for a keyword k, the normalized bids of the competitors decrease automatically so that the ticket agency gets a relatively better ad rank without changing its bid. Also a change in the quality of another competitor would have effects on the sequence of the ads. Although I agree with [Abhishek and Hosanagar, 2013] that in general bids for keywords are rarely updated continuously, our normalized bids are changed continuously because of the adjustments in the quality score. They tested the effect of changes in bids on their bidding policy by computing the mean absolute error (MAE) of the predicted and observed average position and cost-per-click in the after period, where they use the Equations (4.1) to obtain the predicted moments. 'If there is competitive reaction then the predicted average position and CPC would be considerably different from the observed position or CPC as the competitors might change their bids as a response to the changes in bids by the advertiser. If the predicted and observed moments of these quantities are not very different, it suggests that the competitive reaction is subdued.'[20] The MAE of the ticket agency's data is reported in Table 4.5.

Quantity	MAE
position	1.862
CPC	$0.05

Table 4.5: Mean absolute errors of the moments

We can see that the MAE between the predicted and observed position is quite high compared to the value 0.141 in [Abhishek and Hosanagar, 2013], where the one of the CPC is very close to the predicted value. As indicated above, I do not think that this is a result of high competitive reaction during the experimental phase. I rather think that this confirms the impact of the quality changes on the ad rank. The quality affects the position of the ad but not so much the CPC because the normalized bid of the ticket agency does not change with quality and therefore the bid of the next advertiser tends

[20] [Abhishek and Hosanagar, 2013], Appendix, page 39.

to be similar to their bid and thus similar to the bid before the change of the quality. In addition, we assumed that the distribution function of the bids of the competitors $F(\cdot)$ does not change during the estimation period, which is not true because of the function's dependency on the quality scores.

Since October 2012 Google Adwords also provides information about the impression share. 'Impression share (IS) is the number of impressions you've received divided by the estimated number of impressions you were eligible to receive. [...] An easy way to understand the value of impression share is to think of the online advertising landscape as a delicious pie. You and your competitors are each trying to grab the biggest slice of that pie. By tracking your impression share metrics, you're keeping tabs on the size of your slice compared to the whole.'[21] As mentioned in Section 3.1, the impressions are defined as the number of searches for keyword k where the advertiser's ad has been shown. The daily summary measures of the average position, the average cost-per-click, and the click-through rate are only based on the number of impressions and not on the number of searches for a certain search term. If a company wants to be shown at a higher percentage, it has to either increase its budget because the ad is not shown due to insufficient budget, or it has to increase its bid for the keyword if the impression share is low due to a bad ad rank. Thus it is important to observe the impression share and include that percentage in the company's bidding policy.

Another fact which is ignored in the model is that there are more and more people using their mobile phones or tablets to search on the internet. One point is that on a mobile phone there are less advertisements shown. For the search term 'Winterjacke' there are shown seven advertisements on a computer and only three on a mobile phone.[22] Because there are less advertisements of competitors per search query and the companies receive only the summary measures based on the impressions, the click-through rate is considerably higher. Another negative point is that there are fewer people who actually complete a purchase on their mobile phones or on their tablets because it is not so

[21] https://support.google.com/adwords/answer/2497703?hl=en (Visited on August 17, 2014)
[22] This is the result of a the search query on August 8, 2014.

pleasant to type in all the information for payment and delivery with a virtual keyboard. Therefore the revenue per click of mobile phones and tablets is about 60% fewer than the one of computers, and there are less advertisements shown on these devices. The problem is that only 61% of the searches in the estimation period are made from computers and our model does not reflect these differences.

5 Conclusions

The ticket agency for which I carried out the analysis has to decide on and has to adapt regularly the bids for keywords associated with more than 25,000 events. To circumvent this quite time-consuming activity and to improve the performance of its search engine advertising I started to collect a lot of data from the ticket agency's Google Adwords. I introduced the analytical model described by [Abhishek and Hosanagar, 2013], formulated the underlying decision problem, derived the optimality conditions, and described the generalized method of moments estimator, which I needed to estimate the parameters required to solve the decision problem. After that, I divided the data set containing 255 of the ticket agency's keywords into a control group and an experimental group and observed three periods: the 'before', the estimation, and the 'after' period. The bids of the control group were computed during all periods the same way as before, while the bids of the experimental group where changed by the ticket agency weekly during the estimation period. I used my MATLAB program to estimate the required parameters with the generalized method of moments estimator and to compute the optimal bids according to the decision problem and the corresponding optimality conditions. During the 'after' period, the ticket agency used the bids, which I computed for them weekly. I used a Difference-in-Differences (DiD) approach to compare the results of the implemented bidding policy with the advertiser's previous policy and came to the result that the return on investment of the ticket agency's search engine advertising decreased by 500% on a DiD basis in the 'after' period compared to the 'before' period. I pointed out that in my opinion the performance of the ticket agency's Google Adwords campaign deteriorated because the model ignores the difference between the click-through rate at the top of page and at other positions, due to the fact that the quality score plays an important role in the ad rank of each advertiser, because of the importance to observe the impression share and to include that percentage in the company's bidding policy, and finally due to the fact that there are more and more people using their mobile phones where less advertisements are shown and fewer people complete a purchase.

