Analyzing paid search campaigns using keyword-level data and Bayesian statistics

Von der Fakultät Wirtschaftswissenschaften
der Leuphana Universität Lüneburg

zur Erlangung des Grades
Doktor der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)

genehmigte Dissertation von
Tobias-Benedikt Blask

aus
Lüneburg

Eingereicht am: 18.07.2017
Mündliche Verteidigung (Disputation) am: 13.06.2018

Erstbetreuer und Erstgutachter:	Prof. Dr. Burkhardt Funk
Zweitgutachter:	Prof. Dr. Reinhard Schulte
Drittgutachter:	Prof. Dr.-Ing. habil. Jorge Marx Gómez

Die einzelnen Beiträge des kumulativen Dissertationsvorhabens sind oder werden wie folgt veröffentlicht:

Blask, T., Stange, M., Borck, C. & Funk, B (2017). Predicting conversion rates in paid search using sparse keyword-level data. Working Paper.

Blask, T., Funk, B., & Schulte, R. (2011). To bid or not to bid? Investigating retail-brand keyword performance in sponsored search advertising. In International Conference on E-Business and Telecommunications (pp. 129-140). Springer, Berlin, Heidelberg.

Blask, T. (2013). Bayesian Parameter Estimation in Green Business Process Management: A Case Study in Online-Advertising. In GI-Jahrestagung (pp. 852-863).

Blask, T. (2014). Do specific text features influence click probabilities in Paid Search Advertising?. In e-Business (ICE-B), 2014 11th International Conference on (pp. 55-62). IEEE.

Veröffentlichungsjahr: 2018

Veröffentlicht im Onlineangebot der Universitätsbibliothek unter der URL:
http://www.leuphana.de/ub

Danksagung

Ich danke Prof. Dr. Burkhardt Funk für die unermüdliche, intensive Betreuung und das stetige Zutrauen in die Entwicklung meiner Fähigkeiten. Ich danke ihm für die beständige Motivation an mich, das bisher Gelernte und Errungene zu hinterfragen und dabei immer wieder über mich hinaus zu wachsen.

Dankbar bin ich den Professoren Schulte, Niemeyer und Groß für die Begleitung über die gesamte Zeit, die Unterstützung meiner Entwicklung, das ehrliche Interesse, die regelmäßigen Fachgespräche und den nötigen Blick über den Tellerrand. Prof. Marx Gómez danke ich für das erhaltene wertvolle Feedback.

Darüber hinaus gilt mein Dank auch meinen Freunden und Kollegen Martin, Stefan, Florian, Christian, Max, Thomas, Oliver, Stefanos, Thomas, Basti und Patrick sowie natürlich allen aktiven und ehemaligen Mitgliedern des Instituts für Wirtschaftsinformatik, die mich stets unterstützt haben.

Ohne die Unterstützung meiner Familie wäre diese Arbeit nicht möglich gewesen. Daher danke ich meinen Eltern, meinem Bruder Sebastian und meiner Schwiegerfamilie. Besonderer Dank gebührt aber natürlich meiner Frau Kerstin für Zurückstecken, Geduld, Unterstützung und Zuspruch. Meinen Söhnen Felix und Jakob danke ich von Herzen für die unbeschreibliche Motivation und die ganz besondere, neue Perspektive auf wirklich alle Fragen des Lebens …

Abstract

Analyzing paid search campaigns using keyword-level data and Bayesian statistics

by Tobias-Benedikt Blask

Online marketing, especially Paid Search Advertising, has become one of the most important paid media channels for companies to sell their products and services online. Despite being under intensive examination by a number of researchers for several years, this topic still offers interesting opportunities to contribute to the community, particularly because of its large economic impact and practical relevance as well as the detailed and widely unfiltered view of consumer behavior that such marketing offers.

To provide answers to some of the important questions from advertisers in this context, I present four papers in my thesis, in which I extend previous works on optimization topics such as click and conversion prediction. I apply and extend methods from other fields of research to specific problems in Paid Search. After a short introduction, I start with a paper in which we illustrate a new method that helps advertisers to predict conversion probabilities in Paid Search using sparse keyword-level data. We address one of the central problems in Paid search advertising, which is optimizing own investments in this channel by placing bids in keyword auctions. In many cases, evaluations and decisions are made with extremely sparse data, although anecdotal evidence suggests that online marketing is a typical "Big Data" topic. In the developed algorithm presented in this paper, we use information such as the average time that users spend on the advertiser's website and bounce rates for every given keyword. This previously unused data set is shared between all keywords and used as prior knowledge in our proposed model. A modified version of this algorithm is now the core prediction engine in a productive Paid Search Bid Optimization System that calculates and places millions of bids every day for some of the most recognized retailers and service providers in the German market.

Next, I illustrate the development of a non-reactive experimental method for A/B testing of Paid Search Advertising activities. In that paper, we provide an answer to the heavily discussed question of whether and under what circumstances it makes economic sense for brand owners to pay for Paid Search ads for their own brand keywords in Google AdWords auctions.

Finally, I present two consecutive papers with the same theoretical foundation in which I apply Bayesian methods to evaluate the impact of specific text features in Paid Search Advertisements. The first of the two papers covers a topic that is of interest from two perspectives. On the one hand, I examine the impact of the content of specific text features in paid search advertising. On the other hand, I also investigate a topic that has relevance to sustainability research as well. In practice, companies take their responsibilities for a sustainable planet more and more seriously. In the online-retail businesses, a significant share of all CO_2 emissions is generated by delivering goods to their clients. Now various companies are implementing a greener logistic chain into their business processes. A central question for these performance driven companies in this context is whether it pays to invest in additional costs for carbon neutral delivery and if the customers appreciate these steps and prefer retailers that behave in this manner. In the given paper, I apply a non reactive A/B-test that enables me to evaluate the influence of sustainability information on the customers decision to buy a product by clicking on an advertisement. In a further developed version of the previous paper, I examine the influence of the content of text ads in a multivariate setup for a major European car dealer and conclude by finally showing that differences in the formulation of the textual content can have influence on the click probability of Paid Search ads.

Overall, this theses provides contributions to a number of practically and scientifically relevant topics in the Paid Search research community.

Contents

Abstract **iii**

1 Introduction **1**

1.1 Introduction 1
1.2 The users 6
1.3 The Internet search engines 7
1.4 The advertisers 11
1.4.1 Click-Through Rates on the Search Engine Results Page 11
1.4.2 Conversion Rates at the advertiser′s own domain 13
1.4.3 How can advertisers maximize their revenues? 16
1.5 Data sets in Paid Search 18
1.6 Summary 19
References 22

2 Predicting conversion rates in paid search using sparse keyword-level data **28**

2.1 Introduction 30
2.2 Related Work 32
2.3 Estimation of Conversion Rates 33
2.3.1 Feature Selection 33
2.3.2 Clustering of Keywords 34
2.3.3 Estimation of Conversion Rates 36
2.4 Model evaluation 38
2.4.1 Model Benchmark 39
2.4.2 Applying the Approach to a Case Study 40
2.5 Conclusion 42
2.5.1 Contributions and Limitations 42
2.5.2 Outlook 43
References 44

3 To Bid or Not To Bid? Investigating Retail-Brand Keyword Performance in Sponsored Search Advertising. 46
3.1 Introduction . . . 48
3.2 Literature Review . . . 49
3.2.1 Brand Bidding and Piggybacking . . . 49
3.2.2 Blended Search . . . 50
3.3 Hypothesis . . . 51
3.4 Case Study . . . 53
3.5 Results . . . 54
3.5.1 Testing the hypotheses . . . 54
3.5.2 Economic Impact . . . 60
3.6 Conclusions and Outlook . . . 62
References . . . 63

4 Bayesian Parameter Estimation in Green Business Process Management: A Case Study in Online-Advertising 65
4.1 Introduction . . . 67
4.2 Literature Review . . . 68
4.2.1 Green Marketing . . . 68
4.2.2 Sponsored Search Advertising . . . 69
4.2.3 Click probability . . . 70
4.3 Case Study . . . 72
4.4 Conclusions and Outlook . . . 75
References . . . 78

5 Do Specific Text Features Influence Click Probabilities in Paid Search Advertising? 81
5.1 Introduction . . . 82
5.2 Paid Search Advertising . . . 83
5.3 Case Study . . . 84
5.4 Conclusions and Outlook . . . 92
References . . . 94

List of Figures

1.1 Stakeholders in Paid Search Advertising from an advertiser's perspective . . . 3
1.2 Simple model of a Sponsored Search process from the the advertiser's perspective . . . 4
1.3 Screenshot of the appearance of Google's search engine in 1998 . . . 8
1.4 Google Search Engine Results Page 2016 . . . 9

2.1 Conversion-rate for an individual keyword after a given number of accumulated clicks plotted with +/- 1 standard deviation (using binomial errors) . . . 31
2.2 Feature selection based on cross validation with glmnet. The MSE decreases with the number of selected features. . . . 34
2.3 Clustering of keywords with more than one conversion in the whole observation period using recursive partitioning and the given model. The graph shows the cluster number, the mean CVR and the number of keywords per cluster . . . 35
2.4 Graphical representation of the hierarchical model applied here. . . . 36
2.5 Contour plot of the correlation between a^c and b^c. The expected conversion rates ($a^c/(a^c + b^c)$) correspond to the clusters in Figure 2.3. Since a^c is smaller than b^c by at least one order of magnitude, the slopes of the functions $b^c = f(a^c)$ presented here can be interpreted as the expected conversion rates. . . . 37
2.6 Density of deviations from the final conversion rate (CVR(holdout set) - CVR(predicted)) for the tested methods . . . 40
2.7 Development of conversions the advertiser's account before and after the tool was applied. The dashed vertical line represents the start of the intervention. . . . 41
2.8 Posterior distribution of estimated number of conversions after intervention. . . . 42

3.1 Hypotheses of this study in the search and buying process from a user's perspective . . . 52
3.2 Organic (dashed line) vs. sponsored (solid line) clicks during the test period. . . . 57
3.3 Daily sum of all clicks, generated through the search engine via the brand term in the test period (solid line) compared to the weekday values in the reference period. The boxes contain 50% of the values from the reference period. . . . 58
3.4 Conversion rates observed in the test period (solid red line) with standard errors vs. average conversion rates with standard errors on a given weekday (solid line) in the reference period. . . . 59
3.5 Empirical cumulative distribution of the revenues in the observation period (14-day intervals, containing the reference- as well as the test period), the test period is indicated by the vertical line. . . . 60
3.6 Time series of revenues (14-day intervals) during the reference period (dashed line), including the test period (observation point 27, indicated by the vertical line) and a trend line (solid line). . . . 61

4.1 Two variations of an ad, similar to the ones that were used in the A/B test: Carbon Neutral delivery vs. Fast and Reliable delivery . . . 68
4.2 Empirical comparison of CTR's for Group 2 and Group 1 . . . 73
4.3 Group 1 = CTR for ads advertising Carbon Neutral delivery, Group 2 = CTR for ads advertising Fast and Reliable delivery. . . . 74
4.4 Posteriors for Bayesian Parameter Estimation . . . 76

5.1 Variations of Paid Search ads similar the textual ads used for this study 85
5.2 Hierarchical dependencies for two-way BANOVA (Kruschke,2010) . . 87
5.3 Mean CTR values for ads with different text feature combinations . . 89
5.4 beta value posterior distributions for each variation given the data (**x1 = trust seal informaton** (x1:A1 = no trust info, x1:A2 = unspecified test winner information, x1:A3 = concrete trust seal and test winner information), **x2 = pricing information** (x2:B1 = no pricing, x2:B2 = no deposit, x2:B3 = save x%, x2:B4 = concrete pricing information from database)) . . . 90
5.5 contrasts for various levels of trust seal information ($X1$) . . . 91
5.6 contrasts for various levels of pricing information ($X2$) . . . 91

List of Tables

1.1 List of the Publications 20
1.2 List of the Publications 21

2.1 Values of λ, related selected features and resulting MSE. For our analysis we use $\lambda = 0.11$ which results in only four features and in an MSE that is located in the confidence interval of the lowest MSE 35
2.2 Root mean square errors and 90% highest density intervals (HDI) of the distribution of the predictions on the holdout set based on different methods (HR = Heuristic, LR = Linear regression, PR = Recursive partitioning, SVM = Support vector machine regression, RF = Random forest regression, OA = Our approach)) 39

3.1 Brand keyword clicks and revenues (with standard deviations) in the reference period (data are disguised to ensure confidentiality) 53
3.2 Brand keyword visits, conversion rates and revenues in the test period (data are disguised to ensure confidentiality) 55

4.1 typical dataset from Google AdWords (ad level) 70
4.2 summary of click-rates of ads in the given dataset 72
4.3 Estimated parameters of the A/B Test results 75

5.1 Occurences of text features in the dataset (trust information(x1) and pricing information(x2)) 86

Chapter 1

Introduction

1.1 Introduction

Internet search engines like Google, Yandex, Baidu and Bing play an undisputed key role in the modern information society. On the one hand, they serve the information needs of their users, but on the other hand, they represent an important source of customer acquisition for companies in a broad variety of industries and sizes (Jansen et al., 2009; Jansen and Mullen, 2008). Following the "IAB Internet Advertising Revenue Full Year 2016" (IAB, 2017) report, US Internet advertising spending summed up to approximately $72.5 billion. In the United States alone, Paid Search accounted for 48% of the 2016 advertisers' online advertising budgets at around $35 billion, growing about 19% from 2015 ($29.5 billion). Although the technology landscape in the online advertising industry is constantly changing, one can be reasonably certain that Paid Search will remain an important part of advertisers' marketing mix for a long period.

From a user's perspective, search engines provide results whenever the user enters search queries into their given Internet device web browser. These results are generated individually per query, depending, e.g., on the user's location, prior search behavior, and presumed specific intention of the specific query. The search engine displays generic (also known as "organic") results as well as advertisements on the results page (as can be seen in fig 1.4). The search engine marks these results as advertisements and displays them above and alongside the generic results. Higher positions are sold for higher prices by the search engine companies as they are more likely to receive a click for a given advertisement than a lower position.

These ads provide the search engine companies with significant portions of their revenues. While still growing rapidly, Paid Search Advertising already dominates the online media spending of companies that advertise on the Internet. In this form of advertising, advertisers provide search engines with text advertisements and a list of keywords, which can consist of one or more terms alongside which they would like their ads to be displayed. The advertiser usually also provides a number of attributes for each of these keywords, but at the minimum, advertisers define the amount of money they are willing to pay for a click on an ad for this specific keyword (Jansen, 2011). Every time a user types in a query, the search engine generates individually personalized result pages depending on the user's location, search history and other factors. If ads are available that would likely satisfy the need of the user, the search engine displays these ads alongside the organic results. The amount of money that advertisers are willing to pay for any given keyword strongly depends on the expected economic outcome when somebody has clicked on their advertisement. For example, the keyword "Trading Platform" has an average cost-per-click of more than 25 EUR in Germany. Other keywords such as, for example, "Call Center Job" have an average cost-per-click of less than 0.50 EUR. How is this possible? The easiest way to understand this disparity might be to take these average prices per click as an estimate of the advertisers' consideration of the expected revenue that a given user will generate when entering the website via one of those keywords. Assuming that one out of ten users might convert into a customer for a specific provider of trading platforms in Germany when entering the website via the given keyword, if each of those conversions was worth 250 EUR, it would be rational to invest no more than 25 EUR in the acquisition of a user. But how does the overall process and especially pricing work in this environment?

Yao and Mela (2007) contributed a first comprehensive literature review of Paid Search Advertising from the perspective of three stakeholders: **(1) search engine companies, (2) advertisers, and (3) users**. In my thesis, I concentrate on the advertisers' perspective. Nevertheless, it is important to understand the basics of the market and the motivations of the central stakeholders, which I will illustrate in the following chapter. Furthermore, there are a number of additional stakeholders in this marketing channel, as can be seen in fig. 1.1. Some of those come with privacy and policy concerns while measuring the success of the campaigns (Siebert, 2017), and others have a broad variety of interests within this context.

Qin, Chen, and Liu (2015) present a comprehensive review of Paid Search research. They categorize developments in this area into two basic streams. On the one side, they identify papers in which we generally find assumptions such as fully rational

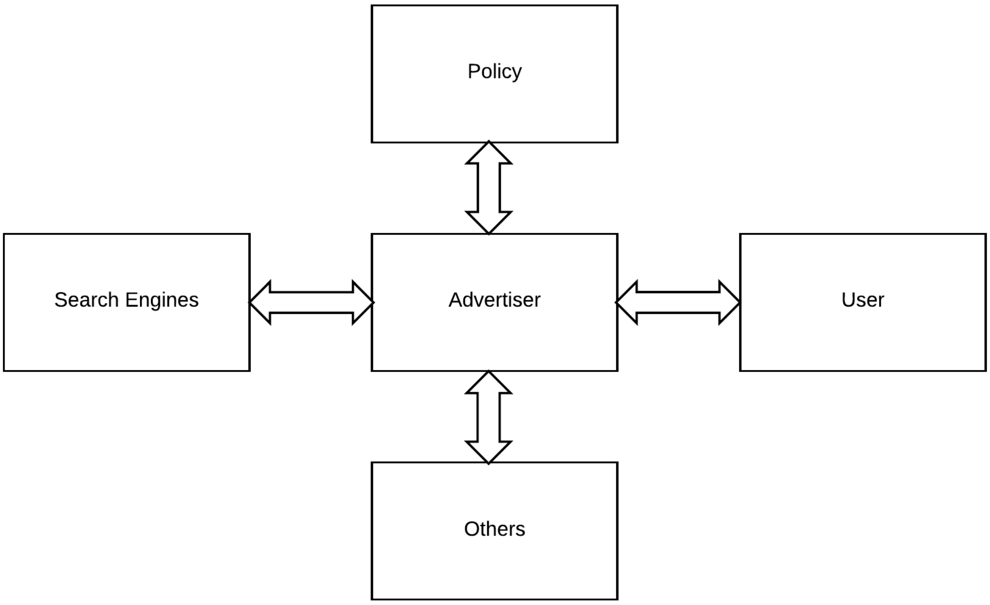

FIGURE 1.1: Stakeholders in Paid Search Advertising from an advertiser's perspective

advertisers that have no constraints in terms of budget. All advertisements have known and independent Click-Through Rates (CTRs), and search queries by users always match the keywords of the advertisers perfectly. On the other side, they identify a large number of papers in which these constraints are relaxed, which makes the research more applicable to practice but often seems to weaken their theoretical foundation. In my contributions, I concentrate on research that is applicable in practice. First, fig. 1.2 illustrates the process and some basic market mechanisms in the context of what it is like to place a keyword on a search engine results page as an advertiser. The figure consists of four parts. In the first part, I describe how the pricing mechanism works. In the second part, I show what else influences the search engine results page. In the third part, I illustrate some of the most significant elements on the advertiser's website itself, and I conclude with the economic evaluation of placing advertisements for a specific keyword.

As an advertiser, you want to be present on the search engine results page whenever somebody enters a query that could result in a conversion on your website. If more than one advertiser is willing to pay for the display of an ad, the search engine auctions the position of these ads among all interested advertisers. In each auction, only the advertiser that is getting a click on an ad is charged by the search engine. As

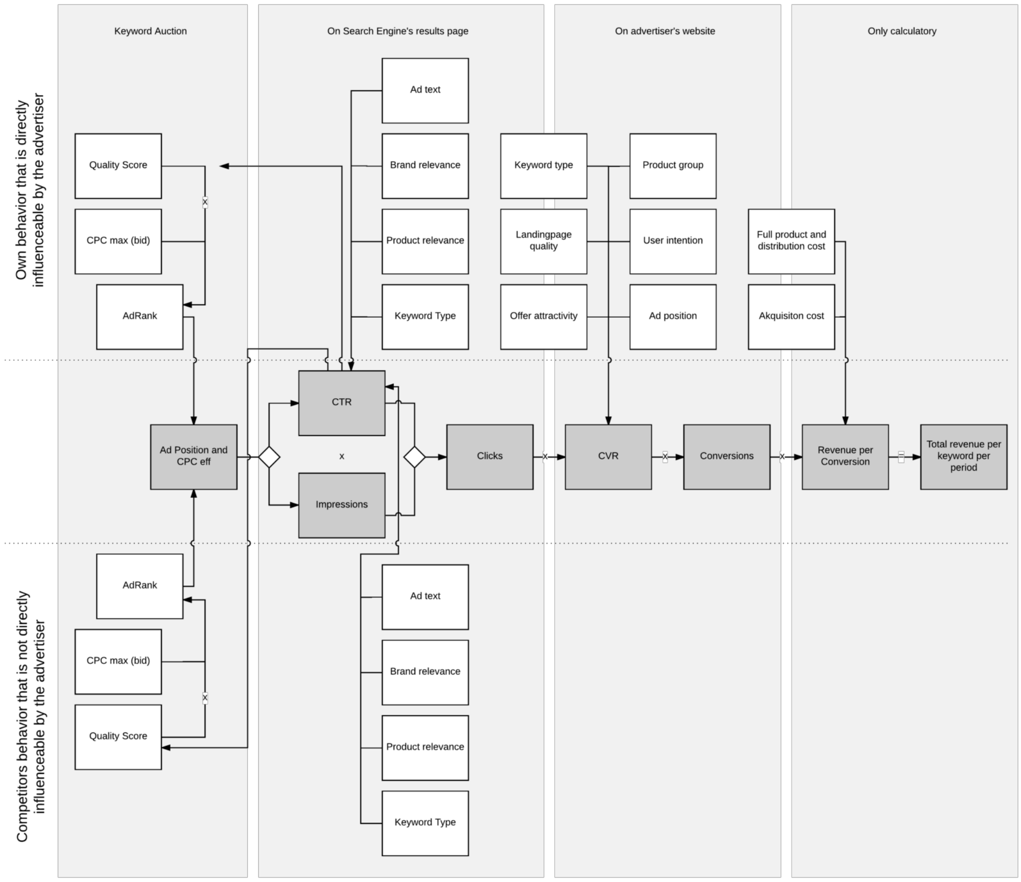

FIGURE 1.2: Simple model of a Sponsored Search process from the the advertiser's perspective

an advertiser, there are generally speaking two instruments to influence the position that your advertisement receives on the search engine results page. The first is the amount of money that you are willing to pay per click, from here on referred to as the "bid"; the second is the optimization of the so-called quality score that the search engines take into consideration in the auctions in order to maximize their profits by rewarding keyword/ad combinations that have a high relevance to users.

The auction design is based on a Generalized Second-Price Auction (GSP). This is the commonly used approach in practice. To provide a better understanding of the underlying mechanism, I will briefly illustrate the basic concept behind GSP auction as described by Qin, Chen, and Liu (2015).

GSP auctions are about pricing and ranking. This auction type usually contains a number of n bidders that want to be displayed in one of k available slots in the auction. In most cases, there are more bidders than available slots ($n > k$). Every bidder i in the auction has a maximum price he or she is willing to pay for a click on the advertisement. This price is usually close to or exactly the expected value of the click to the advertiser v_i. The maximum price is then submitted by the bidder i to the search engine as bid b_i. Advertisers usually take part in more than one auction. For this reason, they submit a multitude of bids to the search engine ($b_{i1}, b_{i2}, \ldots, b_{in}$). Per definition, no advertiser is charged more than b_i in the auction.

Qin, Chen, and Liu (2015) illustrate that the CTR for a given ad is composed from a position-discount term θ_s and an advertiser-specific term q_i that indicates the advertiser's specific probability of getting a click when the given ad has been recognized by a user ($CTR_{i,s} = q_i\theta_s$).

The calculation of prices usually happens in the same way: All bidders in the auction would have to pay the minimum price that they would have had to offer to reach their specific position if their ad was clicked. Following the authors, there are several ways to calculate the position outcome with the underlying position ranking function $y(q_i,b_i)$.

They show two simple but popular variants of this function as follows: $y(q_i,b_i) = b_i$. The advertiser with the bid $i \leq k$ wins position i and has to pay (b_{i+1}). This variant is called rank-by-bid GSP. The other variant that they describe is $y(q_i,b_i) = q_i\, b_i$. The ad positions are ranked by multiplying the previously known or estimated advertiser-specific CTRs and their respective bids. The positions are ordered descending by $q_i b_i$, and for each click they have to pay the price p_i, which is calculated as follows:

$$p_i = \frac{q_{i+1} b_{i+1}}{q_i}$$

In practice, the so-called ad rank that I refer to in fig. 1.2 consists largely of the expected $CTR_{i,s}$. There is a great deal of additional research on optimizations in the auction model from a search engine's perspective. However, from the advertiser's point of view, knowing how prices are set should be sufficient for the moment.

In published research, online marketing, and Sponsored Search especially, has become an established topic with a variety of high quality publications in Computer Science and Information Science as well as in the fields of Operations Research and Marketing. Since 2004, Sponsored Search has continuously become a more important topic in the Online Marketing research area.

On the following pages, we will delve deeper into the interests and motivations of these stakeholder groups and how they affect advertiser behavior.

1.2 The users

Users make the decisions that lead to the success of all the efforts that are taken by search engines and advertisers—it is the users that generate the revenues by first clicking on advertisements and then buying products or services. But what exactly are the decisions that users make, and what impact do these choices have on advertiser behavior? Basically, users are trying to obtain the best insights for their specific concern with the least possible effort. Following Yao and Mela's stakeholder model, users have to make three kinds of decisions while crawling the web with search engines: **(1) Choosing the best engines and keywords, (2) choosing the specific links on the results page and finally (3) deciding whether to make a purchase on the website of a given advertiser**.

The authors illustrate that the way users make decisions in search engines can be observed in other contexts. Similar selection processes were described by Stigler in the 1960s (Stigler, 1961; Stigler, 1962) and Weitzman in the late 1970s (Weitzman, 1979). They describe sequential and non-sequential types of searches. A sequential search is a linear process in which the searcher considers one result after another until he or she has found something that satisfies his or her needs. The non-sequential search process, however, begins with choosing a number of results that will finally be considered by the searcher. Afterwards, the searcher examines all these results and selects the best option within this set. Yao and Mela compare the Internet search process to a conventional shopping trip where users choose relevant stores and categories before deciding for a specific outlet at which to buy their product. They transfer the described behavior to the Internet search engine with the same two types of search patterns, but adopt these in a more concrete way for the Sponsored Search process. In terms of the **sequential** pattern, they simply describe users that investigate one search engine after another. The adoption of the **non-sequential** pattern is characterized by users who choose a number of search engines at a time and evaluate the results by opening a number of these websites at the same time, having all of them in their current consideration set. In practice, we find a mix of these two patterns. It is especially interesting to compare a user's search behavior when it comes to different product types and complexities. The authors describe that the perceived complexity of the topic of interest may have an impact on the way that search engines are used, especially when it comes to the possible size of the consideration set.

After finding the optimal search engines for their specific needs, users select the best possible results on the search engine's result pages. This specific user decision directly impacts the advertiser and search engine behavior. In practice, search engines often provide users with guidance concerning the best suited link. They do this by ordering the results in terms of a best possible guess on what might satisfy the needs of this specific user in terms of individual demographics, geographic location, previous queries, etc. The advertising placements with the highest click probability on the result pages are more valuable to the search engine and the advertiser than any others. For the search engine, this is the placement that eventually produces the highest revenue in a pay-per-click model. It could make sense to sell this placement to the advertiser willing to pay the most for a click. To maximize their long term revenues, the search engine has to provide users with the best possible results at every visit, which may sometimes be in conflict with its own short term monetary interests.

1.3 The Internet search engines

Search engines have two goals they need to balance to reach their business goals. First and foremost, they have to attract people to use the search engine for their queries. In fact, they have a vital interest in having as many users as possible on their websites as this is the only way for them to become relevant for advertisers and their advertising budgets. Therefore, they need to provide the best possible experience for their users. Second, they have to be an attractive partner for advertisers to ultimately maximize their revenues. This becomes especially interesting for them as they usually offer their services completely free of charge to users. The search engines have to generate revenues mainly from monetizing user queries. Search engines went through interesting developments over the past 20 years to get to a point where this business model became successful. This becomes evident when examining their result pages over time. The first figure shows a screenshot of the world's leading search engine, Google, as it looked in 1998 (fig. 1.3). We find only generic results on the website and no obvious monetization model. Since overloaded web directories were the industry standard at the time, search engines, especially Google, had a very strong simplification approach, following the simple goal of providing the best suitable result in the most comfortable way to the user.

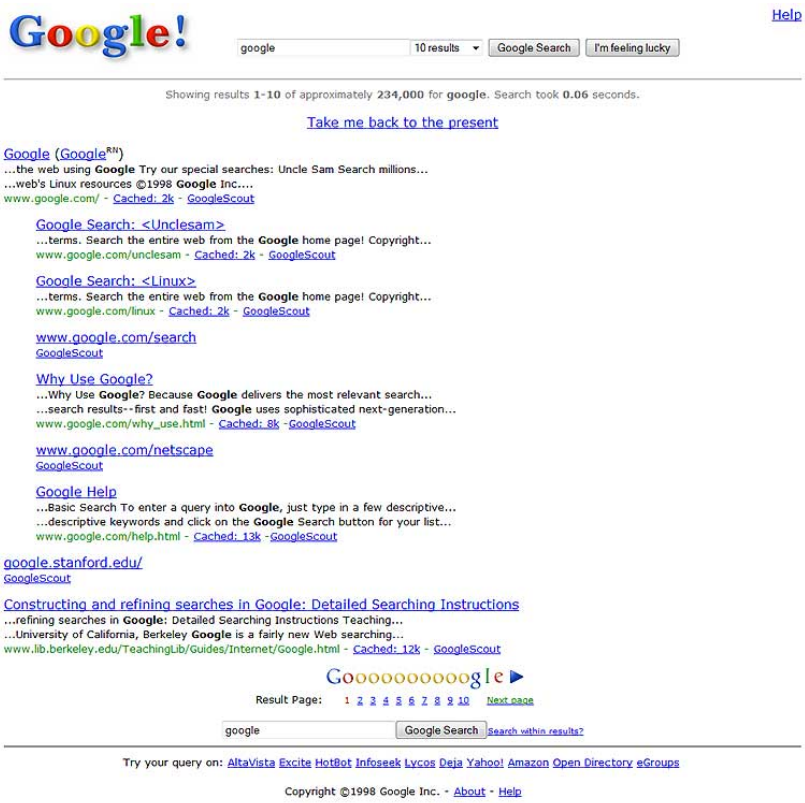

FIGURE 1.3: Screenshot of the appearance of Google's search engine in 1998

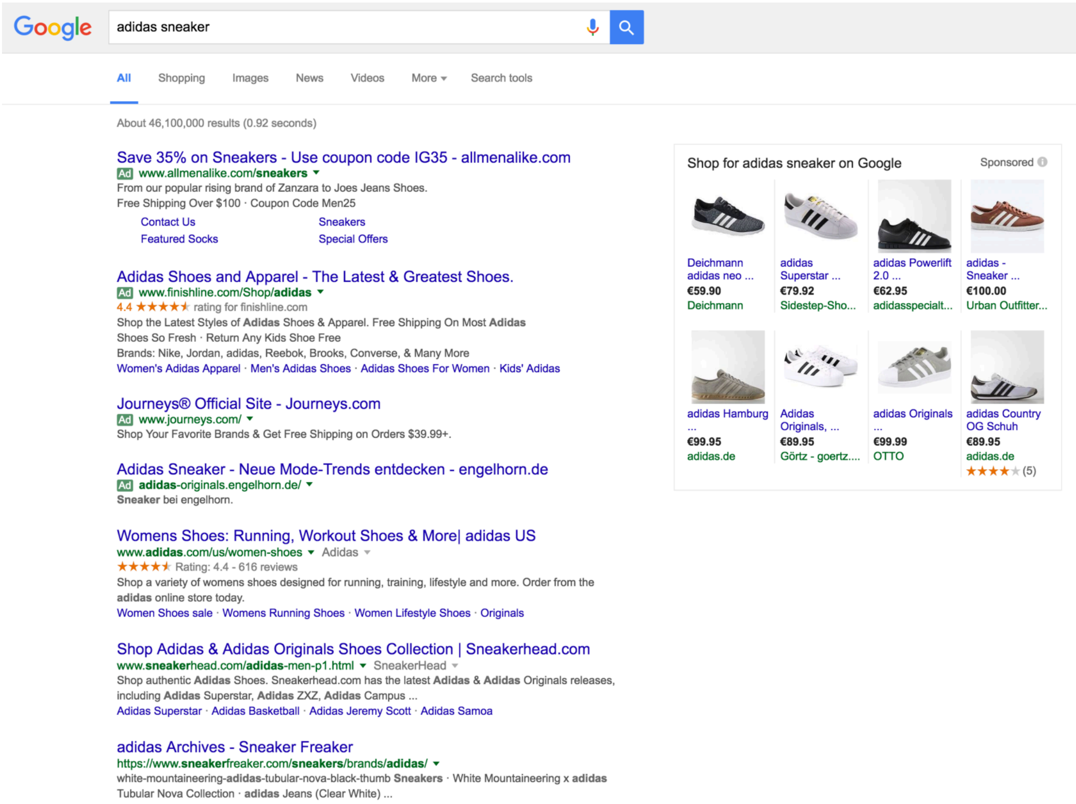

FIGURE 1.4: Google Search Engine Results Page 2016

As stated, this service was provided free of charge for users and did not generate substantial revenues to the search engine company. But that changed over time. A recent screenshot of the same search engine illustrates how the business model developed. In fig. 1.4, we find that the search engine mainly displays advertisements to the user on the largest part of the visible area of the results page.

Considering the central goal of being attractive for users and advertisers, research on a search engine's topics focuses on revenue maximization in the context of this thesis.

Research on the basic design of Paid Search auctions is at the core of many publications. A number of theoretical contributions can be found that develop game-theoretic approaches to equilibrium optimization in GSP auctions as illustrated by various authors (Bu, Deng, and Qi, 2007; Edelman, Ostrovsky, and Schwarz, 2007; Cigler, 2009). There are discussions about the introduction of minimum prices (Gonen and Vassilvitskii, 2008) or the possible benefits of hybrid auction types. In these auctions, the advertisers are given the choice of being charged on a CPC or Cost-Per-Mille (CPM) (Zhu and Wilbur, 2011) or Cost-per-Action (CPA) basis (Edelman and Lee, 2008). None of the important players in the search engine market implement such a choice for advertisers, as they expect their revenues to decline. But advertisers are given the option to set goals on a CPA level, for example, while still being charged by CPC in the Google search engine. Amaldoss, Desai, and Shin (2015) examine the impact on the search engine's profits from the introduction of first-page bid estimates instead of minimum prices for their GSP auctions.

One of the problems that search engines and advertisers share in practice concerns the **long-tail**, which can be explained by the existence of a gap between the very few popular keywords among advertisers (head keywords) and a large number of keywords with only a low degree of competition between advertisers. Those keywords are often not frequently demanded by users. Dasdan et al. (2009) investigate the opportunities that could arise if search engines allowed their advertisers to specify factors other than a list of keywords (input bidding) that determine the delivery of an advertisement (output bidding). This includes booking categories and target groups, which would enable search engines to supply search engine Result Pages (SERPs) with ads that have not been in great demand by advertisers. In practice, this is basically implemented in the context of matchtypes (broad, modified broad, phrase, exact and negative) and to an even greater extent with shopping feeds and remarketing.

Various authors explore **optimization problems with a focus on budget constraints** of the bidders in the auctions (Colini-Baldeschi et al., 2015; Borgs, Chayes, and Immorlica, 2005; Zhou, Chakrabarty, and Lukose, 2008; Abrams, Mendelevitch, and Tomlin, 2007). In practice, search engines have to satisfy the user's query by offering relevant results. Here, we find another example for the triage of interests as the quality of the identification of a query intent is important for search engines first of all, but it is also relevant to advertisers as they have to decide on keywords that have a good chance of ultimately leading to a conversion. This indicates the relevance of having algorithms that classify queries (Jansen, Booth, and Spink, 2008; Guo and Agichtein, 2010; Ashkan A., 2009; Ashkan and Clarke, 2013; Attenberg et al., 2009; Weber and Jaimes, 2011) along "navigational", "informational" and "commercial" dimensions as well as mixed forms of these. In fact, only a small fraction of all queries have a direct commercial background, which makes it important for advertisers to correctly identify these terms in order to avoid wasting expenditure on visitors who have no intention of converting into customers. The better these intentions for each individual query session are understood by search engines, the higher the chance that search engines can precisely match the needs of the current user and can display appropriate advertisements.

1.4 The advertisers

One central topic of interest for advertisers concerns the factors that influence whether the exposure of an ad to a group of users leads to visits to the advertiser's website. What is even more interesting is whether these visits ultimately lead to conversions. My contribution to these topics is presented in **chapter 2**, where my co-authors and I present a new way to predict conversion probabilities based on user behavior, and in **chapter 3**, in which we investigate the economic impact of paid search advertising on brand terms. I also present two consecutive papers in which I investigate the influence of specific text features in Paid Search advertisements on click probabilities in **chapters 4** and **5**.

1.4.1 Click-Through Rates on the Search Engine Results Page

In the second part of fig. 1.2, I illustrate some features that can impact the likelihood that a user will click on an advertisement and be directed to the advertiser's website. This topic has been investigated intensely since the beginnings of the Sponsored Search advertising format.

Four factors can be clearly identified to greatly influence the click probabilities for any given ad/query combination. These are the position of the ad, the perceived relevance of the presented ads, keyword types and the content of the ad.

Various authors find that the **ad position** within the Sponsored Search results has a major influence on its CTR. A number of studies show a correlation between position and CTR (e.g. Richardson, Dominowska, and Ragno, 2007; Agarwal, Hosanagar, and Smith, 2011). This effect is basically undisputed (Narayanan and Kalyanam, 2015) and has received intense research in the past. From an advertiser's perspective, predicting future $CTRs$ for their keyword combinations is of interest. Zhu et al. (2010) even find a significant effect in terms of conversion probability for the top two positions on SERPs.

Users have specific search and crawl patterns on SERPs. This knowledge can be used to help predict click probabilities of results on these pages. Craswell et al. (2008)

present several models for predicting the CTR: the baseline, mixture, examination, and cascade models. The findings were based on organic search results but they are applicable to sponsored search results, as well. The underlying assumption of the baseline model is that a user screens every search result and decides afterwards which one is most appropriate to the query. As a consequence, the click probabilities for each individual search result are identical, independent of their position. The mixture model extends the baseline model and divides user behavior into two groups. One group behaves as described in the baseline model, while the other group randomly clicks on one of the first search results. The examination model refers to findings from eye-tracking studies that state that with declining position, the probability of a click declines as well (Joachims et al., 2005; Joachims et al., 2007). The cascade model is one of the most applied explanation approaches because of its strong ability to explain click data. The basic assumption is that the user scans each search result, from the top to the bottom, comparing the relevance of each ad with the relevance of the previous ad. The user continuously scans the results until the perceived ad relevance reaches a certain level, at which time the user clicks. Jeziorski and Segal (2015) find in their study based on user-level data from Microsoft Live that users tend to click ads in a nonsequential order and that the CTR depends on the identity of the respective ads.

Following Ghose and Yang (2009), the **keyword type** also influences the CTR. According to the study, keywords with retailer-specific content lead to a significant increase in CTR. One challenge is to predict the CTR of keywords or keyword combinations for potential future Sponsored Search ads, much as it is for the search engine companies to maximize their revenues in Paid Search auctions (Gatti et al., 2015). One solution that has been proposed is aggregating historical data from similar keywords (Regelson and Fain, 2006). Here, the CTR is represented as a function of the **position**. The same clustering approach can be applied in optimizing the search engines' profit (Dave and Varma, 2010). There are also models taking the quality score into account (Ilya, 2010; Dembczynski, Kotlowski, and Weiss, 2008). A model developed by Zhu et al. (2010) called the General Click Model focuses on the CTR prediction of long-tail queries, based on a Bayesian network. Dealing with the position bias mentioned before, Zhong et al. (2010) incorporate post-click user behavior data from the respective landing page of the clicked ad into the click model to refine the estimation of the perceived user relevance after clicking on a specific ad. A similar approach using Dynamic Bayesian networks can be found in Chapelle and Zhang (2009). Several models based on historical click data suffer from limitations in terms of failing to consider a possible user-learning effect. Taking Gauzente's results as an example, it has been shown that past user satisfaction with Sponsored Search results influences the current click behavior (Gauzente, 2009). Based on user-level data,Zhang et al. (2014) are able to illustrate sequential click preferences using Recurrent Neural Networks. This level of data, however, is usually only available to the search engine companies themselves and not to advertisers and is thereby not the focus of my thesis.

Besides the incorporation of position data and the perceived relevance of presented ads, the CTR of an ad is also affected by the relationship between Organic and Sponsored Search results. Listing the results of one company at the same time in Sponsored and Organic Search results seems to lead to a higher CTR (Yang and Ghose, 2010) for the paid results. In two papers of this thesis, I empirically investigate the importance of ad content. Wang et al. (2013) find that the existence of specific text patterns in text advertisements in Paid Search advertising influences the likelihood that a given user will click on the ad. In **chapter 4**, we develop and perform a non-reactive A/B-test that enables us to evaluate the influence of sustainability information on the customer's decision to buy a product by clicking on an ad on a SERP. I analyze campaign performance data generated from a European e-commerce retailer and apply a Bayesian parameter estimation to compare the two groups. This study shows that this type of content has an impact on user decisions to click ad buys. More concretely, the findings in **chapter 4** show that the content of an ad has a large impact if it is relevant to the user. Paid Search Advertisers have very few options to influence the user's decision to click on one of their ads. The textual content of the creatives seems to be one important influencing factor, beneath its position on the SERP and the perceived relevance of the given ad to the present search query. In the study in **chapter 5**, we perform a non-reactive multivariate test that enables us to evaluate the influence of specific textual signals in Paid Search creatives. In this case, however, we do not use sustainability information but rather concrete prices as a variant. A Bayesian analysis of variance (BANOVA) is applied to evaluate the influence of various text features on click probabilities. In this case, we conclude by finally showing that differences in the formulation of the textual content can have a substantial influence on the click probability of Paid Search ads.

1.4.2 Conversion Rates at the advertiser's own domain

In most cases, an advertiser wants to generate revenue by selling products or services through Sponsored Search. So, the prediction of a user's likelihood to convert is a major issue in the field of e-commerce. The detailed definition of a conversion depends on the specific website intention but is mostly some type of purchase or membership. Following the results of several studies, the five factors below are among the most important in influencing the conversion rate, CVR. These factors are keyword characteristics, ad position, user intention, quality and content of the landing page and product type.

Obviously, not all the listed factors are directly linked to Sponsored Search. Nonetheless, they have an influence on the conversion rate as will be argued later in this section. It is necessary to keep in mind that the keyword type and ad position influencing factors as well as the quality of the landing page and the included product group are all interconnected. Acknowledging this fact, several studies analyze a combination of these factors. With regard to **keyword characteristics**, the presence of brand information in the keyword decreases the conversion rate; a study of user data from US-online retailers by Ghose and Yang (2009) reveals a rate decrease in the analyzed dataset of more than 40%. Conversely, a retailer-specific keyword leads to an increase. Rutz, Bucklin, and Sonnier (2012) show that the conversion rate is likely to be higher for branded than for generic keywords.

Some authors find that the **ad position** has an influence on conversion probabilities. Agarwal, Hosanagar, and Smith (2011) state that for longer keywords, the conversion rate initially increases but then decreases with the ad position. This effect does not occur, however, with shorter keyword phrases (head), which show continuously decreasing CVR with decreasing position.

Ghose and Yang (2010) find that the CVR is also influenced by a combination of position and product type. Based on a dataset of US retailers, it has been found that several product categories show higher position sensitivity with regard to the conversion rate than others. Similarly to keyword characteristics and ad position, the third factor, **user intention** has a major influence on the conversion rate, as has already been mentioned in the section on market mechanisms. Montgomery et al. (2004) develop a model to predict conversions depending on path data distinguishing between two types of user intention, browse (no real purchase intention) and deliberate (focused purchase intention).

The influence of the **landing page quality** on the conversion probability seems to depend on the **product type**. For example, Huang, Lurie, and Mitra (2009) distinguish goods into experience and search goods. Based on a dataset of more than 50,000 households, they found that providing users with feedback from other users who have already purchased the respective product leads to an increased likelihood of the product being purchased. Furthermore, the analysis of the dataset revealed different degrees of depth and breadth of the search, depending on the intended product type. A limited number of different page visits (breadth) but longer time frames for each visit (depth) are characteristic of the search behavior of customers looking for experience goods. These findings emphasize the relevance of the product type as well as the design and usability of the online shop as an influencing factor on conversion probability. Ye, Aydin, and Hu (2014) even find evidence that increasing and decreasing prices on the landing pages relative to the customers demand in paid search makes sense for advertisers when optimizing their own earnings.

In practice, the search engine advertising accounts of big spenders often consist of hundreds of thousands of keywords that are organized into a large number of ad groups and campaigns. A large proportion of the available media budget is allocated to only the very small fraction of keywords that are responsible for the largest part of the revenue of the advertiser. The advertiser has only very few observations in terms of conversions, clicks or even impressions for most keywords in his account.

In the paper in **chapter 2** of this thesis, we investigate the prediction of conversion probabilities based on what has been called "landing page quality" by other authors. This helps to, first, solve the problem of keyword selection for advertisers (fig. 1.2) and might also support them finding the optimal bid for these keywords in the auction. As mentioned above, choosing the right keywords has an outsize influence on the success of campaigns in Paid Search Advertising. To do so, we estimate click and conversion probabilities for each keyword. In practice, this is easy for frequently demanded keywords but a major challenge in the long-tail, where only relatively few observations per element can be made within a given period. We address this widespread data sparseness problem in **chapter 2**. We use information concerning the average time users spend on the advertiser's website and bounce rates for every given keyword. This previously unused data is shared between all keywords and used as prior knowledge in our proposed CVR prediction model. The developed model enables advertisers to predict individual keyword conversion probabilities in practice significantly better than with commonly used keyword quality assessment approaches and develops a new approach that is used in the core of a PPC Bid Management Software that optimizes millions of keywords for several advertisers from various industries.