Now the question remains: Why did the bidding policy work for the meat distributor treated in [Abhishek and Hosanagar, 2013] but not for the ticket agency? A very important point is that the field experiment with the data set of the meat distributor was made in 2011 and the one with the data of the ticket agency in 2014. Google is continuously changing the algorithms, the way in which the search results are displayed, adding features showing extra business information with the ads (such as phone number, address, ratings, apps, or additional links), adding new bidding strategies, and making a lot of other smaller and bigger changes. Additionally, I do not know the previous bidding policy of the meat distributor, whether the company used only manual bidding, or if they let Google support them with a bidding strategy like enhanced cost-per-click or automatic CPC bidding. I just know that the average number of unique bids per keyword were 1.12 during three months, which is much less than the one of the ticket agency. Maybe it is still possible to improve the Google Adwords performance of a company using the bidding policy of [Abhishek and Hosanagar, 2013], e.g., if the company normally does not change their bids very often and does not use the new features, but – as I demonstrated in my diploma thesis – not for a company that changes the bids regularly and is aware of the multiple changes in Google Adwords.

Probably it is possible to compute a better bidding policy and automize the bid optimization by extending the model and considering the facts I mentioned in Section 4.5 as well as possible. It would be interesting to implement those changes, where I could even use the same data of the estimation period, compute the bids using the enhanced model, compare again the results and check, whether or not it is possible to improve the performance of the ticket agency's search engine advertising using the new bidding policy.

List of Figures

List of Tables

Bibliography

[Abhishek and Hosanagar, 2013] Abhishek, V. and Hosanagar, K. (2013). Optimal bidding in multi-item multislot sponsored search auctions. Operations Research, 61(4):855–873.

[Abhishek et al., 2012] Abhishek, V., Hosanagar, K., and Fader, P. S. (2012). On aggregation bias in sponsored search data: existence and implications. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1490169. (Visited on August 7, 2014).

[Agarwal et al., 2011] Agarwal, A., Hosanagar, K., and Smith, M. D. (2011). Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets. Journal of Marketing Research, 48(6):1057–1073.

[Cao, 2006] Cao, Z. (2006). File exchange - matlab central. http://www.mathworks.com/matlabcentral/fileexchange/12114-gmm/content/gmmestimation.m. (Visited on August 7, 2014).

[Edelman et al., 2007] Edelman, B., Ostrovsky, M., and Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. American Economic Review, 97(1):242–259.

[Ghose and Yang, 2009] Ghose, A. and Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. Management Science, 55(10):1605–1622.

[Google, 2014a] Google (2014a). Ad rank - adwords help. https://support.google.

com/adwords/answer/1752122?hl=en. (Visited on August 19, 2014).

[Google, 2014b] Google (2014b). Enhanced cost-per-click (ecpc) - adwords help. https://support.google.com/adwords/answer/2464964?hl=en. (Visited on August 6, 2014).

[Google, 2014c] Google (2014c). Google search for ”schuhe”. https://www.google.at/?gws_rd=ssl#q=schuhe. (Visited on August 14, 2014).

[Google, 2014d] Google (2014d). Google search for ”winterjacke”. https://www.google.at/?gws_rd=ssl#q=Winterjacke. (Visited on August 14, 2014).

[Google, 2014e] Google (2014e). Quality score - adwords help. https://support.google.com/adwords/answer/140351?hl=en. (Visited on August 19, 2014).

[Google, 2014f] Google (2014f). Tracking impression share - adwords help. https://support.google.com/adwords/answer/2497703?hl=en. (Visited on August 17, 2014).

[Google, 2014g] Google (2014g). Using automatic bidding - adwords help. https://support.google.com/adwords/answer/2470106?hl=en. (Visited on August 6, 2014).

[Hall, 2010] Hall, A. R. (2010). Generalized Method of Moments (GMM). Encyclopedia of Quantitative Finance, pages 836–840. (Cont R., ed.) John Wiley & Sons Ltd.

[Hansen, 1982] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50(4):1029–1054.

[Hansen et al., 1996] Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. Journal of Business & Economic Statistics, 14(3):262–280.

[Imbens and Wooldridge, 2007] Imbens and Wooldridge (2007). Difference-in-differences estimation. http://www.nber.org/WNE/lect_10_diffindiffs.pdf. (Visited on August 12, 2014).

[Katona and Sarvary, 2010] Katona, Z. and Sarvary, M. (2010). The race for sponsored links: Bidding patterns for search advertising. Marketing Science, 29(2):199–215.

[Knorn, 2005] Knorn, F. (2005). M-code latex package - file exchange - matlab central. http://www.mathworks.com/matlabcentral/fileexchange/8015-m-code-latex-package. (Visited on August 1, 2014).

[Newey and West, 1987] Newey, W. K. and West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3):703–708.

[Tragler, 2012] Tragler, G. (2012). Nichtlineare Optimierung. Lecture Notes, TU Wien.

[Varian, 2007] Varian, H. (2007). Position Auctions. International Journal of Industrial Organization, 25(6):1163–1178.

[Varian, 2009] Varian, H. (2009). Online Ad Auctions. American Economic Review, 99(2):430–34.

[Yang and Ghose, 2010] Yang, S. and Ghose, A. (2010). Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? Marketing Science, 29(4):602–623.