1.4.3 How can advertisers maximize their revenues?

One major challenge is to identify the incremental contribution of Sponsored Search ads to conversions in the context of multi-touchpoint advertising user journeys in combination with other formats such as display advertising. Often, the users also receive various additional advertising formats such as banner advertising. There is a strong indication that their sole impression might influence the users decisions as well. Recent studies have shown that banner ads that have been displayed but not clicked can influence future user behavior (Chatterjee, 2008). Adapting this result to the models presented in my thesis, the return of a conversion can be seen as a relative value whose actual amount depends on the degree of influence that various touchpoints have on the conversion. Although user journey cross channel effects are outside the scope of this thesis, it is important to keep them in mind for future research on the profitability of Sponsored Search (Dinner, Van Heerde, and Neslin, 2014; Joo, Wilbur, and Zhu, 2016; Yang et al., 2016). In many cases, the questions are still directly related to optimizing the marketing channel Sponsored Search. Gaining and maximizing revenues is the central interest of advertisers in this context. The return of a conversion, which depends on the product price and the acquisition costs of the conversion, is a major factor for future investments in certain advertising formats (Szymanski and Lee, 2006). Only few authors concentrate on this topic. In fact, only a few publications have studied revenue maximization from an advertiser's perspective (Auerbach, Galenson, and Sundararajan, 2008). In some cases, Paid Search might not work at all. Schlangenotto and Kundisch (2017) find evidence that there is no measurable effect when it comes to the effectiveness of Paid Search Advertising for brick-and-mortar businesses.

Revenue maximization is often the objective when it comes to keyword bidding strategies. Some early work has been done by Kitts and Leblanc (2004), who attempt to maximize profits given a budget constraint. Applying their model to real world keyword auctions, they were able to increase clicks by a factor of four, given a specific budget.

The best-response bidding strategy by Cary et al. (2007), for example, illustrates that the best path for any given bidder in the next round of a repeated auction is to place bids that will probably maximize their utility, simply assuming that the bids of the the competitive bidders remain the same in the next round of the auction. The authors also present a so called balanced bidding strategy in their paper. The interesting feature of this strategy is that it takes multiple possible goals of the advertisers into account. Referring to the balanced bidding strategy, the advertisers want to maximize their own utility while also causing maximal harm to their competitors in the auction. Given the GSP principle in the Paid Search auctions, it could be applied by placing bids that are high enough to make the competitor ranked directly above the advertiser pay as much as possible for a click on their ad while simultaneously preventing a change in positions, as this would lead to the opposite outcome. Kominers (2009) find a generalization for the balanced bidding strategy in the context of user search behavior and a dynamic position auction. Vorobeychik and Reeves (2008) evaluate various strategies like the balanced bidding strategy, a cooperative strategy and a collusion strategy. They demonstrate that the balanced bidding strategy is highly stable to deviations in both the advertiser's and the competitor's behavior. In practice, we learned that bidding at scale usually works best when placing bids selfishly while ignoring the competitor's behavior completely.

Using additional information for the bidding process is suggested by Wang, Li, and Kaafar (2016), who propose using social media trends to optimize Paid Search bidding portfolios. This idea is also applied in our research in **chapter 2**, although we use other data from the advertiser's campaigns and the user's on-site behavior instead of information from social media and trends.

Borgs et al. (2007) present a bidding algorithm that optimizes the utility for bidders by equalizing the return-on-investment for each advertiser across all keywords and compares approaches in first- and second-price auctions.

Another research stream focuses on the outcome comparison of Sponsored and Organic Search results as presented in the section about CTR with respect to the keyword type, e.g., branded or generic (Ghose and Yang, 2008). Based on a comprehensive dataset of US-retail chain advertising on Google, similar to the findings on the influence of the keyword type on CVR, queries with retailer-specific keywords also lead to an increase in order value and revenue. Following the influence of the keyword on the revenue, Rusmevichientong and Williamson (2006) develop an algorithm to select the profit-optimizing keyword-set under budget constraints. However, profit in this case is based only on expected $CTRs$, with no further attention paid to $CVRs$.

Besides demonstrating the influence of the keyword type on revenue, research on the relation between ad position and profit reveals equally interesting findings. For longer keywords, in contrast to the CTR, profit first increases and then decreases with ad position. These results are in contrast to the widespread assumption that a higher position has a higher value for the advertiser (Agarwal, Hosanagar, and Smith, 2011).

The interplay between organic and paid results on SERPs is of high interest for a number of researchers (Jerath, Ma, and Park, 2014). Agarwal, Hosanagar, and Smith (2015), for example, find that an increase in organic competition on SERPs leads to a decrease in the performance of Paid Search ads of an advertiser. The objective of the paper in **chapter 3** is to determine whether and under what circumstances it makes sense, in economic terms, for brand owners to pay for sponsored search ads for their brand keywords. This issue is the subject of a heated debate in business practice, especially when the company is already placed prominently in the organic search results. In this paper we describe and apply a non-reactive method that is based on an A/B-test. It was employed in a case study of a European Internet pharmacy. The results of this study indicate that the use of sponsored search advertising for their own brand name enables advertisers to generate more visitors (>10%), resulting in higher sales volumes at relatively low advertising costs even when the company is already listed in first position in the organic part of the respective SERP. Interestingly, another experimental study by Blake, Nosko, and Tadelis (2015) has completely opposite results. Experimental results at eBay illustrate that in this case, almost all brand-term Paid Search conversions are substituted by organic traffic leading to almost the same amount of conversions. In fact, they find that for eBay, Paid Search has almost no significant effect on the number of conversions they generate over all. This paper is also interesting due to the current trend in Marketing research to apply large scale experimental research setups. Gordon et al. (2016) also compare the results from randomized controlled trial experimental research with other commonly used techniques to measure the effect of advertising in their research based on massive data from Facebook ad campaigns.

1.5 Data sets in Paid Search

All papers in my thesis are based on real advertising campaign data. But which data are available to researchers and practitioners in this area? In the following, I will give a brief overview of the available data sources. This data may be produced by the advertisers or the search engines themselves. Generally, quantitative research is conducted with three types of datasets: (1) Search engine query data, (2) aggregated media and e-commerce statistics and (3) individual user journey data.

Search engine query datasets are the rarest form of available data for researchers who are not directly affiliated with the search engines, as this type of data can only be collected by the search engine companies themselves. Although every search engine company generates masses of this type of data, few datasets are available for academic use. One of those is the well-known AOL dataset. It consists of approximately 20 million completely non-censored web queries collected from about 650,000 users over a three month period, arranged by anonymous individual IDs. This dataset has been extensively examined since 2006 (e.g., by Pass, Chowdhury, and Torgeson, 2006; Adar, 2007; Strohmaier, Prettenhofer, and Kröll, 2008; Brenes, Gayo-Avello, and Pérez-González, 2009).

Aggregated media and e-commerce statistics are generated by the advertisers themselves during their ad campaigns. One way this kind of data is produced is by the campaigning tool itself (e.g., Google AdWords) or the advertiser's respective software solution. The data are usually aggregated on the campaign, ad group and keyword levels and contain variables like the total number of clicks, impressions, CTR, CPC, and CVR. This is the type of data I used for my research in this thesis.

The third sort of available data enables researchers to understand individual user behavior. Individual user journey datasets include information about all measured touch-points that an specific user has with an advertiser. These datasets make the development of attribution-models possible where every conversion success can be allocated to the ad contacts a user had (Kireyev, Pauwels, and Gupta, 2016). Like the other types of data, user journey data are always subject to several types of bias, such as bias caused by media discontinuities.

1.6 Summary

In my thesis, I contribute to selected scientifically and practically important problems that advertisers face when they have to make decisions in one of their most important marketing channels: Paid Search. I use aggregated e-commerce media data and primarily Bayesian Statistics to develop a model that helps advertisers predict click probabilities in **chapter 2**. Then, I contribute to the area of revenue optimization in **chapter 3** by developing an alternative approach to A/B testing the success of marketing activities. Finally, I shed light on aspects of the influence of textual contents on click probabilities in **chapters 4** and **5**.

TABLE 1.1: List of the Publications

Title	Authors/Journal/Proceedings	Status	In this Thesis
Should Companies Bid on their Own Brand in Sponsored Search ?	Blask, T., Funk, B. & Schulte, R. 2011 Beitrag in International Conference on E-Business, Sevilla, Spanien Juli 18, 2011 - Juli 21, 2011. in Marca, D. A., Shishkov, B. & van Sinderen, M. (Hrsg.): ICE-B 2011: Proceedings of the International Conference on e-Business SciTePress Digital Library (S. 14-21).	Published	No
To Bid or Not To Bid? Investigating Retail-Brand Keyword Performance in Sponsored Search Advertising	Blask, T-B., Funk, B. & Schulte, R. 2012 E-Business and Telecommunications. Obaidat, M. S., Sevillano, J. L. & Filipe, J. (Hrsg.). Springer-Verlag, Band 314, S. 129-140 (Communications in Computer and Information Science; Band 314)	Published	Yes
Bayesian Parameter Estimation in Green Business Process Management: A Case Study in Online-Advertising	Blask, T-B. 2013 Beitrag in Informatik 2013, Koblenz, Deutschland September 16, 2013 - September 20, 2013. in Hornbach, M. (Hrsg.): INFORMATIK 2013 - Informatik angepasst an Mensch, Organisation und Umwelt Gesellschaft für Informatik (S. 852-863). (GI-Edition: Lecture notes in informatics; Band 220)	Published	Yes

TABLE 1.2: List of the Publications

Title	Authors/Journal/Proceedings	Status	In this Thesis
Applying Bayesian Parameter Estimation to A/B Tests in e-Business Applications: Examining the Impact of Green Marketing Signals in Sponsored Search Advertising	Blask, T-B. 2013 Beitrag in International Conference on E-Business, Reykjavik, Island Juli 29, 2013 - Juli 31, 2013. in Obaidat, M. S. (Hrsg.): DC-NET 2013 ICE-B 2013 OPITICS 2013 ICETE 2013: Proceedings of the 4th International Conference on Data Communication Networking, 10th International Conference on e-Business and 4th International Conference on Optical Communication SciTePress Digital Library (S. 312-319)	Published	No
Investigating the Promotional Effect of Green Signals in Sponsored Search Advertising Using Bayesian Parameter Estimation	Blask, T. 2014 in: Information Technology in Environmental Engineering Springer-Verlag . (Environmental Science and Engineering)	Published	No
Do Specific Text Features Influence Click Probabilities in Paid Search Advertising?	Blask, T-B. 2014 in: Proceedings of the International Conference on e-Business (ICEB), Vienna, Austria, 28-29 August, 2014, ICE-B is part of ICETE, the 11th International Joint Conference on e-Business and Telecommunications.	Published	Yes
Predicting conversion rates in paid search using sparse keyword-level data	Blask, T., Stange, M.,Borck, C. & Funk, B (2017), Working Paper	To be submitted	Yes

References

Abrams, Zoe, Ofer Mendelevitch, and John Tomlin (2007). “Optimal delivery of sponsored search advertisements subject to budget constraints”. In: *Proceedings of the 8th ACM conference on Electronic commerce - EC '07*. New York, New York, USA: ACM Press, p. 272.

Adar, E (2007). “User 4xxxxx9: Anonymizing query logs”. In: *Proc of Query Log Analysis Workshop, International Conference on World Wide Web - WWW '07*.

Agarwal, Ashish, Kartik Hosanagar, and Michael D Smith (2011). “Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets”. In: *Journal of Marketing Research* 48.6, pp. 1057–1073.

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2015). “Do Organic Results Help or Hurt Sponsored Search Percaformance?” In: *Information Systems Research* 26.4, pp. 695–713.

Amaldoss, Wilfred, Preyas S. Desai, and Woochoel Shin (2015). “Keyword Search Advertising and First-Page Bid Estimates: A Strategic Analysis”. In: *Management Science* 61.3, pp. 507–519.

Ashkan, Azin and Charles L. A. Clarke (2013). “Impact of query intent and search context on clickthrough behavior in sponsored search”. In: *Knowledge and Information Systems* 34.2, pp. 425–452.

Ashkan A., Clarke C L A (2009). “Characterizing commercial intent”. In: *International Conference on Information and Knowledge Management, Proceedings*, pp. 67–76.

Attenberg, Josh et al. (2009). *Modeling and Predicting User Behavior in Sponsored Search*, pp. 1067–1075.

Auerbach, Jason, Joel Galenson, and Mukund Sundararajan (2008). “An empirical analysis of return on investment maximization in sponsored search auctions”. In: *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising - ADKDD '08*. New York, New York, USA: ACM Press, pp. 1–9.

Blake, Tom Thomas, Chris Nosko, and Steven Tadelis (2015). “Consumer Heterogeneity and Paid Search Effectiveness: a Large Scale Field Experiment”. In: *Econometrica* 83.1, pp. 155–174.

Borgs, Christian, Jennifer Chayes, and Nicole Immorlica (2005). “Multi-unit auctions with budget-constrained bidders”. In: *6th ACM Conference on Electronic Commerce*. Vancouver, BC, Canada.

Borgs, Christian et al. (2007). “Dynamics of bid optimization in online advertisement auctions”. In: *Proceedings of the 16th international conference on World Wide Web - WWW '07*. New York, New York, USA: ACM Press, p. 531.

Brenes, David J., Daniel Gayo-Avello, and Kilian Pérez-González (2009). "Survey and evaluation of query intent detection methods". In: *Proceedings of the 2009 workshop on Web Search Click Data - WSCD '09*. New York, New York, USA: ACM Press, pp. 1–7.

Bu, T M, X Deng, and Q Qi (2007). "Dynamics of strategic manipulation in ad-words auction". In: *3rd Workshop on Sponsored Search Auctions*. Vol. 8. Citeseer.

Cary, Matthew et al. (2007). "Greedy bidding strategies for keyword auctions". In: *Proceedings of the 8th ACM conference on Electronic commerce - EC '07*, pp. 262–271.

Chapelle, Olivier and Ya Zhang (2009). "A dynamic bayesian network click model for web search ranking". In: *Proceedings of the 18th international conference on World wide web - WWW '09*. New York, New York, USA: ACM Press, p. 1.

Chatterjee, Patrali (2008). "Are Unclicked Ads Wasted ? Enduring Effects of Banner and Pop-Up Ad Exposures on Brand Memory and Attitudes". In: *Journal of Electronic Commerce Research* 9.1, pp. 51–61.

Cigler, L. (2009). *Semester Project: Bidding Agent for Advertisement Auctions*. Tech. rep., pp. 1–11.

Colini-Baldeschi, Riccardo et al. (2015). "On Multiple Keyword Sponsored Search Auctions with Budgets". In: *ACM Transactions on Economics and Computation* 4.1, pp. 1–34.

Craswell, Nick et al. (2008). "An experimental comparison of click position-bias models". In: *Proceedings of the international conference on Web search and web data mining - WSDM '08*. New York, New York, USA: ACM Press, p. 87.

Dasdan, Ali et al. (2009). "Output Bidding: A New Search Advertising Model Complementary to Keyword Bidding". In: *Fifth Workshop on Ad Auctions (in conjunction with the ACM Conference on Electronic Commerce EC'09)*. Stanford InfoLab, pp. 1–5.

Dave, Kushal S. and Vasudeva Varma (2010). "Learning the click-through rate for rare/new ads from similar ads". In: *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. New York, New York, USA: ACM Press, p. 897.

Dembczynski, Krzysztof, Wojciech Kotlowski, and Dawid Weiss (2008). "Predicting Ads' Click-Through Rate with Decision Rules". In: *Workshop on Targeting and Ranking in Online Advertising, WWW '08*.

Dinner, Isaac M., Harald J. Van Heerde, and Scott A. Neslin (2014). "Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising". In: *Journal of Marketing Research* 51.5, pp. 527–545.

Edelman, Benjamin and Hoan Soo Lee (2008). "CPC/CPA hybrid bidding in a second price auction". In: *Harvard Business School Working Paper*.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz (2007). "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords". In: *American Economic Review* 97.1, pp. 242–259.

Gatti, Nicola et al. (2015). "Truthful learning mechanisms for multi-slot sponsored search auctions with externalities". In: *Artificial Intelligence* 227, pp. 93–139.

Gauzente, Claire (2009). "Information search and paid results — proposition and test of a hierarchy-of-effect model". In: *Electronic Markets* 19.2-3, pp. 163–177.

Ghose, Anindya and Sha Yang (2008). "Comparing performance metrics in organic search with sponsored search advertising". In: *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising - ADKDD '08*. New York, New York, USA: ACM Press, pp. 18–26.

– (2009). "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets". In: *Management Science* 55.10, pp. 1605–1622.

– (2010). "Modeling Cross-Category Purchases in Sponsored Search Advertising". In: *SSRN Electronic Journal*.

Gonen, Rica and Sergei Vassilvitskii (2008). "Sponsored Search Auctions with Reserve Prices: Going Beyond Separability". In: *Internet and Network Economics*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 597–608.

Gordon, Brett et al. (2016). "A Comparison of Approaches to Advertising Measurement : Evidence from Big Field Experiments at Facebook".

Guo, Qi and Eugene Agichtein (2010). "Ready to buy or just browsing?" In: *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. New York, New York, USA: ACM Press, p. 130.

Huang, Peng, Nicholas H Lurie, and Sabyasachi Mitra (2009). "Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods". In: *Journal of Marketing* 73.2, pp. 55–69.

IAB (2017). *IAB internet advertising revenue report 2016 full year results*. Tech. rep. IAB (Interactive Advertising Bureau), PriceWaterhouseCoopers.

Ilya, G (2010). "Forecasting click-through rates based on sponsored search advertiser bids and intermediate variable regression". In: *ACM Transactions on Internet Technology* 10.3.

Jansen, Bernard J., Danielle L. Booth, and Amanda Spink (2008). "Determining the informational, navigational, and transactional intent of Web queries". In: *Information Processing & Management* 44.3, pp. 1251–1266.

Jansen, Bernard J. and Tracy Mullen (2008). "Sponsored search: an overview of the concept, history, and technology". In: *International Journal of Electronic Business* 6.2, p. 114.

Jansen, Bernard J. et al. (2009). "The Components and Impact of Sponsored Search". In: *Computer* 42.5, pp. 98–101.

Jansen, Jim (2011). *Understanding sponsored search: Core elements of keyword advertising.* Cambridge University Press.

Jerath, Kinshuk, Liye Ma, and Young-Hoon Park (2014). "Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity". In: *Journal of Marketing Research* 51.4, pp. 480–486.

Jeziorski, Przemyslaw and Ilya Segal (2015). "What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising". In: *American Economic Journal: Microeconomics* 7.3, pp. 24–53.

Joachims, Thorsten et al. (2005). "Accurately interpreting clickthrough data as implicit feedback". In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05*. New York, New York, USA: ACM Press, p. 154.

Joachims, Thorsten et al. (2007). "Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search". In: *ACM Transactions on Information Systems* 25.2, 7–es.

Joo, Mingyu, Kenneth C. Wilbur, and Yi Zhu (2016). "Effects of TV advertising on keyword search". In: *International Journal of Research in Marketing* 33.3, pp. 508–523.

Kireyev, Pavel, Koen Pauwels, and Sunil Gupta (2016). "Do display ads influence search? Attribution and dynamics in online advertising". In: *International Journal of Research in Marketing* 33.3, pp. 475–490.

Kitts, Brendan and Benjamin Leblanc (2004). "Optimal Bidding on Keyword Auctions". In: *Electronic Markets* 14.3, pp. 186–201.

Kominers, Scott Duke (2009). "Dynamic Position Auctions with Consumer Search". In: *Algorithmic Aspects in Information and Management*, pp. 240–250.

Montgomery, Alan L. et al. (2004). "Modeling Online Browsing and Path Analysis Using Clickstream Data". In: *Marketing Science* 23.4, pp. 579–595.

Narayanan, Sridhar and Kirthi Kalyanam (2015). "Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach". In: *Marketing Science* 34.3, pp. 388–407.

Pass, Greg, Abdur Chowdhury, and Cayley Torgeson (2006). "A picture of search". In: *Proceedings of the 1st international conference on Scalable information systems - InfoScale '06*. New York, New York, USA: ACM Press, 1–es.

Qin, Tao, Wei Chen, and Tie-Yan Liu (2015). "Sponsored Search Auctions: Recent Advances and Future Directions". In: *ACM Trans. Intell. Syst. Technol.* 5.4, 60:1–60:34.

Regelson, Moira and Daniel C Fain (2006). "Predicting Click-Through Rate Using Keyword Clusters". In: *EC'06*. Ann Arbor, MI.

Richardson, Matthew, E Dominowska, and Robert Ragno (2007). "Predicting clicks: estimating the click-through rate for new ads". In: *Proceedings of the 16th ...* Pp. 521–529.

Rusmevichientong, Paat and David P. Williamson (2006). "An adaptive algorithm for selecting profitable keywords for search-based advertising services". In: *Proceedings of the 7th ACM conference on Electronic commerce - EC '06*. New York, New York, USA: ACM Press, pp. 260–269.

Rutz, Oliver J, Randolph E Bucklin, and Garrett P Sonnier (2012). "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising". In: *Journal of Marketing Research* 49.3, pp. 306–319.

Schlangenotto, Darius and Dennis Kundisch (2017). "Over-Paid Search: When Bricks-and-Mortar Retailers Should Not Use Paid Search". In: pp. 1267–1270.

Siebert, Sören (2017). *Google Analytics und Google Universal Analytics rechtssicher nutzen: Was Webseitenbetreiber jetzt tun müssen*.

Stigler, G. J. (1961). "The economics of information". In: *The journal of political economy*, pp. 213–225.

Stigler, G. J (1962). "Information in the labor market". In: *The journal of political economy*, pp. 94–105.

Strohmaier, Markus, Peter Prettenhofer, and Mark Kröll (2008). "Acquiring Explicit User Goals from Search Query Logs". In: *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. IEEE, pp. 602–605.

Szymanski, B.K. and J.S. Lee (2006). "Impact of ROI on bidding and revenue in sponsored search advertisement auctions". In: *Second Workshop on Sponsored Search Auctions*. Vol. 25. 6, pp. 1163–1178.

Vorobeychik, Yevgeniy and Daniel M. Reeves (2008). "Equilibrium analysis of dynamic bidding in sponsored search auctions". In: *International Journal of Electronic Business* 6.2, p. 172.

Wang, Dong, Zhenyu Li, and Mohamed-ali Kaafar (2016). "From Social Trends to Profit : An Optimisation Model for Search Engine Adwords Management". In:

Wang, Taifeng et al. (2013). "Psychological advertising: Exploring User Psychology for Click Prediction in Sponsored Search". In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*, p. 563.

Weber, Ingmar and Alejandro Jaimes (2011). "Who uses web search for what". In: *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*. New York, New York, USA: ACM Press, p. 15.

Weitzman, Martin L. (1979). "Optimal search for the best alternative." In: *Econometrica: Journal of the Econometric Society*, pp. 641–654.

Yang, Sha and Anindya Ghose (2010). "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" In: *Marketing Science* 29.4, pp. 602–623.

Yang, Shuai et al. (2016). "Brand engagement on social media: will firms' social media efforts influence search engine advertising effectiveness?" In: *Journal of Marketing Management* 32.5-6, pp. 526–557.

Yao, Song and Carl F. Mela (2007). "Sponsored Search Auctions: Research Opportunities in Marketing". In: *Foundations and Trends in Marketing* 3.2, pp. 75–126.

Ye, Shengqi, Goker Aydin, and Shanshan Hu (2014). "Sponsored Search Marketing: Dynamic Pricing and Advertising for an Online Retailer". In: *Management Science*.

Zhang, Yuyu et al. (2014). "Sequential click prediction for sponsored search with recurrent neural networks". In: *arXiv preprint*, pp. 1369–1375.

Zhong, Feimin et al. (2010). "Incorporating post-click behaviors into a click model". In: *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. New York, New York, USA: ACM Press, p. 355.

Zhou, Yunhong, Deeparnab Chakrabarty, and Rajan Lukose (2008). "Budget constrained bidding in keyword auctions and online knapsack problems". In: *Proceeding of the 17th international conference on World Wide Web - WWW '08*, p. 1243.

Zhu, Yi and Kenneth C. Wilbur (2011). "Hybrid Advertising Auctions". In: *Marketing Science* 30.2, pp. 249–273.

Zhu, Zeyuan Allen et al. (2010). "A novel click model and its applications to online advertising". In: *Proceedings of the third ACM international conference on Web search and data mining - WSDM '10*. New York, New York, USA: ACM Press, p. 321.

Chapter 2

Predicting conversion rates in paid search using sparse keyword-level data

Tobias Blask, Martin Stange, Christian Borck, Burkhardt Funk

Abstract

Background: Placing optimal bids in paid search auctions has a huge influence on the success of search advertising campaigns. Optimal bidding behavior requires advertisers to accurately estimate conversion probabilities for every keyword in their search campaign. In practice, this is easy for frequently searched keywords but a challenge in the long-tail where searches per keyword are sparse.

Aim: We propose a method that supports advertisers to estimate conversion probabilities per keyword early on, even in case there have been only a few searches so far.

Method: First, we use lasso regression to select relevant features. Second, we cluster keywords based on these features. Third, we estimate conversion rates for each keyword using a hierarchical Bayesian model based on these clusters.

Results: Applying the method, we demonstrate that conversion rates of long-tail keywords can be estimated with higher precision than it is possible by only considering clicks and conversions per keyword. We implement our method at a major European advertiser and show that the overall number of conversions increased significantly by placing bids in accordance to the predicted conversion rates.

This article has been published as follows: Tobias Blask, Martin Stange, Christian Borck, Burkhardt Funk (2017). "Predicting conversion rates in paid search using sparse keyword-level data". In: *Working Paper 2017*

2.1 Introduction

Paid Search has become one of the most important sources for the generation of leads and conversions in many industries (Yang, Lu, and Lu, 2014; Qin, Chen, and Liu, 2015).

One of the major challenges for advertisers in paid search advertising is determining the maximum price they are willing to pay for clicks on their keyword-specific ads. This maximum cost per click CPC_{max} depends on many different factors. A very important factor is the (predicted) performance of a keyword-specific ad in terms of conversion rates (CVRs) or return on investment. Search campaigns typically consist of only a few keywords that account for the majority of impressions, clicks and conversions. These keywords are referred to as "head" keywords. Using the data produced by these keywords to predict their performance in terms of CVRs, for instance, is rather simple. On the contrary, the majority of keywords belongs to the so-called long-tail. These keywords are more specific in their meaning than head keywords (e.g. "nike air max size 6" vs. "shoes"). For these keywords even predicting click-through rates (CTRs) is already a challenge (Regelson and Fain, 2006). Since conversions usually occur much less frequent than clicks, predicting their CVRs is even more challenging (Rutz, Bucklin, and Sonnier, 2012; Rutz and Bucklin, 2007). In this context, many approaches have been proposed to optimize the choice of keywords and to optimize maximum costs per clicks, i.e., to determine optimal bid sizes (Broder et al., 2011; Skiera and Nabout, 2013; Kitts and Leblanc, 2004).

Since estimating individual CVRs is central to judge long term profitability of keywords, advertisers often use rule based heuristics. An often used heuristics in practice works as follows: A keyword is bid for as long as the average cost per order (CPO) is below a certain threshold, say twice the average target CPO of a campaign. If no conversions are observed, the CPO can obviously not be calculated. In this case, the keyword is bid for until the money spent on this keyword is twice the target CPO. Having in mind that some advertisers have more than 1 million keywords in their search campaigns, employing this heuristic to long-tail keywords needs time and significant budgets. Figure 2.1 exemplifies how the uncertainty of CVR estimates decrease as the number of clicks increases.

In this paper, we take a different approach to predict CVRs for keywords in the long-tail that is based on the similarity of different keywords. In the literature different approaches for this kind of keyword clustering have been proposed, for example semantic clustering or hierarchical clustering (Rutz, Bucklin, and Sonnier, 2012). In

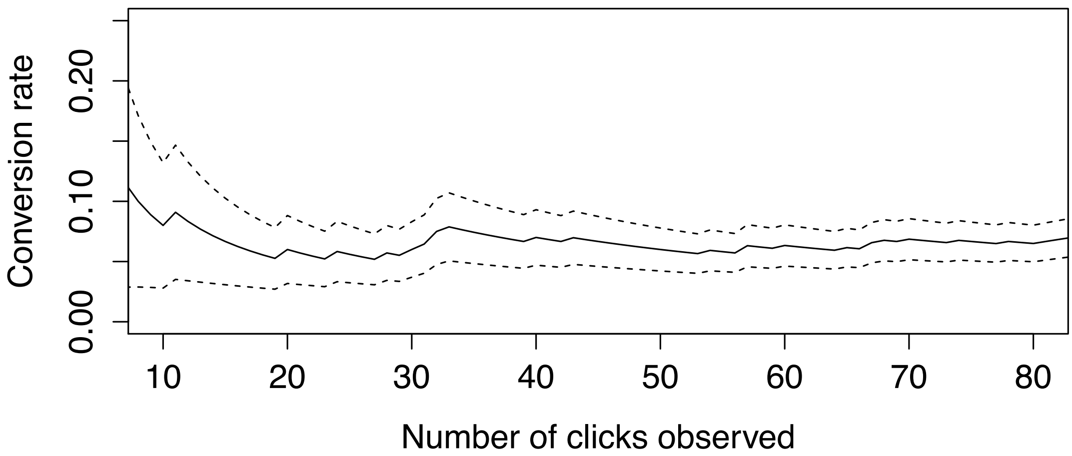

FIGURE 2.1: Conversion-rate for an individual keyword after a given number of accumulated clicks plotted with +/- 1 standard deviation (using binomial errors)

our work, we use the information on keyword level that is available to advertisers, for instance, the percentage of new visits of the advertiser's website or bounce rates. Here, bounce rates represent the ratio of users immediately leaving the advertiser's website after visiting the landing page. An advantage of this type of information is that it is often tracked in practice and that it can be determined accurately with only a few clicks for a given keyword. As our main contribution, we propose and evaluate an approach to predict keyword specific CVRs when only little information is available. As demonstrated in a case study these prediction can be used to support bid management in paid search.

In our approach, we first apply a lasso regression to determine relevant features for predicting CVRs. Second, we create a decision tree for clustering keywords based on the selected features. Third, we estimate CVRs by conducting a Bayesian analysis for each cluster. We update CVR estimates as new data about keywords is collected, by including the observed CVR as likelihood in the Bayesian model.

The remainder of this paper is structured as follows: First, we review recent literature dealing with predicting CTRs and CVRs when only little data is available. Second, we describe our modeling approach in detail an present results of each step of the analysis. Third, we apply the approach to a case study and show that it increases CVR prediction quality significantly and outperforms other methods and heuristics. Finally, we conclude our work and outline its implications.

2.2 Related Work

Recent work about search engine advertising often points out the importance of historical CTRs to allow for accurate predictions of future CTRs (Cetin et al., 2011; Hillard et al., 2011; Regelson and Fain, 2006). However, for many (particularly long-tail) keywords not sufficient historical information, and, naturally, for new keywords no information is available at all. Regelson and Fain (2006) propose to create clusters of similar keywords to overcome the problem of missing data. They argue that the more closly related two keywords are the more likely it is that their CTRs are similar. They conclude that building hierarchical clusters of keywords improves predictive accuracy in terms of CTRs, when only little historical data about some of the keywords is available. Xiong et al. (2012) use logging data provided by a commercial search engine to predict CTRs for certain search engine ads. In contrast to other studies, they use surrounding ads as independent variables and show that these significantly influence the CTRs of the ads. However, this information is not available for advertisers in general.

The aforementioned studies often focus on predicting the keywords' CTRs. Thereby, however, they ignore the fact that different keywords might also imply different CVRs (Rutz, Bucklin, and Sonnier, 2012). Rutz and Bucklin (2007) compare different models to predict conversion rates when only little information about individual keywords is available. They show that a model that considers heterogeneous CVRs across keywords outperforms models that assume a fixed CVR over all keywords in terms of predictive accuracy. Using the same data set, Rutz, Bucklin, and Sonnier (2012) develop a model to measure keyword performance that allows for keyword-specific correlations between CTRs, keyword semantics, keyword positions and CVRs. They find that higher positions typically lead to higher CVRs and that generic keywords typically have lower CVRs than more specific keywords. Ghose and Yang (2009) develop a model that involves keyword-level data such as CTR, CVR, and position as well as landing-page specific information. They find that the quality of a landing page may boost CTR and CVR. However, they admit that landing page quality is usually determined by proprietary algorithms employed by search engines, which might change over time. Factors that account for landing page quality involve the adherence to source code standards, average time on site or bounce rates. These measures are more transparent than a quality factor and, therefore, more suitable in a scientific context. They can be tracked using tools such as Google Analytics (Hu et al., 2009). In our approach, we include measures related to landing page quality such as average time on site, bounce rates, and the percentage of new visits to predict CVRs in search engine advertising campaigns. In contrast to the hierarchical models proposed in the literature (Rutz and Bucklin, 2007; Ghose and Yang, 2009), we use a pragmatic 3-step learning approach to predict CVRs.

The knowledge of a keyword's CVR can be used to calculate appropriate bids (Broder et al., 2011; Skiera, Gerstmeier, and Stepanchuk, 2008; Kitts and Leblanc, 2004) and to optimize the distribution of budget over time (Zhang et al., 2014). Proposed methods often focus on maximizing profits. For our case study, however, we follow a simple method to estimate the size of a bid which works as follows: Given an advertiser who defines maximum costs per acquisition of 50 EUR, and a CVR of 1%, the advertiser would be willing to pay 0.50 EUR for each click, which limits the size of a bid in a second price auction to 0.51 EUR.

2.3 Estimation of Conversion Rates

The goal of the analysis conducted here is to estimate conversion probabilities for individual keywords. Our modeling approach consists of three sequential steps: First, we apply a lasso regression to determine relevant keyword characteristics with respect to CVR. Second, we cluster all keywords using a decision tree that is based on selected features from the first step. This step requires a limited set of variables to result in meaningful clusters, which is why the feature selection performed in the first step of the analysis is important. Third, we estimate CVRs for each cluster from the second step using a hierarchical Bayesian model. A Bayesian approach is feasible because the model includes multiple variables that are non-normally distributed. The hierarchical structure of the model allows a given cluster of keywords to learn from other clusters.

2.3.1 Feature Selection

To determine which of the available information is relevant for predicting conversion probabilities, we conduct a lasso regression (Tibshirani, 1996) with cross validation using the R package glmnet (Friedman, Hastie, and Tibshirani, 2010).

This method works as follows. For each value of the regularization term λ, a ten fold cross validation is conducted and the mean squared error is recorded. The smaller the regularization term, the more variables are selected by the algorithm. Figure 2.2 shows the mean square error depending on the size of the regularization term λ, or

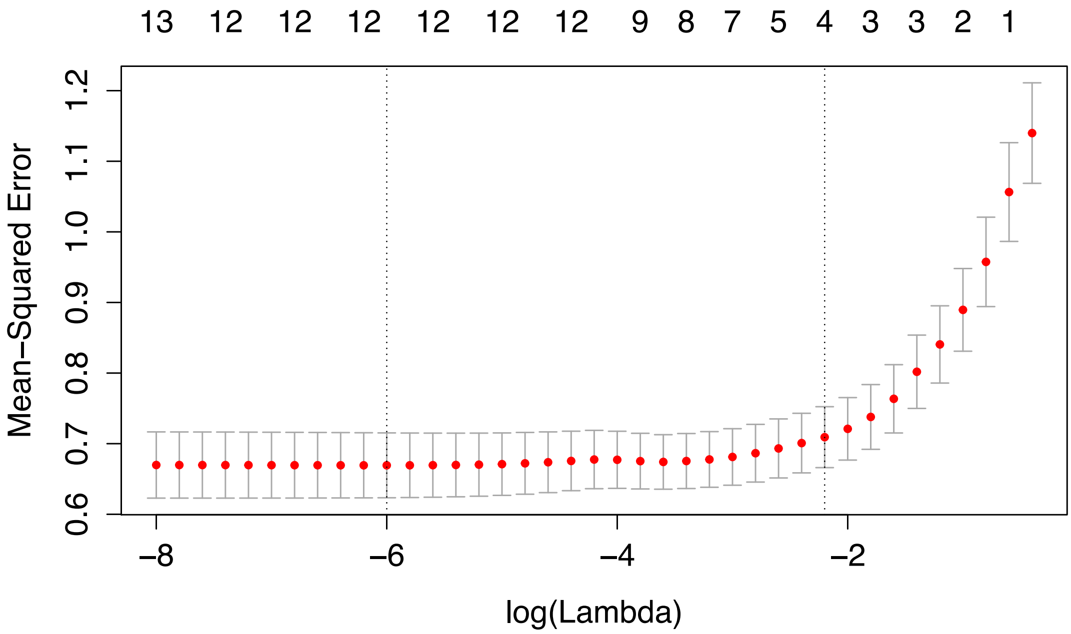

FIGURE 2.2: Feature selection based on cross validation with glmnet. The MSE decreases with the number of selected features.

in other words, the number of selected variables. For the next step of the analysis, we selected the four most relevant features that are determined by the regularization term resulting in a mean square error that is located within one standard deviation (right dashed in Figure 2.2) of the best fit , containing all but one available variables (left dashed in Figure 2.2).

Table 2.1 shows the results of the lasso regression. For our further analysis, we use the percentage of new visits, bounce rate, average position of the ad, and CTR.

2.3.2 Clustering of Keywords

Next, we use a decision tree to find clusters of similar keywords. We use the variables identified as relevant in the previous step.

The keywords are sorted into clusters using a decision tree algorithm that employs a recursive partitioning algorithm (Therneau, Atkinson, and Ripley, 2015). This step is conducted to build clusters of keywords. This effort is made to take the similarity of keywords into account when predicting conversion probabilities for every keyword.

Using the four selected features, the decision tree analysis results in five keyword clusters (Figure 2.3), for which we estimate the conversion rates in the next step

λ	0.55	0.37	0.25	0.11	0.07	0.06	0.04	0.02	0.02	0.01	0
MSE	1.06	0.89	0.8	0.71	0.69	0.69	0.68	0.68	0.68	0.68	0.67
Number of Impressions							✓	✓	✓	✓	✓
Average Time on Site					✓	✓	✓	✓	✓	✓	✓
Bounce Rate		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Percentage of New Visits	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Average Position			✓	✓	✓	✓	✓	✓	✓	✓	✓
Total Costs						✓	✓	✓	✓	✓	✓
Pages per Session									✓	✓	✓
Average CPM								✓	✓	✓	✓
First Page CPC											✓
Top of Page CPC									✓	✓	✓
External Quality Score						✓	✓	✓	✓	✓	✓
Click Through Rate				✓	✓	✓	✓	✓	✓	✓	✓
Average CPC										✓	✓

TABLE 2.1: Values of λ, related selected features and resulting MSE. For our analysis we use $\lambda = 0.11$ which results in only four features and in an MSE that is located in the confidence interval of the lowest MSE

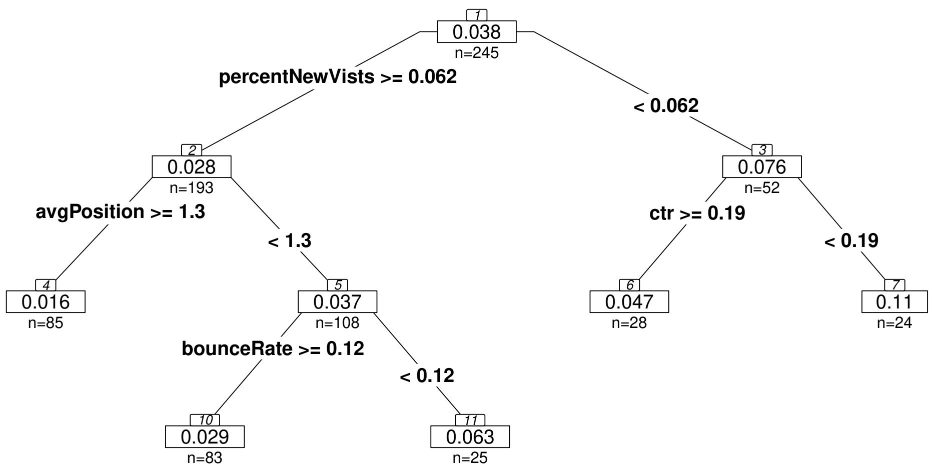

FIGURE 2.3: Clustering of keywords with more than one conversion in the whole observation period using recursive partitioning and the given model. The graph shows the cluster number, the mean CVR and the number of keywords per cluster

using a Bayesian model. We set a lower limit on the number of instances in each leaf and used a cutoff value of 0.005.

The tree already produces estimates of CVR's and one could stop here now. We decided to continue with adding additional methods here. That makes the predictions significantly better as can be seen in table 2.2.

2.3.3 Estimation of Conversion Rates

We estimate conversion rates of each keyword i in cluster c using the hierarchical Bayesian model presented in Figure 2.4 and Equation 2.1. We model the number of conversions of a keyword $CONV_i^c$ to be binomially distributed according to Binomial(CVR_i^c, CL_i^c), with CL_i^c representing the number of clicks of keyword i of cluster c. Since the true CVR of a keyword is unknown, we assume that the CVR_i^c from a cluster a distributed according to a Beta-distribution. The prior parameters of this Beta distribution (a^c and b^c) are sampled from Γ distributions that are parameterized by the hyperpriors $\alpha_{1,2} = 0.01$ and $\beta_{1,2} = 0.001$. We chose smaller values for $\beta_{1,2}$ because $a^c \leq b^c$, i.e. clicks occur more frequent than conversions. We are interested in the parameters a^c and b^c because they can later be used for predicting keyword-specific conversion rates.

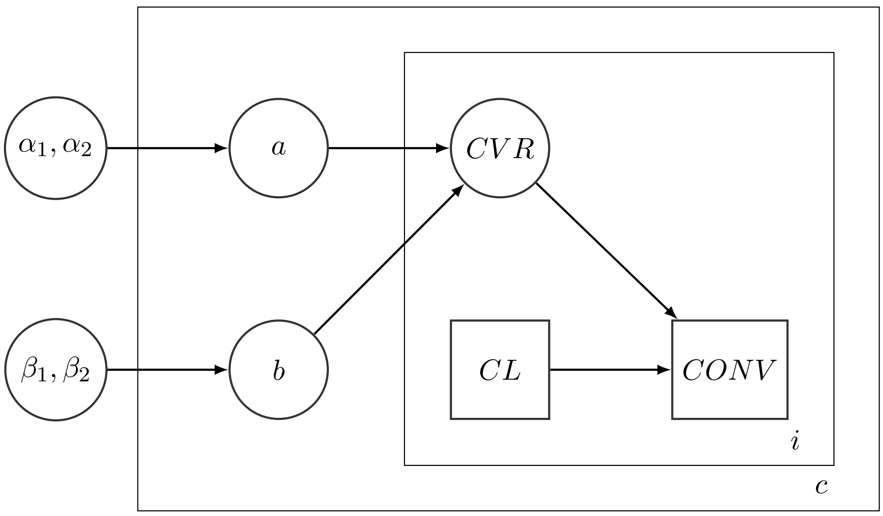

FIGURE 2.4: Graphical representation of the hierarchical model applied here.

We estimate the parameters a^c and b^c using the software package JAGS (Plummer, 2003) in accordance with Equation 2.1.

$$
\begin{aligned}
CONV_i^c &\sim \text{Binomial}(CVR_i^c, CL_i^c) \\
CVR_i^c &\sim \text{Beta}(a^c, b^c) \\
a^c &\sim \Gamma(\alpha_1, \alpha_2) \\
b^c &\sim \Gamma(\beta_1, \beta_2) \\
\alpha_1, \alpha_2 &\sim \Gamma(.01, .01) \\
\beta_1, \beta_2 &\sim \Gamma(.001, .001)
\end{aligned} \tag{2.1}
$$

We run 5000 burn-in steps and 5000 sampling steps. Figure 2.5 shows the correlation between a^c and b^c for each cluster. The higher the values of a^c and b^c, the greater is the influence of the prior in the calculation of the posterior value for CVR.

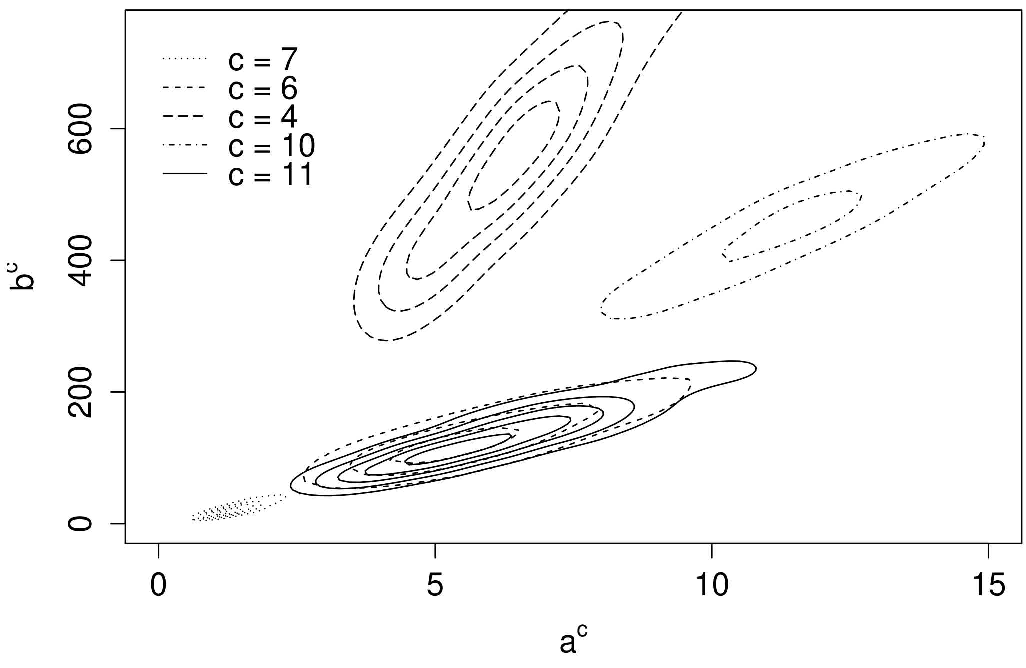

FIGURE 2.5: Contour plot of the correlation between a^c and b^c. The expected conversion rates ($a^c/(a^c+b^c)$) correspond to the clusters in Figure 2.3. Since a^c is smaller than b^c by at least one order of magnitude, the slopes of the functions $b^c = f(a^c)$ presented here can be interpreted as the expected conversion rates.

The posterior distribution of a^c and b^c can be used to calculate the CVR for each keyword i of each cluster c by evaluating the expected value E in accordance with Equation 2.2.

$$E(CVR_i^c) = \frac{CONV_i^c + a^c}{CONV_i^c + a^c + CL_i^c + b^c} \tag{2.2}$$

The expected value of CVR_i^c is based on prior knowledge of the cluster c and directly observed data (CL_i^c and $CONV_i^c$). In this approach, prior knowledge as well as likelihoods are updated when new observations (keyword-related or cluster-related) are made.

2.4 Model evaluation

To evaluate our approach, we use keyword-level data from an advertiser in the European market. The dataset has been collected using "Google Adwords" and "Google Analytics".

It covers one year of observations of search engine advertising campaigns with several thousand keywords and budgets in the millions resulting in tens of thousands of conversions. In addition, it contains on-page information. All data is available on a daily keyword level and on an aggregated level that meet the specific requirements of the different steps of the analysis. As reported by Abhishek, Hosanagar, and Fader (2011), analyses based on keyword level on a daily basis can be characterized by an aggregation bias because of systematic differences in click-through or conversion rates over the course of a day. However, if we would control for hourly variations, our data would become even more sparse.

We first compare the predictive performance of our model to other commonly used methods using the validation set. Next, we illustrate the practical implications and performance of the given approach in a case study in which we place bids for keywords in accordance with the predictions of our model. Thereby, we evaluate whether an increase of the advertiser's conversions can be observed that can be attributed to implementing the bidding agent.

For both purposes we use the same data set from a major European advertiser that has been collected during regular search engine advertising campaigns. The details of this advertiser cannot be revealed due to a non disclosure agreement. To fulfill the company's requirements, data is sanitized in this paper.

2.4.1 Model Benchmark

First, we use the observed conversion rates for each keyword in the training set as predictions for each keyword in the holdout set, and, thus do not make use of any statistical model. After that, we apply several conventional statistical methods to predict conversion rates in the holdout set: linear regression, recursive partitioning, support vector machine regression and random forest regression. We always use the identical variables (Percent New Visits, Click Through Rate, Average Position and Bounce Rate) in the model to predict CVRs with every method. To perform the test, we split the data into a training set and a holdout set. The training data consists of the first two months (61 days) while the validation data contains the rest of the dataset. We use the 245 keywords with at least one conversion (430,353 clicks resulting in 8,733 conversions in total) in the training period and predict the CVR of the 642 keywords (1,796,416 clicks resulting in 32,741 conversions in total) in the holdout set that meet the same restriction.

Finally, we calculate the root mean square errors of the respective predictions in the holdout set. The results can be found in table 2.2. In Figure 2.6 we illustrate how the different approaches under- and overestimate the true CVR in the holdout test to illustrate the accuracy of the predictions with our approach.

Prediction	HR	LR	RP	SVM	RF	OA
90% lower HDI	-0.033	-0.049	-0.064	-0.028	-0.041	-0.023
90% upper HDI	0.013	0.030	0.036	0.022	0.014	0.021
Root Mean Square Error	0.042	0.038	0.033	0.027	0.027	0.020

TABLE 2.2: Root mean square errors and 90% highest density intervals (HDI) of the distribution of the predictions on the holdout set based on different methods (HR = Heuristic, LR = Linear regression, PR = Recursive partitioning, SVM = Support vector machine regression, RF = Random forest regression, OA = Our approach))

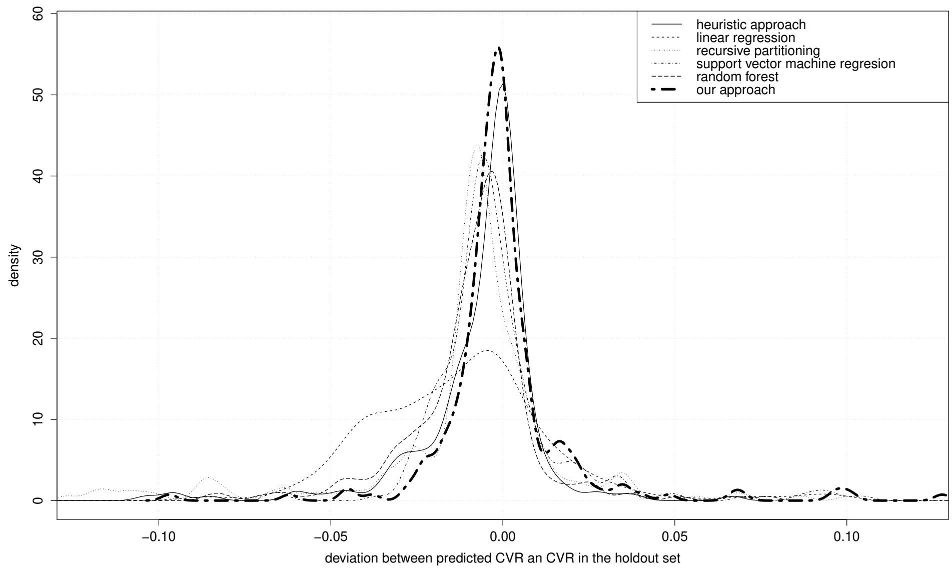

FIGURE 2.6: Density of deviations from the final conversion rate (CVR(holdout set) - CVR(predicted)) for the tested methods

2.4.2 Applying the Approach to a Case Study

We then apply our approach in practice by implementing it into a real world "Paid Search bidding agent". One additional step required to use the predictions in practice, is to calculate the bids for the keyword auctions. To test our approach, we calculate the actual price that we are willing to pay for a click on a given keyword by multiplying the predicted CVR with the desired maximum cost per acquisition. The bids for every keyword are then submitted to Google AdWords on a daily basis. The results illustrated from here on are based on the advertiser's Paid Search account that has already been described above. Figure 2.7 illustrates the 30-days periods before and after the bidding based on the predictions of our model was introduced to the advertiser's paid search account (bid management period). No other substantial changes were made to the campaigns during that time.

To measure the impact of the implementation of the described approach into the advertiser's bidding agent, which we will refer to as "intervention" from here on, we use the Bayesian structural time-series model proposed by Brodersen et al. (2015). This model infers the impact of an intervention by predicting the market response if this intervention would not have taken place. The causal impact of our intervention is defined as the difference between observed conversions and the number of conversions that would have been expected without intervention.

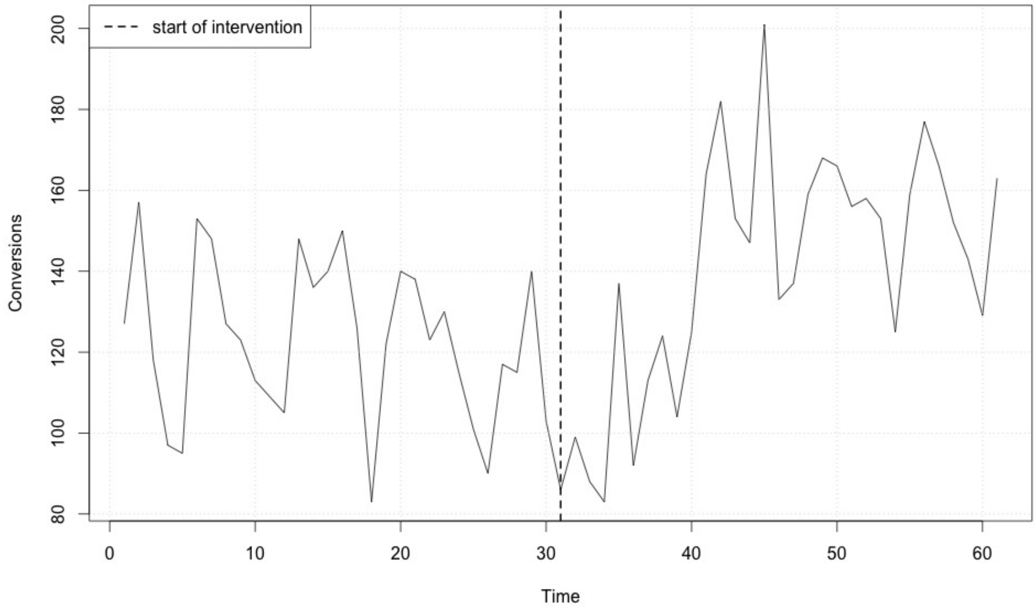

FIGURE 2.7: Development of conversions the advertiser's account before and after the tool was applied. The dashed vertical line represents the start of the intervention.

During the bid management period, we observe a total number of about 4.300 conversions. By contrast, if the implementation had not taken place, we would have expected a sum of 3.810 conversions. The 90% interval of this prediction is [3.560, 4.060]. So, the number of conversions showed an increase of +14%. The 90% interval of this percentage is [+7%, +21%].

These findings suggest that the positive effect observed during the bid management period is statistically significant and unlikely to be due to random fluctuations. As illustrated in Figure 2.8, the probability of obtaining this effect by chance is very small (tail-area probability p = 0.001). This indicates that the implementation of the algorithm into the company's bidding decisions has a significantly positive influence on their campaign performance.

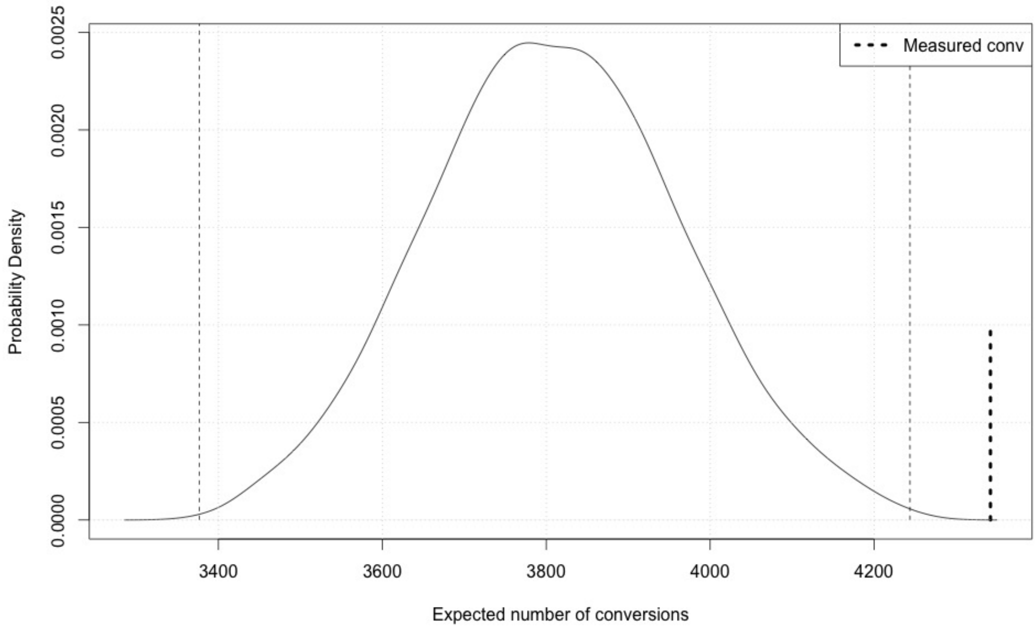

FIGURE 2.8: Posterior distribution of estimated number of conversions after intervention.

2.5 Conclusion

2.5.1 Contributions and Limitations

We contribute to research and practice in Paid Search by developing and evaluating a sequential modeling approach using feature selection, clustering and a hierarchical Bayesian model. By including information on keyword auctions and on aggregated user behavior, we develop a hierarchical Bayesian model that uses cluster-level data as follows: If only little specific information on keyword level is available, the model uses knowledge from the respective cluster to predict its conversion probability; as more and more data on keyword level is available, the prediction of its conversion rate is increasingly based on this specific data. In practice, the approach described here can be used to predict conversion probabilities for individual keywords, even if only sparse data is available. This is particularly important for the calculation of bids in Paid Search auctions.

Our research has primarily two limitations: First, the available conversion data was generated by the "Google AdWords conversion tracking". In the observed time period, conversions have been attributed by using a "last click" attribution model in which 100% of each conversion are attributed to the last Paid Search keyword clicked by a user. On the contrary, conversions of users who type-in the company's URL in the browser directly after having clicked on a Paid Search ad initially (Rutz, Trusov, and Bucklin, 2011) are not considered in our model. One important trend in marketing is related to using statistical models for the attribution of conversion success. Taking more realistic conversion data into account could improve results for advertisers when using the predictions of our approach for calculating their bids in Paid Search auctions. Furthermore, although the the decision tree analysis conducted here results in more than one hierarchical levels, we only use one hierarchical level in our Bayesian model. The extension of our model to allow for Bayesian belief propagation across hierarchical clusters as done by (Regelson and Fain, 2006) could possibly further improve predictive performance of our approach.

2.5.2 Outlook

Further extensions to improve the prediction quality of the given method could be the estimation of conversion probabilities when no conversion or even click data is available. This could, for example, be addressed by taking the properties of the ad copy into account. Kang et al. (2011) refer to these features as "language attractiveness, URL attractiveness, and query-snippet matching attractiveness". One could also use the knowledge concerning the influence of general, industry or advertiser specific features to predict and optimize their advertisements conversion probabilities. In addition, the hierarchical structure of our Bayesian model could be extended by considering keyword semantics. This might further improve the predictive performance of our model because of the spillover effect from more generic to specific keywords (Rutz and Bucklin, 2011).

Contributing to the presented topic has major implications for the success of advertisers' campaigns. With our case study, we show that the developed model is able to significantly increase conversion rates in real-life search engine advertising campaigns.

In summary, our approach shows how to come closer to accurate predictions in Paid Search advertising when only little information is available.

References

Abhishek, Vibhanshu, K Hosanagar, and P Fader (2011). “Aggregation bias in sponsored search data: The curse and the cure”. In: *SSRN 1490169*.

Broder, Andrei et al. (2011). “Bid generation for advanced match in sponsored search”. In: *Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11*.

Brodersen, Kay H. et al. (2015). “Inferring causal impact using Bayesian structural time-series models”. In: *The Annals of Applied Statistics* 9.1, pp. 247–274.

Cetin, Ozgur et al. (2011). “Missing Click History in Sponsored Search: A Generative Modeling Solution”. In: *17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1, pp. 1–22.

Ghose, Anindya and Sha Yang (2009). “An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets”. en. In: *Management Science* 55.10, pp. 1605–1622.

Hillard, Dustin et al. (2011). “The sum of its parts: reducing sparsity in click estimation with query segments”. In: *Information Retrieval* 14.3, pp. 315–336.

Hu, Jian et al. (2009). “Understanding user's query intent with wikipedia”. In: *Proceedings of the 18th international conference on World wide web*, pp. 471–480.

Kang, Changsung et al. (2011). “Modeling Perceived Relevance for Tail Queries without Click-Through Data”. In: *arXiv:1110.1112*.

Kitts, Brendan and Benjamin Leblanc (2004). “Optimal Bidding on Keyword Auctions”. In: *Electronic Markets* 14.3, pp. 186–201.

Plummer, Martyn (2003). “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling”. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna*.

Qin, Tao, Wei Chen, and Tie-Yan Liu (2015). “Sponsored Search Auctions: Recent Advances and Future Directions”. In: *Transactions on Intelligent Systems and Technology* 5.4, p. 60.

Regelson, Moira and D Fain (2006). “Predicting click-through rate using keyword clusters”. In: *Proceedings of the Second Workshop on Sponsored Search Auctions*.

Rutz, O. J., M. Trusov, and R. E. Bucklin (2011). “Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?” In: *Marketing Science* 30.4, pp. 646–665.

Rutz, Oliver J and Randolph E Bucklin (2007). "A Model of Individual Keyword Performance in Paid Search Advertising". In: *SSRN 1024765*.

– (2011). "From Generic to Branded: A Model of Spillover in Paid Search Advertising". In: *Journal of Marketing Research* 48.1, pp. 87–102.

Rutz, Oliver J., Randolph E. Bucklin, and Garrett P. Sonnier (2012). "A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising". In: *Journal of Marketing Research* 49.3, pp. 306–319.

Skiera, Bernd, Eva Gerstmeier, and Tetyana Stepanchuk (2008). "Stichwort Suchmaschinenmarketing". In: *Die Betriebswirtschaft : DBW* 68.1, pp. 113–117.

Skiera, Bernd and Nadia Abou Nabout (2013). "Practice Prize Paper PROSAD : A Bidding Decision Support System for Profit Optimizing Search Engine Advertising". In: *Marketing Science* 32.2, pp. 213–220.

Therneau, Terry, Beth Atkinson, and Brian Ripley (2015). "rpart: Recursive Partitioning and Regression Trees". In: *r-project.org, R package version 4.1-9*.

Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society: Series B* 58.1, pp. 267–288.

Xiong, Chenyan et al. (2012). "Relational click prediction for sponsored search". In: *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*. New York, New York, USA: ACM Press.

Yang, S.a, S.a Lu, and X.b Lu (2014). "Modeling competition and its impact on paid-search advertising". In: *Marketing Science* 33.1, pp. 134–153.

Zhang, Jie et al. (2014). "Dynamic dual adjustment of daily budgets and bids in sponsored search auctions". In: *Decision Support Systems* 57.1, pp. 105–114.

Chapter 3

To Bid or Not To Bid? Investigating Retail-Brand Keyword Performance in Sponsored Search Advertising.

Tobias Blask, Burkhardt Funk, Reinhard Schulte

Abstract

In Sponsored Search Advertising companies pay Search Engines for displaying their text advertisements in the context of selected queries on their Search Engine Results Pages (SERPs). The position of the ads is auctioned among all interested advertisers every time a query is executed by a user. The results are displayed separately from the organic results. Whether it is profitable for advertisers to pay for advertisements in the context of directly retail-brand related queries is a heated debate in practice between media agencies and budget responsible managers in companies. Anecdotal evidence implies that users who are searching for a specific retail-brand are conducting navigational searches and would end up on the brand owners website anyway, especially when the company is already placed prominently in the organic search results. The objective of the present research is to determine whether and under what circumstances it makes sense, in economic terms, for brand owners to pay for sponsored search ads for their own brand keywords. Using an exclusively available dataset from a major European internet-pharmacy we describe a non reactive method that is based on an A/B-test which enables us to investigate the economic

value of the described behaviour. We are able to prove that the advertiser benefits significantly from these additional Sponsored Search ads which enable the company to generate more visitors (> 10 %), resulting in higher sales volumes at relatively low advertising costs even when the com pany is already listed in first position in the organic part of the respective SERP.

This article has been published as follows: T. Blask, B. Funk, R. Schulte (2012) To Bid or Not To Bid? Investigating Retail-Brand Keyword Performance in Sponsored Search Advertising. In: *Communications in Computer and Information Science , vol. 314, p. 129-140*

3.1 Introduction

In the information society, Internet search engines play a key role. They serve the information needs of their users and are an important source for advertising companies in terms of customer acquisition and activation (Jansen and Mullen, 2008). Search engine companies like Google generate most of their revenue through sponsored search (Hallerman, 2008). At the interface of computer science, economics, business administration, and behavioral sciences, search engine marketing has been established as an interdisciplinary research topic and has seen a growing and diverse number of publications during the last years (Edelman, Ostrovsky, and Schwarz, 2007; Skiera, Gerstmeier, and Stepanchuk, 2008; Varian, 2007; Varian, 2009). Selected decision problems are examined from the perspective of three stakeholder groups (i) users, (ii) search engines and (iii) advertising companies (Yao and Mela, 2009). Beside the optimal bidding behavior in sponsored search auctions (Kitts and Leblanc, 2004), one of the key decision problems for advertisers is the selection of keywords appropriate for their campaigns (Abhishek and Hosanagar, 2007; Fuxman et al., 2008).

So far little research has been conducted on the use of brand names in sponsored search (Rosso and Jansen, 2010b). What is the subject of a heated debate in business practice is whether companies should bid on their own brand name or whether this only substitutes clicks from organic listings on the SERP. To answer this question, we apply a non-reactive experimental method and use it in a case study of an online pharmacy that is ranked first with its brand name in the organic search results in Google (Unrau, 2010).

The contribution of this paper is the development and application of a method for measuring the impact of bidding on brand names in a partially controlled experiment. From a theoretical point of view, we make a contribution to understanding keyword selection in blended search. We begin with a review of the literature on the competitive importance of brands in search engine marketing. On this basis we derive four hypotheses which we examine using the methods described in chapter 4. In chapter 5 we discuss outcomes and business implications of this paper and finally give an outlook in chapter 6.

3.2 Literature Review

There are two streams of research which are important for our work. The first studies bidding behavior of competitors in sponsored search. The second stream – blended search – analyses user preferences for organic and sponsored results as well as the interactions between them.

3.2.1 Brand Bidding and Piggybacking

Although brand terms bidding behavior is of great relevance in business practice, there have only been very few scientific publications on the topic. As a first step, a distinction has to be drawn between the bids on the own brand and those on other companies brands. Previous research on sponsored search brand keyword advertising by Rosso and Jansen (2010b), which was based on the global top 100 brands included in the well-known WPP BrandZ survey, reveals that 2/3 of the brand names examined were used by other firms while only 1/3 of the brand owners analyzed advertise in the context of their own brand names on SERPs. Bidding on other companies' brand names is referred to as piggybacking, for which three different types of motivation have been isolated: (i) competitive: piggybacking by an obvious, direct competitor; (ii) promotional: e.g. by a reseller; and (iii) orthogonal: e.g. by companies that offer complementary services and products for the brand owners' products. While retail, fast food and consumer goods brands are greatly affected by piggybacking, this practice is rarely observed in the field of luxury brands and technology (Rosso and Jansen (2010a)).

The Assimilation-Contrast Theory (ACT) (Sherif and Hovland, 1961) and the Mere Exposure Effect (Zajonc, 1968) are models that offer an explanation of the circumstances under which bids on one's own or third party brand names could be economically valuable. In sponsored search advertising the use of other companies' brand names seems to be advantageous when the perceived difference between the own and other brands is low from a user's point of view (ACT), while the value of bidding on own brand terms depends on the degree of the Exposure Effect, i.e. the display frequency that a brand needs in order to influence the purchasing decisions of users positively. Until now the empirical validations of these models for brand-bidding have been based on user surveys (Shin, 2010) and can therefore be subject to the problem of method bias. However, for the first time we are able to present results that are based on data that were collected in a non-reactive setup.

3.2.2 Blended Search

From the search engines' perspective, the question is about the extent to which the free presentation of results in the organic part of the SERP counteracts their own financial interests in sponsored search as they generate essential parts of their profits in this area (Xu, Chen, and Whinston, 2009). While a high perceived quality in the organic search results helps search engines to distinguish themselves from their competitors and to gain new customers, it is exactly this high quality in the organic results that may lead to cannibalization effects between organic and sponsored results (White, 2008).

From the users' point of view, the question has to be asked which preferences and intentions they have when making their choice whether to use organic or sponsored results. Depending on their personal experience of this particular advertising channel and their motivation to search, Gauzente (2009) shows that consumers do not only tolerate sponsored search as just one more channel for advertising on the Internet but do sometimes even consider these sponsored results more relevant than the organic ones. This is particularly true for transactional-intended queries, i.e. the so-called commercial- navigational search, in which the search engine is used instead of manually typing the URL into the browser's address bar. The same strong preference for sponsored results can also be found in the context of, for advertisers even more attractive, commercial- informational queries where users, although they have a strong intention to buy, are nevertheless still looking for the best matching result for their specific commercial interest (Ashkan et al., 2009).

Along with the multiplicity of intentions that individual users have when typing queries into search engines, there are significant variances of key performance indicators (KPI) that search engines and advertisers pay attention to. Ghose and Yang (2008c) compare organic and sponsored search results in respect to conversion rate, order value and profitability. In fact, the authors note that both conversion rate and order values are significantly higher through traffic that has been generated by sponsored search results than those generated by visitors that have clicked on organic results. It seems that the combination of relevance and the clearly separated presentation of organic and sponsored results as well as their explicit labeling are factors that lead to a greater credibility of the search engine and thus increases the willingness to click on the sponsored results, which are often not inferior to organic results (Brown, Jansen, and Resnick, 2007).

Studies on the interaction between these two types of results indicate that their simultaneous presence in both the organic and sponsored results leads to a higher overall click probability (Jansen and Spink, 2007). More specifically, a high similarity between the content in the respective snippets leads to a higher click probability in the context of informational queries while users who are searching with transactional intentions seem to be more likely to click on one of the results when the similarity is low (Danescu-Niculescu-Mizil et al., 2010). Yang and Ghose (2010) confirm this observation and point out that this effect is much more pronounced in the context of brand-keywords with only little competition (e.g. retail brands / names of online retailers) than it is in a highly competitive environment.

In conclusion, and in contrast to a widely held opinion in business practice it has to be noted that previous research indicates that the placement of advertisements on SERPs is useful for advertisers even where the company is already represented in the organic results for the respective keyword. For the special - and for e-commerce queries most interesting - case of commercially intended queries, these studies indicate that the simultaneous occurrence in both result lists increases the overall probability to be clicked. The verification of these findings to brand terms has however not been accomplished so far and is the key contribution of this paper.

3.3 Hypothesis

The following hypotheses are formulated with reference to the online search and buying process. We assume that, when a user searches for the brand name of a company, both organic as well as sponsored results are displayed. These results contain links to the brand owner's website as well as links to other companies' websites. The user has three options to choose from (as shown in figure1): he may click on one of the two links that lead to the website of the company or click on a link that takes him to a different website, which makes him leave the area of observation of the study.

Due to partial substitution effects, the following hypothesis is almost self-evident as the studied brand occupies the first result in the organic part of the SERP for queries that contain the brand name:

> **H1:** The number of visitors from organic search results decreases when brand owners engage in sponsored search for their own

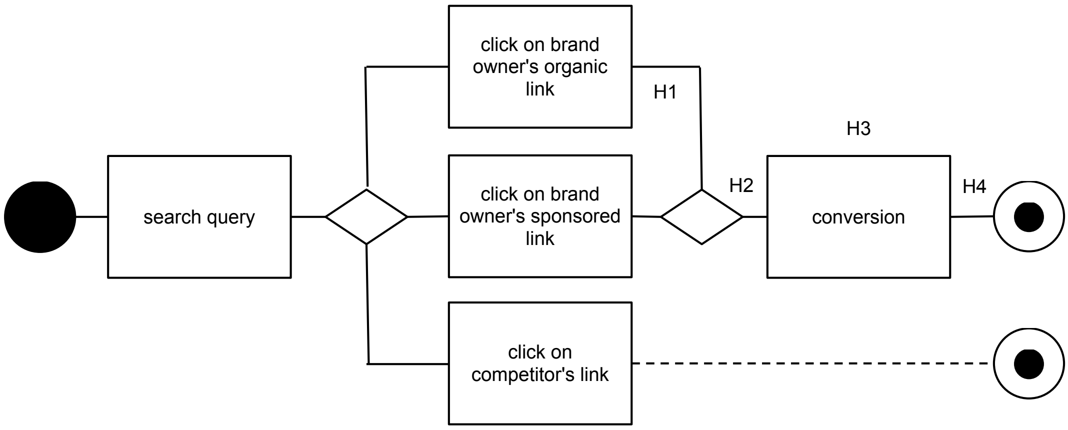

FIGURE 3.1: Hypotheses of this study in the search and buying process from a user's perspective

brand keywords.

In his paper Jansen and Spink (2007) assumes that the simultaneous appearance in the organic and the sponsored results has a positive impact on the overall click rate of the companies' advertisements. This leads to:

H2: The overall number of visitors through brand name queries from a search engine increases when companies engage in sponsored search for respective keywords.

It is important to point out again that this statement is by no means self-evident, since it would be possible that the sponsored clicks generated through a brand term advertisement would merely substitute organic clicks that would come for free when no sponsored search is employed. In business practice it is exactly this point that is the subject of an intense and controversial debate between advertisers, agencies, and search engines. In their study Ghose and Yang (2008b) point out that the conversion rate of commercial-navigationally intended queries is higher for sponsored than for organic results. Consequently, the following hypotheses can be derived:

H3: The conversion rate of keyword traffic from own brand keywords decreases when companies decide not to place sponsored

TABLE 3.1: Brand keyword clicks and revenues (with standard deviations) in the reference period (data are disguised to ensure confidentiality)

Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Ad Status	OFF	ON	OFF	ON	OFF	ON	OFF
Sum of all visiors	562.3	543.6	497.2	452.2	376	283	431.6
Standard Deviation	93.7	99.9	101.7	119.8	89.2	69.7	103.2
Revenue in EUR	8285	7119	6855	6162	4771	3843	7627
Standard Deviation	2117	1924	2022	1903	1630	1608	2537

search ads for these keywords.

Based on hypotheses H2 and H3 and other things being equal the following hypothesis on the number of sales and revenue derived from brand oriented search can be made:

> **H4:** The overall number of sales and the respective revenue increase when companies bid on their own brand names in sponsored search.

3.4 Case Study

The study covers a 14 day test period in which sponsored search for brand keywords is switched on and off on alternate days. Below, the respective states in the test period are called “ON” (sponsored search for brand keywords is employed) and “OFF" days. A full two weeks test period was chosen to allow us to monitor each weekday in both of the two possible states to ensure an acceptable consideration of the well-known weekday variations in e-commerce. The test period does not contain any holidays or other predictable events which could be relevant for the search engine traffic and conversions in this time span.

The company we study uses Google Analytics to collect data on the number and origin of users (organic as well as the sponsored results). In order to leverage existing data as a reference we decided to also use Google Analytics for our study. The reference period (Table 1) stretches from April 2009 to August 2010 with the omission of the test period which was chosen to be from April 12, 2010 till April 25, 2010, starting with an "OFF" day (Monday). The alternation of “OFF" and "ON" in the test period was executed manually each morning at eight o’clock.

Google Analytics assigns recognized re-visitors to the origin of their first visit. For example, a user who first reached the company’s website on an “ON" day via a sponsored search result would also be associated with this type of result for his future visits and will thus be assigned to the sponsored search visitors regardless of whether he arrives via an organic search result or by typing the address into browser manually. This is the main reason why there are sponsored search visitors on "OFF" days. To derive statements on the effect of self-bidding, the data from the test period is compared with a reference period that has no overlap with the test period and contains continuous self-bidding activities for the brand keyword. As will be argued in the next section, the main question about the data is whether the results are statistically significant. Using a Monte-Carlo-Simulation, we examine the validity of the observations especially with respect to hypotheses H2.

Even though the applied method does obviously influence the behavior of involved users and could therefore be categorized as ‘reactive’ in terms of social sciences, it shares common criteria with non-reactive methods since individual users have no knowledge of the investigation of his behavior.

3.5 Results

3.5.1 Testing the hypotheses

Hypothesis H1 predicts that the placement of sponsored search ads for the own brand name leads to a substitution of clicks that would have otherwise been generated without costs through clicks on organic results. This is clearly confirmed in the data. The magnitude and significance of this effect is clearly illustrated in figure 2. Comparing the composition of the sum of all clicks generated on "ON" days with the clicks on those days without self-bidding activities, we find more than double the number of organic clicks on "OFF" days (2392 clicks) than on "ON" days (1060 clicks).

TABLE 3.2: Brand keyword visits, conversion rates and revenues in the test period (data are disguised to ensure confidentiality)

Day	04/13	04/14	04/15	04/16	04/17	04/18	04/19	04/20	04/21	04/22	04/23	04/24	04/25
Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Ad status	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON
Paid visitors	376	56	340	64	184	44	436	92	340	108	292	68	252
Org. visitors	204	340	124	292	88	396	176	436	248	304	124	136	96
Sum visitors	580	396	464	356	272	440	612	528	588	412	416	204	348
Rev in EUR	5736	4704	6420	3328	3096	3720	8928	7796	6280	5832	4620	1112	7064
CVR	19%	22%	23%	19%	16%	12%	24%	23%	17%	17%	19%	10%	36%

It is, again, noticeable, and illustrated in figure 2, that we find sponsored clicks in the data that were generated on "OFF" days where we actually would not expect any. This can be explained by two effects: first, the status change was made manually from “ON” to “OFF” and vice versa every day at eight o’clock in the morning in the test-period so that sponsored search advertisements were served until eight o’clock in the morning even on “OFF” days, accounting for the minor part of these clicks. Second, as argued before the cookie based tracking contributes to the occurrence of sponsored clicks on “OFF” days. It is obvious that the existence of sponsored search clicks on “OFF” days could never generate or strengthen but would on the contrary weaken the findings that are presented in this paper, since they tend to blur a potential effect.

In summary, it is clear that these findings are consistent with the expectation of a substitution of organic by sponsored search results (H1).

The second hypothesis H2 deals with the question of whether the sum of all sponsored and organic clicks that are generated through the use of the brand name as keyword in search engines can be increased through the use of sponsored search advertising. For this, we compare data from the test period with the data of the reference period (figure 3).

Beginning with an "OFF" day, figure 3 shows the values that were generated on a daily basis in the test period as well as the weekday values of the reference period, both representing the sum of sponsored and organic traffic via the brand keyword from the Google SERPs. The observations of the test period mainly fall into the 50% percentiles of the reference period and thus follow the overall weekday cycle.

However, one can clearly recognize an overlaying pattern in the test period that is most likely driven by the alternation of the status of "OFF" and "ON". Overall, the expected pattern of more clicks on "ON" days than on the surrounding "OFF" days could be observed in 11 of 13 possible daily changes.

What is the likelihood that this pattern occurs by chance? To answer this question we conduct a Monte-Carlo-Simulation, in which 1,000,000 random 14-day samples were generated, each representing a random test period. To generate each 14-day time series, we use the Poisson distribution and take weekday means from the reference period as the mean of the distribution. What is remarkable is that a fraction of only 0.2% of the randomly generated test periods fit the observed (alternating) pattern with at least 11 or more changes. Employing this measure, it can be concluded with a probability of 99.8% that the placement of sponsored search advertisements for the

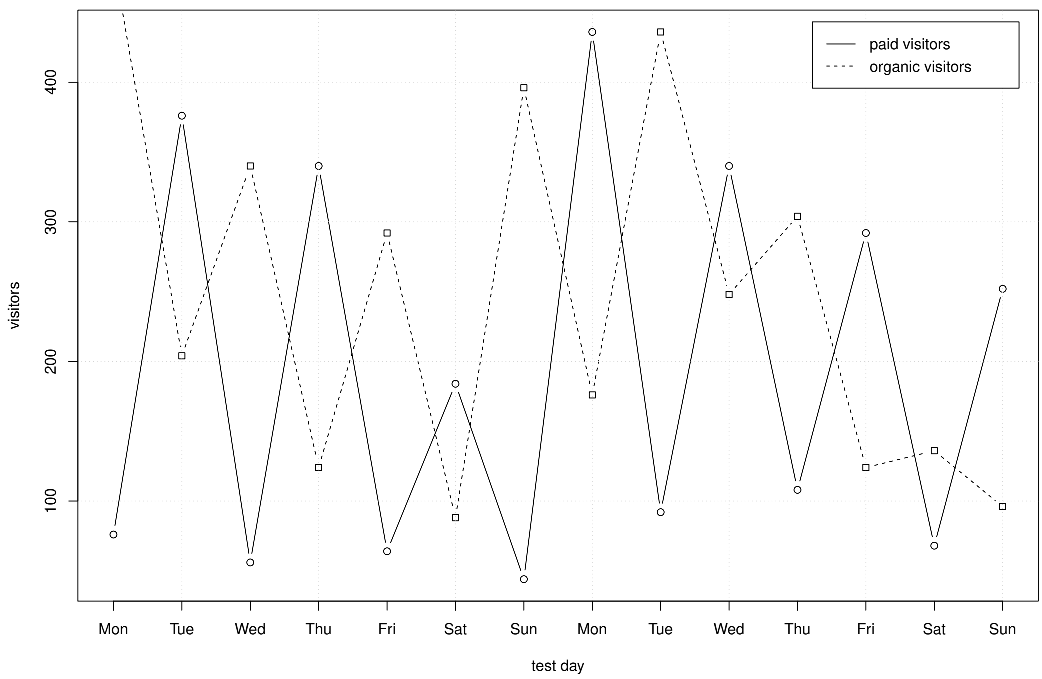

FIGURE 3.2: Organic (dashed line) vs. sponsored (solid line) clicks during the test period.

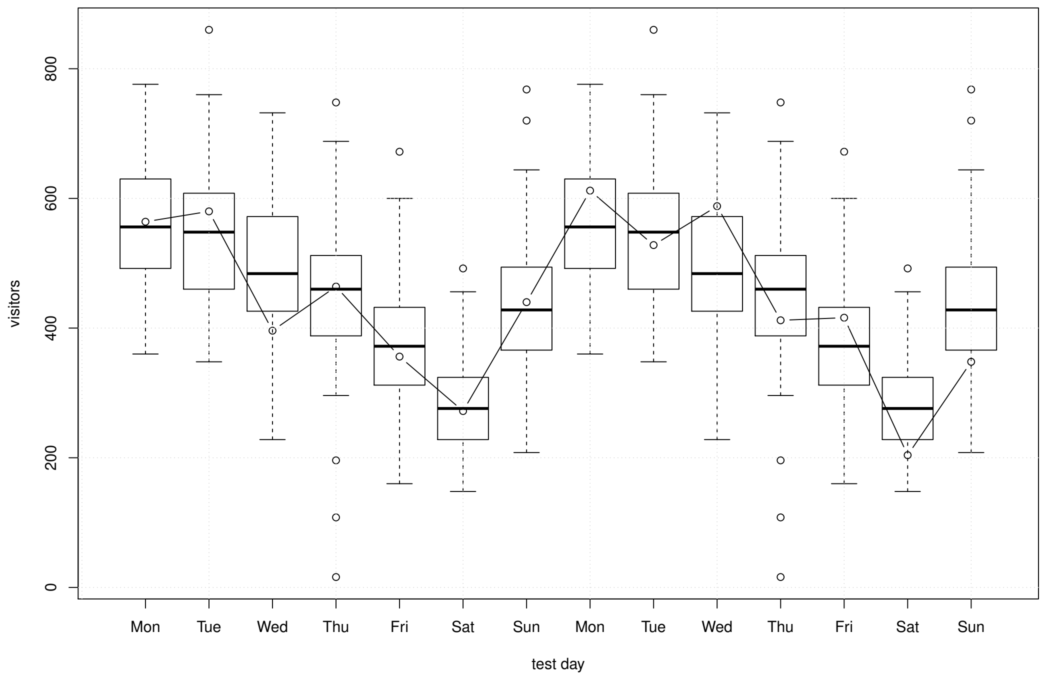

FIGURE 3.3: Daily sum of all clicks, generated through the search engine via the brand term in the test period (solid line) compared to the weekday values in the reference period. The boxes contain 50% of the values from the reference period.

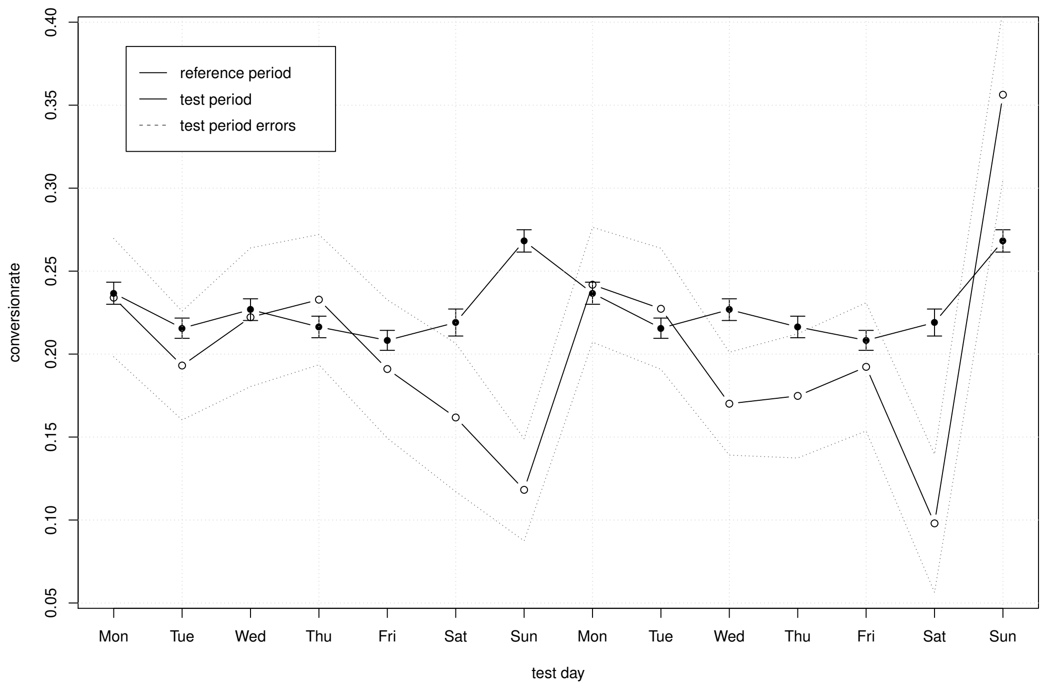

FIGURE 3.4: Conversion rates observed in the test period (solid red line) with standard errors vs. average conversion rates with standard errors on a given weekday (solid line) in the reference period.

own brand name actually leads to an increase in the total number of visitors for this keyword.

From the third hypotheses (H3), we would expect the conversion rate to be lower on days without sponsored search advertising than on the other days in the test. Given the average conversion rate of 22.7 % $\pm$ 0.3% in the reference period (figure 4) we find a lower conversion rate for the test period of 20.1% $\pm$ 1.6%, consistent with the study of Ghose and Yang (2008a) , who observed a lower conversion rate for traffic from organic listings. It should be mentioned, that due to the low number of transactions per day (and the corresponding statistical error) we cannot observe a consistent difference of the conversion rate between "ON" and "OFF" days as for the overall clicks (figure 3).

Following the proven hypothesis H2 (more visitors through sponsored search advertising for the brand name) and the lower conversion rate observed in the context of hypothesis H3, we expect less sales and reduced revenues in the test period. In fact, the revenue via the brand keyword in the test period (EUR 77,200) is lower than

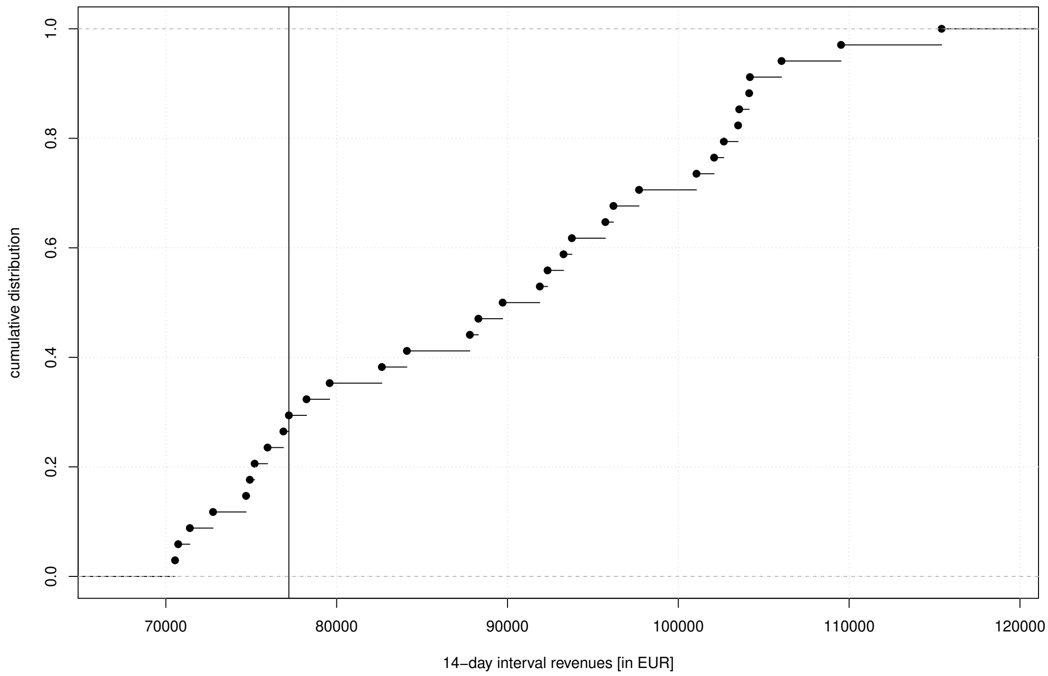

FIGURE 3.5: Empirical cumulative distribution of the revenues in the observation period (14-day intervals, containing the reference- as well as the test period), the test period is indicated by the vertical line.

70% of all comparable 14-day intervals in the reference period (figure 5).

Considering the revenue trend over the reference period, the relatively low revenue in the test period becomes significant since the revenue in the reference period shows a rising trend as shown in figure 6 (two-week revenue mean after New Year's Eve without the test period: EUR 99,130 with a standard deviation of ± EUR 6,107.89). A similar reduction of sales can only be observed in the two-week period around Christmas and New Year's Eve 2009 corresponding to observation point 19 in figure 6. Thus, we interpret the lower revenue as a consequence of not employing sponsored search for brand keywords.

3.5.2 Economic Impact

We now estimate the economic value of sponsored search for own brand names. During the test period each weekday was observed in both states, "ON" and "OFF". The number of additional visitors can be estimated by the sum of all clicks on "ON"

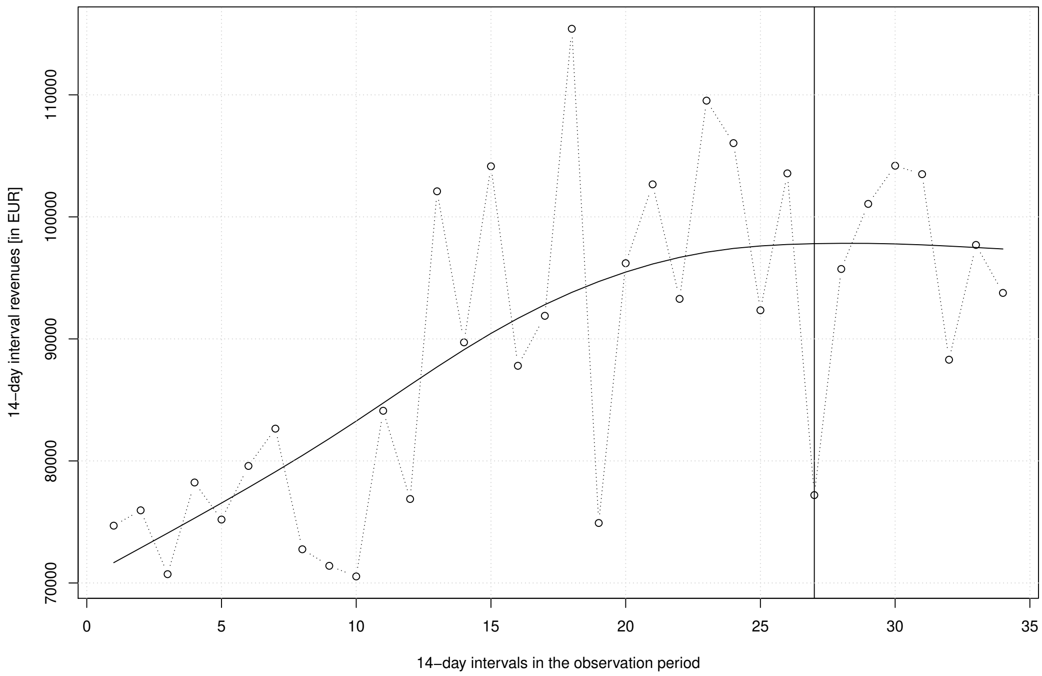

FIGURE 3.6: Time series of revenues (14-day intervals) during the reference period (dashed line), including the test period (observation point 27, indicated by the vertical line) and a trend line (solid line).

days minus the sum of all clicks on "OFF" day in the test period equal the total number of additional visitors for one week. In the current study, this results in 380 additional visitors per week. This is a significant growth of more than 10% achievable through sponsored search for own brand keywords.

Given the average conversion rate of 22.7% (reference period) and an average value per transaction of EUR60.88 this leads to an increase in sales of about EUR 275,000 per year. The average cost per click for the brand keyword in the test period was EUR 0.03, leading to additional costs of about EUR 600 per year. To sum up: Even if there were only very moderate margins for online pharmacies we would recommend the use of sponsored search advertising for brand keywords.

In general, it seems to be likely that sponsored search for own brands lead to more visitors and accordingly to more sales and higher revenues for the brand owner. The low prices per click for brand keywords and a higher conversion rate make brand name advertising economically profitable in the context of sponsored search.

3.6 Conclusions and Outlook

It is plausible to argue that users who search for a specific retail brand name in a search engine have already decided where their search is going to end (the website of the retailer). Yet, evidence from this study suggests that this is not the case for all users. Some users apparently find other advertisements or organic results on the SERP more interesting so that they can get lost for the brand owner if he is not present in the sponsored search results.

We expect that the extent to which the described effect occurs in practice for other companies depends on a number of factors. E.g., the intensity of competition – defined by the number of competitors who are also bidding on the brand name – is likely to have an influence on the observed effect. This is of special interest, because since September 2010 (in the European Union) companies can not ban other advertisers to bid for their brand keywords (Bechtold, 2011) which will lead to a more intense competition. In the light of this change the present research gains in importance for a whole range of advertisers. Other factors may be the price level of sponsored search clicks, the reputation and brand value of the advertiser and product characteristics.

Considerably more research is needed to determine the extent to which these factors have an impact on the described effect. Besides that, the authors currently work on a project that will help to understand user behavior in this context.

References

Abhishek, V. and K. Hosanagar (2007). "Keyword generation for search engine advertising using semantic similarity between terms". In: *Proceedings of the ninth international conference on Electronic commerce*. x. ACM, pp. 89–94.

Ashkan, A. et al. (2009). "Classifying and Characterizing Query Intent". In: *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, pp. 578–586.

Bechtold, S. (2011). "Google AdWords and European trademark law". In: *Communications of the ACM* 54.1, pp. 30–32.

Brown, A, B Jansen, and M Resnick (2007). "Factors relating to the decision to click on a sponsored link". In: *Decision Support Systems* 44.1, pp. 46–59.

Danescu-Niculescu-Mizil, Cristian et al. (2010). "Competing for users' attention". In: *Proceedings of the 19th international conference on World wide web - WWW '10*. New York, New York, USA: ACM Press, pp. 291–300.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz (2007). "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords". In: *American Economic Review* 97.1, pp. 242–259.

Fuxman, Ariel et al. (2008). "Using the wisdom of the crowds for keyword generation". In: *Proceeding of the 17th international conference on World Wide Web - WWW '08*. New York, New York, USA: ACM Press, pp. 61 –70.

Gauzente, C. (2009). "Information search and paid results, proposition and test of a hierarchy-of-effect model". In: *Electronic Markets* 19.2, pp. 163–177.

Ghose, A. and S. Yang (2008a). "An empirical analysis of sponsored search performance in search engine advertising". In: *Proceedings of the international conference on Web search and web data mining*. ACM New York, NY, USA, pp. 241–250.

– (2008b). "Modeling cross-category purchases in sponsored search advertising". In: *papers.ssrn.com*, pp. 1–35.

Ghose, Anindya and Sha Yang (2008c). "Comparing performance metrics in organic search with sponsored search advertising". In: *Proceedings of the 2nd International Workshop on Data Mining and Audience Intelligence for Advertising - ADKDD '08*. September 2007. New York, New York, USA: ACM Press, pp. 18–26.

Hallerman, D. (2008). *Search Engine Marketing: User and Spending Trends*.

Jansen, B.J. and T. Mullen (2008). "Sponsored search: an overview of the concept, history, and technology". In: *International Journal of Electronic Business* 6.2, pp. 114–131.

Jansen, B.J. and A. Spink (2007). "Sponsored search: is money a motivator for providing relevant results?" In: *Computer* 40, pp. 52–57.

Kitts, B. and B. Leblanc (2004). "Optimal bidding on keyword auctions". In: *Electronic Markets* 14.3, pp. 186–201.

Rosso, MA and BJ Jansen (2010a). "Brand Names as Keywords in Sponsored Search Advertising". In: *Communications of the Association for Information Systems* 27.1, p. 6.

– (2010b). "Smart marketing or bait & switch: competitors' brands as keywords in online advertising". In: *Proceedings of the 4th workshop on Information credibility*. ACM, pp. 27–34.

Sherif, M. and CI. Hovland (1961). "Social judgment: assimilation and contrast effects in communication and attitude change". In: *Yale University Press*.

Shin, W. (2010). "The Company that You Keep: When to Buy a Competitors Keyword". PhD thesis. The Wharton School, University of Pennsylvania.

Skiera, Gerstmeier, and Stepanchuk (2008). "Stichwort Suchmaschinenmarketing". In: *DBW Die Betriebswirtschaft* 68, pp. 113–117.

Unrau, E. (2010). "Wechselwirkungen zwischen bezahlter Suchmaschinenwerbung und dem organischen Index". In: *Master Thesis, Leuphana University, unpublished*.

Varian, H (2007). "Position auctions". In: *International Journal of Industrial Organization* 25.6, pp. 1163–1178.

Varian, H.R. (2009). "Online ad auctions". In: *American Economic Review* 99.2, pp. 430–434.

White, A. (2008). "Search Engines: Left Side Quality versus Right Side Profits". Toulouse.

Xu, Lizhen, Jianqing Chen, and Andrew Whinston (2009). "Too Organic for Organic Listing? Interplay between Organic and Sponsored Listing in Search Advertising". Austin, Texas.

Yang, S. and A. Ghose (2010). "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" In: *Marketing Science* 30.1, pp. 1–22.

Yao, S. and C.F. Mela (2009). "A Dynamic Model of Sponsored Search Advertising Preliminary and Incomplete". In: *faculty.insead.edu*.

Zajonc, R.B. (1968). "Attitudinal effects of mere exposure". In: *Pers. Soc. Psycho* 9.2. Part 2, pp. 1–27.

Chapter 4

Bayesian Parameter Estimation in Green Business Process Management: A Case Study in Online-Advertising

Tobias Blask

Abstract

Companies take their responsibilities for a sustainable planet more and more seriously. For online-retail businesses a significant share of all CO2 emissions is generated by delivering goods to their clients. Now various companies are implementing a greener logistic chain into their business processes. What is a central question for these performance driven companies in this context is whether it pays to invest in additional costs for carbon neutral delivery and if the customers appreciate these steps and prefer retailers that behave in this manner. We develop and perform a non reactive A/B-test that enables us to evaluate the influence of sustainability information on the customers decision to buy a product by clicking on an ad on a search engine results page (SERP). We analyze campaign performance data generated from a European e-commerce retailer, apply a Bayesian parameter estimation to compare the two groups, and demonstrate the advantages of the given Bayesian approach in comparison to the application of Null Hypothesis Significance Testing ($NHST$).

This article has been published as follows: Tobias Blask (2013). "Bayesian Parameter Estimation in Green Business Process Management: A Case Study in Online-Advertising". In: *Informatik 2013, 43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt, 16.-20. September 2013, Koblenz.* Additional references to Kruschke (2012) have been added on pp. 80, 81 and 85.

4.1 Introduction

Internet search engines like Google, Yahoo! or Bing play an undisputed key role in the modern information society. On the one hand they serve the information needs of their users, on the other hand they represent an important source of customer acquisition for companies in a broad variety of industries and sizes (Jansen and Mullen, 2008; Alby and Funk, 2011). They also provide the search engine companies with significant amounts of their revenues through Sponsored Search Advertising. While still growing rapidly Sponsored Search Advertising already dominates the online media-spendings of companies that advertise on the internet. In this form of advertising, developed in 1998 by Overture, advertisers provide search engines with text-advertisements and a list of keywords, which can consist of one or more terms, they want to be displayed. The advertiser usually also provides attributes to each of these keywords, but at the very least the amount of money he is willing to pay for a click on an ad for this specific keyword (CPC_{max}) (Jansen et al., 2009). Every time a user types in a query the search engine generates individual personalized result pages, depending on the users' location, his search history and other factors. If ads are available that could probably satisfy the need of the user, the search engine displays these ads alongside the organic results. If more than one advertiser is willing to pay for the display of an ad the search engine auctions the position of these ads among all interested players typically based on a Generalized Second Price Auction (GSP) (Jansen, 2011; Varian, 2009). In each auction only the advertiser that wins the auction by getting a click on an ad is charged by the search engine. The effective Cost-Per-Click (CPC_{eff}) is basically the maximum bid of the advertiser with the subsequent highest bid plus a small additional fee. In practice search engine companies use a more robust mechanism to maximize their profits by rewarding keyword/ad combinations that have a high relevance to users (often referred to as the quality score). Although detailed calculations are not disclosed, the key metric is claimed to be the historic Click-Through-Rate (CTR) where available, otherwise an expected click probability for the specific advertiser-ad-keyword combination is used.

In the present paper we concentrate on the advertisers' perspective and the direct impact of green signals in text advertisements. We evaluate the probability that a user will click on a given Sponsored Search text advertisement containing the promise of Carbon Neutral delivery vs. another one offering generic information on reliable fast delivery and conduct a Bayesian parameter estimation approach to analyze the data. In this context we illustrate the advantages one can get by applying a

FIGURE 4.1: Two variations of an ad, similar to the ones that were used in the A/B test: Carbon Neutral delivery vs. Fast and Reliable delivery

Bayesian Parameter Estimation analysis compared to conventional Null Hypothesis Significance Testing.

4.2 Literature Review

There are two streams in literature which are important for the present research. The first is green marketing. The second studies the various influences on Sponsored Search advertising effectiveness.

4.2.1 Green Marketing

Green marketing has been a widely recognized trend for international firms over the last years. One can clearly identify strong efforts in the development of sustainable brand images in a number of branches. One trend Leonidou et al identify in their review of developments in green advertising research and practice from 1988 to 2007 is a strategy shift from communicating environmental aspects within the production process to the communication of sustainable consumption by the customers themselves. An other important expansion of this field is observed in the intensification of efforts by B2C businesses in communicating green messages in their advertising activities. (Leonidou et al., 2011) The use of ecolabels is a well known tactic is to provide the potential consumer with independent confirmation of the green efforts of the respective advertiser. In fact Rex and Baumann state that there is still lack of empirical knowledge about the consumers reception in this area. (Rex and Baumann, 2007) Recent studies indicate that a number of consumers may be willing to pay higher prices for products they identify as environmental friendly. (Haytko and Matulich, 2008) What is still unanswered is the whether these green signals still have an impact direct buying decisions in situations in E-Commerce situations. A first approach to answer this question with a significantly smaller fraction of the given dataset has been made (Blask, 2013). However, more data is needed to ensure reliable results from the given study setup.

4.2.2 Sponsored Search Advertising

In published research, Online Marketing and Sponsored Search especially has become an established topic with a variety of high quality publications in Computer- and Information Science as well as in the fields of Operations Research and Marketing. Since 2004, Sponsored Search has become a continuously more and more important topic in the Online Marketing research area (Evans, 2008; Evans, 2009; Jansen and Mullen, 2008).

The search engine auctions the positions of the ads on the Search Engine Result Page (SERP) between all advertisers that placed a bid (CPC_{max}) for the given keyword. The ad position is the result of the combined CPC_{max} and so called quality scores of the players. The amount of money a specific advertiser has to pay for the click (CPC_{eff}) depends on the advertisers bid and the ones provided by the other advertisers in the auction as well as the quality score of the ad / query combination.

Many publications in this area have an empirical basis. Basically quantitative research is conducted with three types of datasets: **(a)** Search engine query data **(b)** aggregated media and e-Commerce statistics and **(c)** individual user journeys. Search engine query data is the rarest form of available data for researchers who are not directly affiliated to the search engines as it can only be collected by the search engine companies themselves. Although every search engine company generates masses of this type of data, there are only few datasets available for academic use. One of those is the well known AOL dataset. It consists of about 20 million completely non-censored web queries collected from about 650,000 users over a three month period, arranged by anonymous individual IDs. This dataset has been extensively examined since 2006 (Pass, Chowdhury, and Torgeson, 2006; Adar, 2007; Strohmaier et al., 2007; Strohmaier, Prettenhofer, and Kröll, 2008; Brenes and Gayo-Avello, 2009).

Aggregated media and e-Commerce statistics are generated by the advertisers themselves during their ad campaigns. One way this kind of data is produced is by the campaigning tool itself (e.g. Google AdWords) or the advertiser's respective software solution. The data is usually aggregated on campaign, adgroup and keyword-level and contains variables like the total number of clicks, impressions, CTR, and conversionrate (CVR) as can be seen in table 4.1.

TABLE 4.1: typical dataset from Google AdWords (ad level)

Ad	Clicks	Impr.	Avg.CPC	Cost	Avg.Position	Conversions
ad 1	132	2,198	1.32	174.08	2	16
ad 2	421	2,893	2.32	976.72	3	21
...	...	...	...	...	...	...

The third sort of available data enables researchers to understand individual user behavior. User journey conversion datasets include information about all measured touch-points that an individual user has with a specific advertiser. These datasets make the development of attribution-models possible where every conversion success can be allocated to the ad-contacts a user has had. Like the other types of data too, user journey data is always subject to several types of bias, such as caused by media discontinuities.

4.2.3 Click probability

Click probabilities have been widely studied since the early beginning of the advertising format Sponsored Search. However, due the lack of possibilities to observe the user behaviour while using a search engines, a complete coverage of all factors influencing the CTR is no easy task.

Evidence suggests that one of the most influencing factors is the ad position within the Sponsored Search results, which depends among other facts on the advertisers CPC_{max} and the so called quality score. The quality score, used by search engines to determine the quality of an advertisement, is based primarily on the historical CTR. A large number of studies has shown the correlation between decreasing position and a decreasing CTR and vice versa (Richardson, Dominowska, and Ragno, 2007; Agarwal, Hosanagar, and Smith, 2011). It should be emphasized, that the highest positions leads to high $CTRs$ but not mandatorily to the highest conversion rates. From an advertiser's perspective, a topic of interest is to predict the future CTR of sponsored ads. As argued before, the position has a major influence on the CTR, called the position bias. In the course of research, several models have been developed to explain the influence of the position bias on the CTR.

Crasswell, Zoeter and Taylor (Craswell et al., 2008) present several models for predicting the CTR: (a) baseline model, (b) mixture model, (c) examination model, and (d) cascade model . The findings were originally based on organic search results but, they are applicable to Sponsored Search results as well (Agarwal, Hosanagar, and Smith, 2011). The underlying assumption of the (a) baseline model is that a user screens every search result and decides afterwards, which one fits the best to the query. As a consequence, the click probabilities for each individual search result are identically, independently of its position. The (b) mixture model extends the baseline model and divides user behavior into two groups. One group behaves as described in the baseline model, the other group clicks randomly on one of the first search results. The (c) examination model refers to findings from eye tracking studies which state that with declining position, the probability of a click declines as well (Joachims et al., 2005; Joachims et al., 2007). The (d) cascade model is, owing to the high degree of explanation by click data, one of the most applied explanation approaches. The basic assumption is that the user scans each search result, beginning from the top to the bottom, comparing the relevance of each ad with the relevance of the ad before. The user continuous scanning the results until the perceived ad relevance reaches a certain level and the user clicks.

As mentioned above one challenge is to predict the CTR of keywords or keyword combinations for potential future Sponsored Search ads. One solution that has been proposed is aggregating historical data from similar keywords. Here, the CTR is represented as a function of position, independent of a bid. In doing so, the developed models do not focus on a certain advertiser. The same clustering approach can be applied in optimizing the search engines' profit (Dave and Varma, 2010). There are also models taking the quality score into account (Gluhovsky, 2010). A model developed by Zhu et al. (Zhu et al., 2010) called General Click Model focuses on the CTR prediction of long-tail queries, based on a Bayesian network. Dealing with the position bias mentioned before, Zhong et al. (Zhong et al., 2010) incorporate post-click user behaviour data from the respective landing page of the clicked ad into the click model to refine the estimation of the perceived user relevance after clicking on a specific ad. A similar approach, using Dynamic Bayesian networks can be found in Chappelle and Zhang (Chapelle and Zhang, 2009). Several models based on historical click data suffer from limitations in terms of lacking consideration of a possible user learning effect. Taking Gauzente's results as an example, it has been shown that past user satisfaction with Sponsored Search results influences the current click behaviour (Gauzente, 2009). Besides the incorporation of position data and the perceived relevance of presented ads, the CTR of an advertiser is also affected by the relationship between organic and Sponsored Search results. Listing the results of one company at the same time in sponsored and organic search results leads to a higher CTR and vice versa (Yang and Ghose, 2010; Blask, Funk, and Schulte, 2011).

4.3 Case Study

This study covers a test period over several weeks in which a single element in selected Sponsored Search text advertisements hast been alternated for a number of queries that users type into the Google search engine to eventually buy products in the advertiser's online shop as can be seen in fig. 5.1. The advertisers' products can be classified as B2C Fast Moving Consumer Goods. The selected keywords include **(a)** variations of the retailer brand, **(b)** the brand names of product manufacturers as well as **(c)** several clear-cut descriptions of selected products in the online-shop. The data was generated directly by Google Adwords as part of the normal campaign evolution of the advertiser.

TABLE 4.2: summary of click-rates of ads in the given dataset

data	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Group 2	0.00000	0.07639	0.14290	0.14490	0.19700	0.42860
Group 1	0.02941	0.10380	0.15000	0.17590	0.19050	0.50000

The test has been carried out in the first half of 2013. The resulting dataset contains a large number of Sponsored Search key performance indicators (KPI) for the given period as exemplified in table 4.1. The content of the unfiltered dataset as well as the exact dates of the test period cannot be revealed to ensure confidentiality for the advertiser and are of no importance for what follows from here. To ensure that only the impact of the specific text alternation is analyzed and to exclude other factors that would blur the results, especially the strong position effects we describe above, we only analyze the advertisements that were displayed above the organic search results and that were part of the described A/B test. The updated dataset, which is only a small fraction of the advertisers' regular Sponsored Search campaign, includes a total of 110 advertisements of which 49 advertise "Carbon Neutral delivery" **(Group 1)** while the other 61 advertise "Fast and Reliable delivery" **(Group 2)** in the third row of the advertisement as illustrated in fig 5.1. It contains a total number of 42,364 impressions and 5,775 clicks. What is used for the analysis is the aggregated CTR for each ad over the whole test period.

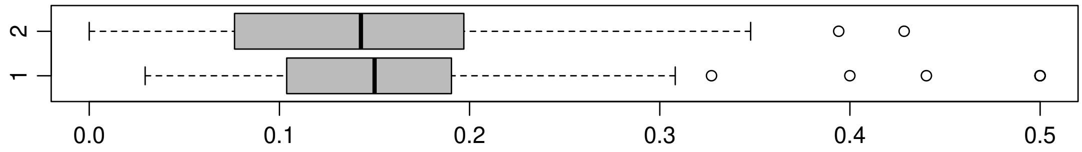

FIGURE 4.2: Empirical comparison of CTR's for Group 2 and Group 1

Analytic approach: Traditionally one makes probabilistic assumptions about the magnitude of the difference between two observed groups by using null hypothesis significance testing ($NHST$). (Kruschke, 2012) Applying an unpaired t-test, the two-tailed P value equals 0.1420 for the given data. This difference is considered to be not statistically significant. The mean of Group 1 minus Group 2 equals 0.031. The 95% confidence interval of this difference ranges from -0.0105 to 0.0725. In fact this is almost all the information one gets from conventional $NHST$ statistics.

We, however, apply a Bayesian approach to answer the question whether there is a positive, negative or zero impact of sustainability information in ad texts in Sponsored Search advertising and await richer information by unsing Bayesian Methods. To do so, we are comparing the two groups of users that took part in the described A/B test. Even though the applied method possibly influences the behavior of the involved users and could therefore be categorized as reactive in terms of social sciences, it shares common criteria with non-reactive methods since individual users have no knowledge of the investigation of their behavior.

We follow Kruschke and describe the data using mean and standard deviation parameters for t-distributions representing both groups individually and add a normality parameter that is common for both groups. The prior allocation of credibility across the parameters is vague, so that the prior has minimal influence on the estimation, to let the data dominate the inference. Taking the data into account the Bayesian estimation reallocates credibility to parameter values that represent the observed data best. The resulting distribution is a joint distribution across the five parameters, thereby revealing combinations of the five parameter values that are credible, given the data (Kruschke, 2012). The two histograms in the top right in fig. 4.3 are representations of empirical data and display the two observed groups (group 1 = "Carbon Neutral delivery", group 2 = "Fast and Reliable delivery"), with curves of representative examples of posterior predictive t-distributions. In the left column you will find marginals of the posterior distributions of credible values of means of group 1 and 2 as well as the same for the respective standard deviations and a distribution of credible values for the the combined normality parameter. Lower right

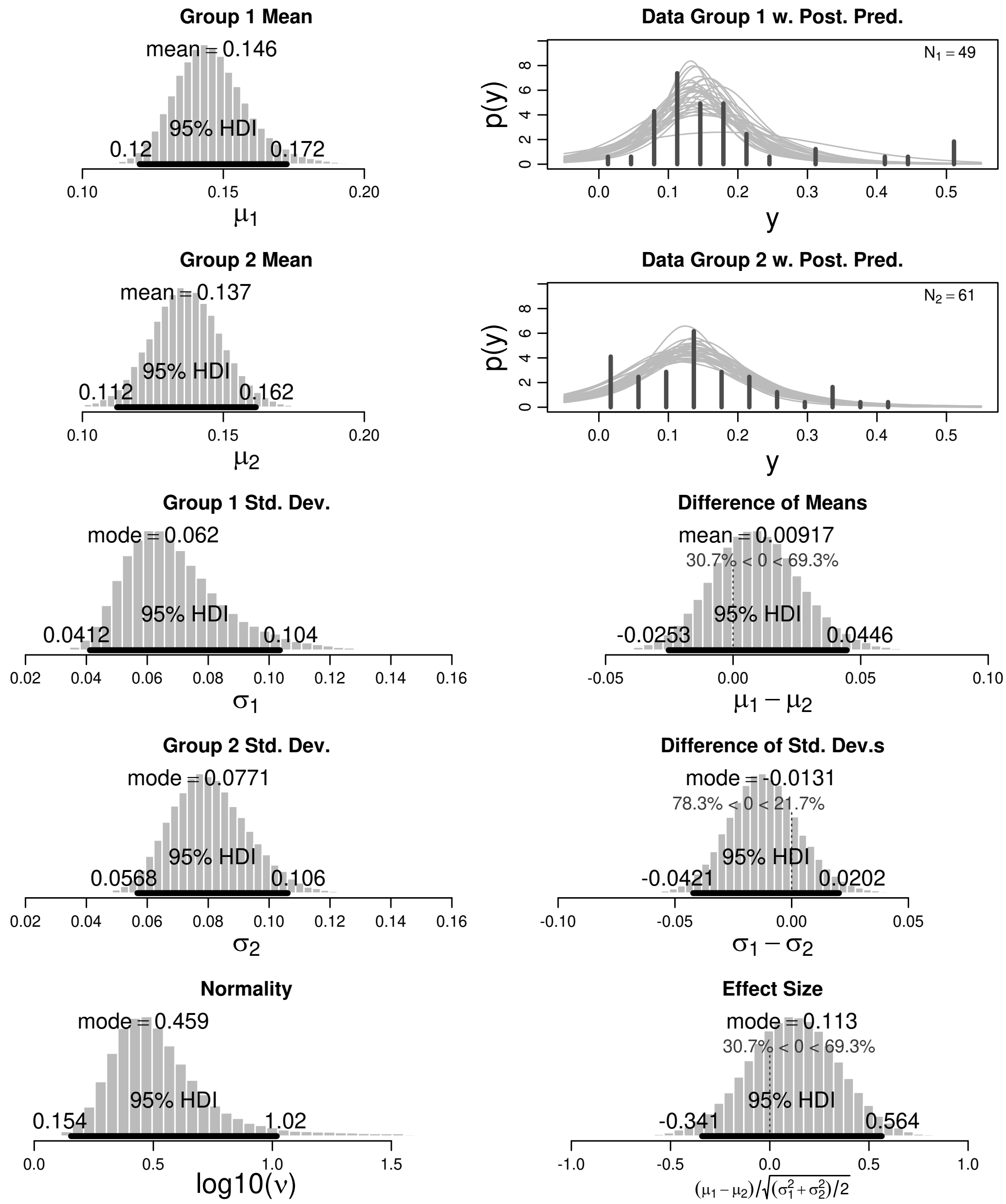

FIGURE 4.3: Group 1 = CTR for ads advertising Carbon Neutral delivery, Group 2 = CTR for ads advertising Fast and Reliable delivery.

shows posterior distribution of differences in means and effect size. (Kruschke, 2012) Fig. 4.4 displays pairwise plots of the parameters for the given study.

4.4 Conclusions and Outlook

Taking a first look at the data as displayed in table 4.2 we find a slightly lower empirical mean CTR over all ads on ads that advertise "Fast and Reliable delivery" (14.39%) than on the "Carbon Neutral delivery" ads (15.94%). These values are not to be confused with those in the top left histograms in fig. 4.3 which represent the simulated mean parameters of t-distributions to fit the empirical distribution. So, in the data we observe a 1.55% higher empirical mean CTR for "green" ads which would eventually make us accept the hypotheses that ads with green marketing signals have a higher click probability than their counterparts in the A/B test. What is the central question is whether this result is significant and if it enables us to derive inferences about the "real" long-term distribution.

TABLE 4.3: Estimated parameters of the A/B Test results

	mean	median	mode	HDIlow	HDIhigh	pcgtZero
mu1	0.15	0.14	0.14	0.12	0.17	
mu2	0.14	0.14	0.13	0.11	0.16	
muDiff	0.01	0.01	0.01	-0.03	0.04	69.26
sigma1	0.07	0.07	0.06	0.04	0.10	
sigma2	0.08	0.08	0.08	0.06	0.11	
sigmaDiff	-0.01	-0.01	-0.01	-0.04	0.02	21.67
nu	4.23	3.11	2.49	1.22	9.82	
nuLog10	0.54	0.49	0.46	0.15	1.02	
effSz	0.12	0.12	0.11	-0.34	0.56	69.26

To answer this question 100,000 parameter combinations for t- distributions that are credible given the data are generated by Markov Chain Monte Carlo simulation ($MCMC$). One gets a good insight by comparing the distribution of credible values for μ_1 which has a mean of 0.146 and a 95% Highest Density Interval (HDI) from 0.120 to 0.173 with μ_2 which has a mean of 0.136 with a 95% HDI from 0.112 to 0.161 as can be seen in tab. 4.3. The exact difference $\mu_1 - \mu_2$ is 0.009 on average as can be found in the plot in the middle of the right column of fig. 4.3. One can see that 69.3% of the 95% HDI for $\mu_1 - \mu_2$ is positive. What is even more relevant for the analysis is that all computed values within the 95% HDI fall into the Region of Practical Equivalence ($ROPE$) which spreads from -0.1 to 0.1. So, these results

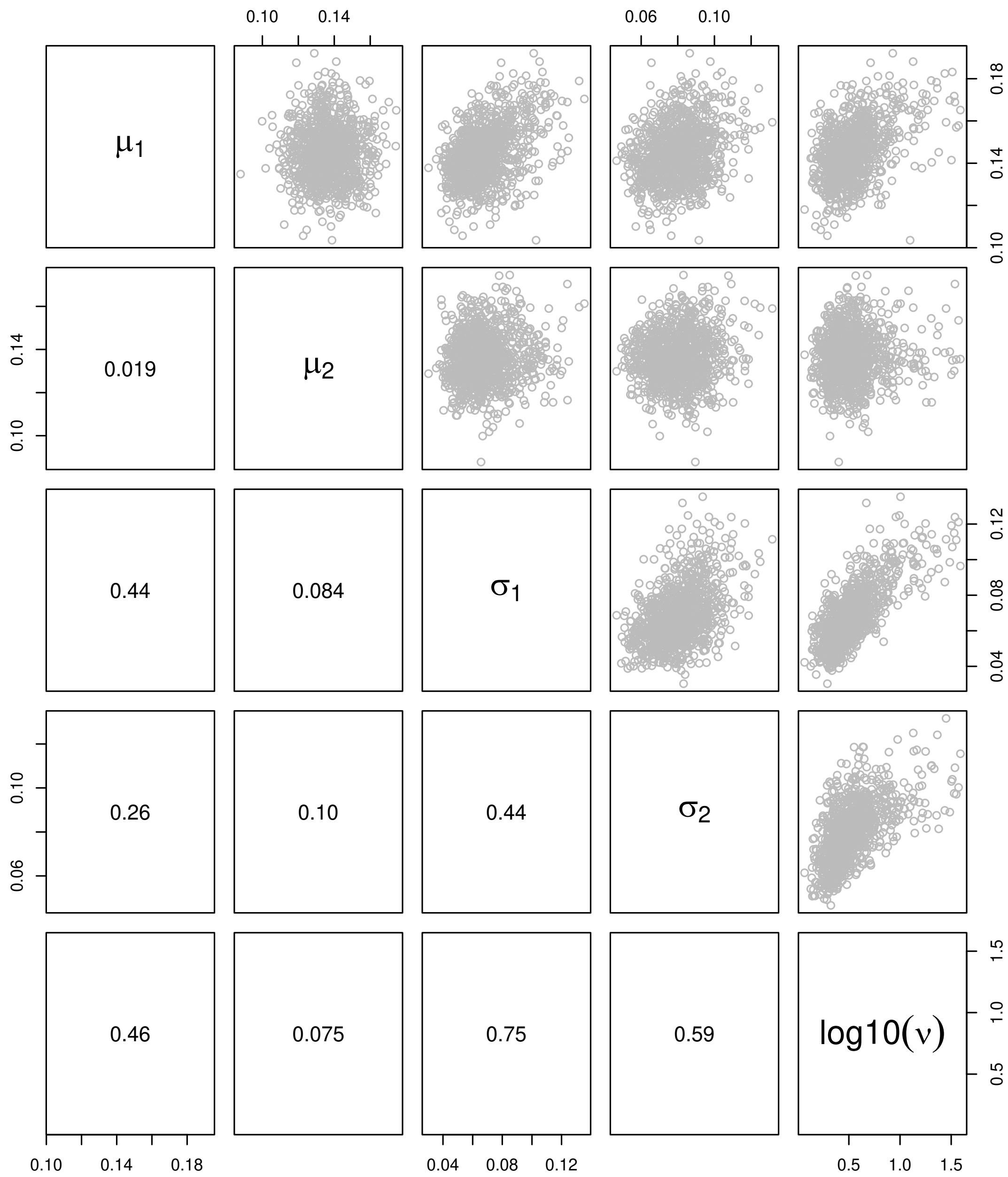

FIGURE 4.4: Posteriors for Bayesian Parameter Estimation

imply that there is a 69.3% chance that the "real" mean of group 1 is grater than the "real" mean of group 2. Nevertheless the difference of means is so small that there is a high probability that the groups are not credibly different from each other in this aspect. Comparing the distribution of credible values for σ_1 and σ_2 one can see that these groups do not credibly differ too. This can be seen in the respective histogram in fig. 4.3 where all computed values for $\sigma_1 - \sigma_2$ are found in the $ROPE$ with 78.3% being negative and 21.7% being positive. This suggests that there is a 78.3% probability that the standard deviation for group 2 is grater than for group 1.

The lower right panel of fig. 4.3 shows the distribution of credible effect sizes, given the data. For each combination of means and standard deviations, the effect size is computed. The histogram of 100,000 credible effect sizes has a mode of 0.123 and the zero included in the 95% HDI. 69.3% of all computed outcomes are positive while 30.7% are negative. (Kruschke, 2012)

What can we derive from that? What is true is that there is some probability that there is absolutely no effect caused by the different signals in the advertisements as we do not observe strongly significant unambiguous results. If any effect is presumed, it will have a higher probability of being positive for "green signals" in Sponsored Search ads, given the observed data. How can this outcome be explained? One argument could be that ad texts do not influence users on SERPs at all. Although we know about various other effects, like the strong position bias described above, that do affect the user there are too many indications that ad texts do have influence on click decisions to let this be true.

In fact, these results need to be interpreted with caution. One possible explanation for this is that users might not be as green in their decisions as marketers would like them to be. In this case the promise of "Fast and Reliable delivery" seams to lead to a slightly lower motivation to click on an ad than the green signals the advertiser sends out to his potential customers. This A/B test should be repeated over a number of various branches before one can derive implications for the whole e-Commerce industry. What is an even more interesting outcome of this paper is that more future research should be conducted on the general impact of texts in Sponsored Search ads considering a variety of branches and containing more diversity in texts to make sophisticated assumptions on the impact of text-details on click probabilities.

References

Adar, E. (2007). "User 4xxxxx9: Anonymizing query logs". In: *Proc of Query Log Analysis Workshop, International Conference on World Wide Web - WWW '07*.

Agarwal, Ashish, Kartik Hosanagar, and Michael D Smith (2011). "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets". In: *Journal of Marketing Research* 48.6, p. 1057.1073.

Alby, T. and B. Funk (2011). "Search Engine Marketing in Small and Medium Companies: Status Quo and Perspectives." In: *E-Business Managerial Aspects, Solutions and Case Studies*. Ed. by M. Cruz-Cunha and J. Varajão. igi-global, pp. 206–221.

Blask, T., B. Funk, and R. Schulte (2011). "Should Companies Bid on their Own Brand in Sponsored Search?" In: *Proceedings of the International Conference on e-Business (ICE-B 2011)*. Seville, Spain, pp. 14–21.

Blask, Tobias (2013). "Applying Bayesian Parameter Estimation to A / B Tests in E-Business Applications : Examining the Impact of Green Marketing Signals in Sponsored Search Advertising". In: *Proceedings of the International Conference on e-Business (ICE-B 2013)*.

Brenes, David J. and Daniel Gayo-Avello (2009). "Stratified analysis of AOL query log". In: *Information Sciences* 179.12, pp. 1844–1858.

Chapelle, Olivier and Ya Zhang (2009). "A dynamic bayesian network click model for web search ranking". In: *Proceedings of the 18th international conference on World wide web - WWW '09*. New York, New York, USA: ACM Press, p. 10.

Craswell, Nick et al. (2008). "An experimental comparison of click position-bias models". In: *Proceedings of the international conference on Web search and web data mining - WSDM '08*. New York, New York, USA: ACM Press, p. 8.

Dave, Kushal S. and Vasudeva Varma (2010). "Learning the click-through rate for rare/new ads from similar ads". In: *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. July. New York, New York, USA: ACM Press, p. 897.

Evans, David S (2008). "The Economics of the Online Advertising Industry". In: *Review of Network Economics* 7.3, pp359–391.

– (2009). "The Online Advertising Industry: Economics, Evolution, and Privacy". In: *Journal of Economic Perspectives* 23.3, pp. 37–60.

Gauzente, Claire (2009). "Information search and paid results proposition and test of a hierarchy-of-effect model". In: *Electronic Markets* 19.2-3, pp. 163–177.

Gluhovsky, Ilya (2010). "Forecasting Click-Through Rates Based on Sponsored Search Advertiser Bids and Intermediate Variable Regression". In: *October* 10.3, pp. 1–28.

Haytko, DL and Erika Matulich (2008). "Green advertising and environmentally responsible consumer behaviors: Linkages examined". In: *Journal of Management and Marketing Research* 1, pp. 2–11.

Jansen, Bernard J. et al. (2009). "The Components and Impact of Sponsored Search". In: *Computer* 42.5, pp. 98–101.

Jansen, B.J. and T. Mullen (2008). "Sponsored search: an overview of the concept, history, and technology". In: *International Journal of Electronic Business* 6.2, pp. 114–131.

Jansen, Jim (2011). "The Serious Game of Bidding on Keywords". In: *Understanding Sponsored Search Core Elements of Keyword Advertising*. Cambridge University Press. Chap. 8, pp. 176–201.

Joachims, Thorsten et al. (2005). "Accurately interpreting clickthrough data as implicit feedback". In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05*. New York, New York, USA: ACM Press, p. 154.

Joachims, Thorsten et al. (2007). "Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search". In: *ACM Transactions on Information Systems* 25.2, pp. 1–26.

Kruschke, John K (2012). "Bayesian Estimation Supersedes the t Test." In: *Journal of experimental psychology. General*.

Leonidou, Leonidas C. et al. (2011). "Evaluating the green advertising practices of international firms: a trend analysis". In: *International Marketing Review* 28.1, pp. 6–33.

Pass, Greg, Abdur Chowdhury, and Cayley Torgeson (2006). "A picture of search". In: *Proceedings of the 1st international conference on Scalable information systems - InfoScale '06*. New York, New York, USA: ACM Press, 1–es.

Rex, Emma and Henrikke Baumann (2007). "Beyond ecolabels: what green marketing can learn from conventional marketing". In: *Journal of Cleaner Production* 15.6, pp. 567–576.

Richardson, M., E. Dominowska, and R. Ragno (2007). "Predicting clicks: estimating the click-through rate for new ads". In: *Proceedings of the 16th international conference on World Wide Web*. ACM New York, NY, USA, pp. 521–530.

Strohmaier, M. et al. (2007). "How do users express goals on the web?-An exploration of intentional structures in web search". In: *LECTURE NOTES IN COMPUTER SCIENCE* 4832, p. 67.

Strohmaier, Markus, Peter Prettenhofer, and Mark Kröll (2008). “Acquiring Explicit User Goals from Search Query Logs”. In: *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. Vol. 8. Ieee, pp. 602–605.

Varian, H.R. (2009). “Online ad auctions”. In: *American Economic Review* 99.2, pp. 430–434.

Yang, S. and A. Ghose (2010). “Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?” In: *Marketing Science* 30.1, pp. 1–22.

Zhong, Feimin et al. (2010). “Incorporating post-click behaviors into a click model”. In: *Annual ACM Conference on Research and Development in Information Retrieval*, pp. 355–362.

Zhu, Z.A. et al. (2010). “A novel click model and its applications to online advertising”. In: *Proceedings of the third ACM international conference on Web search and data mining*. ACM, pp. 321–330.

Chapter 5

Do Specific Text Features Influence Click Probabilities in Paid Search Advertising?

Tobias Blask

Abstract

Paid Search Advertisers have only very few options to influence the user's decision to click on one of their ads. The textual content of the creatives seems to be one important influencing factor beneath its position on the Search Engine Results Page (SERP) and the perceived relevance of the given ad to the present search query. In this study we perform a non reactive multivariate test that enables us to evaluate the influence of specific textual signals in Paid Search creatives. A Bayesian Analysis of Variance (BANOVA) is applied to evaluate the influence of various text features on click probabilities. We conclude by finally showing that differences in the formulation of the textual content can have influence on the click probability of Paid Search ads

This article has been published as follows: Tobias Blask (2014). "Do Specific Text Features Influence Click Probabilities in Paid Search Advertising?". In: *Proceedings of the international Conference on E-Business 2014*. One additional reference to Kruschke (2010) has been added on p. 87.

5.1 Introduction

Internet search engines play a key role in the modern information society. Not only do they serve the information needs of their users but they also represent an important source of customer acquisition for a variety of companies (Jansen and Mullen, 2008; Alby and Funk, 2011). Internet search engine companies also generate significant amounts of revenue through Paid Search Advertising. While still growing rapidly, Paid Search Advertising already dominates the online media spending of companies that advertise on the internet. Developed in 1998 by Overture, this form of advertising uses text advertisements and a list of keywords. The advertiser also usually provides attributes to each of these keywords, but always indicates the amount of money he is willing to pay for a click on an ad for a specific keyword (CPC_{max})(Jansen et al., 2009). Every time a user types a query into a search engine, personalized result pages are generated based on the user's location, search history and other factors. If ads are available that could probably satisfy the need of the user, the search engine displays these ads alongside the organic results. If more than one advertiser is willing to pay for the display of an ad, the search engine typically auctions the position of these ads among all interested parties through a Generalized Second Price Auction (GSP) (Jansen, 2011; Varian, 2009). Only the advertiser that wins the auction by getting a click on its ad is charged by the search engine. The effective Cost-Per-Click (CPC_{eff}) is the maximum bid of the advertiser with the subsequent highest bid plus a small additional fee. In practice, search engine companies use a more robust mechanism to maximize their profits by rewarding keyword/ad combinations that have a high relevance to users (often referred to as the quality score). Although detailed calculations are not disclosed, the key metric is claimed to be the historic Click-Through Rate (CTR) where available, otherwise an expected click probability for the specific advertiser-ad-keyword combination is used. An interesting issue for advertisers is how to maximize the probability that a given user will click on one of their advertisements, ultimately fulfilling a defined goal on their website. In practice there are only limited options to do so. One is to optimize the relevance of an advertisement by only choosing keyword / landing page combinations that provide a suitable offer to the respective query of a given user. Additionally, advertisers can maximize click probability by influencing the position of an ad on the SERP via the CPC_{max} and by improving the ad quality. Finally, optimizing the wording of creatives to communicate advantages over the competition may help users with their decision on which ad to click. In the present paper we concentrate on the impact of various signals in text advertisements. We analyze a non-reactive multivariate test in which users are confronted with some defined variations of ads containing information on **trust** and **pricing** that has been differently formulated or completely omitted. We evaluate the probability that a user will click on a given Paid Search text ad by using a Bayesian Analysis of Variance (BANOVA). Finally, we illustrate that variations in the text of a creative have significant influence on click probabilities in Paid Search.

5.2 Paid Search Advertising

An interesting topic for the current research is evaluating the factors that influence the probability that a given user will click on a specific advertiser's ad. These factors have been widely studied since Paid Search Advertising first began; however, due to a lack of opportunity to observe search engine user behavior, complete coverage of all factors influencing the click probability is no easy task. Evidence suggests that one of the most influential factors is the ad position within the Paid Search results, which depends on the CPC_{max} and quality score. The quality score is used by search engines to determine the quality of an advertisement and is based primarily on the historical CTR of a given keyword/ad combination. There is a strong correlation between decreasing ad position and a decreasing CTR (Richardson, Dominowska, and Ragno, 2007; Agarwal, Hosanagar, and Smith, 2011). In principle, the top positions lead to high $CTRs$. From an advertiser's perspective, it is appealing to be able to predict the future CTR of a given ad or even better to find rules for predicting click probability in advance. As research has progressed, several models have been developed to explain the influence of the position bias on the CTR.

Crasswell, Zoeter and Taylor (Craswell et al., 2008) present several models for predicting the CTR: (a) baseline model, (b) mixture model, (c) examination model, and (d) cascade model . The findings were originally based on organic search results, but they are applicable to Paid Search results as well (Agarwal, Hosanagar, and Smith, 2011). The underlying assumption of the (a) baseline model is that a user screens every search result and then decides, which one best fits the query. As a result, the click probabilities for each individual search result are identical and independent of its position. The (b) mixture model extends the baseline model and divides user behavior into two groups. One group behaves as described in the baseline model, while the other group clicks randomly on one of the first search results. The (c) examination model refers to findings from eye tracking studies, which state that, with declining position, the probability of a click also declines (Joachims et al., 2005;

Joachims et al., 2007). The (d) cascade model is, owing to the high degree of explanation by click data, one of the most applied explanation approaches. The basic assumption is that the user scans each search result from top to bottom, comparing the relevance of each ad with the relevance of the ad before it. The user continues scanning the results until the perceived ad relevance reaches a certain level and the user clicks.

One challenge is to predict the CTR of keywords or keyword combinations for potential future Paid Search ads. A proposed solution is aggregating historical data from similar keywords. Here, the CTR is represented as a function of position, independent of a bid. The resulting developed models do not focus on a certain advertiser. The same clustering approach can be applied in optimizing the search engines' profit (Dave and Varma, 2010). There are also models that take into account the quality score (Gluhovsky, 2010). The General Click Model model developed by Zhu et al. (Zhu et al., 2010), focuses on the CTR prediction of long-tail queries, based on a Bayesian network.To address the the aforementioned position bias, Zhong et al. (Zhong et al., 2010) incorporate post-click user behavior data from the clicked ad's landing page into the click model to refine the estimation of the perceived user relevance after clicking on a specific ad. A similar approach using Dynamic Bayesian networks can be found in Chappelle and Zhang (Chapelle and Zhang, 2009). Several models based on historical click data are limited in that they lack consideration of a possible user learning effect. Taking Gauzente's results as an example, it has been shown that past user satisfaction with Paid Search results influences current click behavior (Gauzente, 2009). In addition to the incorporation of position data and the perceived relevance of presented ads, the CTR of an advertisement is also affected by the relationship between organic and Paid Search results. Listing the results in Paid and organic search results for one company at the same time leads to a higher CTR (Yang and Ghose, 2010; Blask, Funk, and Schulte, 2011). What has often been overlooked is the influence that specific text patterns have on click probabilities.

5.3 Case Study

It is part of a Paid Search Advertising manager's daily routine to test different versions of a specific ad. In practice, at least two variations of an ad are tested against each other in each ad group. One commonly method is to replace the weaker performing version of an ad with the stronger variation after enough clicks are generated to identify which one is performing better. Anecdotal evidence and personal experience often play an important role in this process, and the knowledge gained from tests is often not preserved within an organization. In this paper we present a method for multivariate tests based on historical data that is able to enhance the options of working with unbalanced study designs in A/B and multivariate tests. This ultimately improves advertisers' ability to recognize low performing ads sooner than with conventional ANOVA methods. What makes these models interesting is the ability to take prior knowledge into account where only sparse data are available.

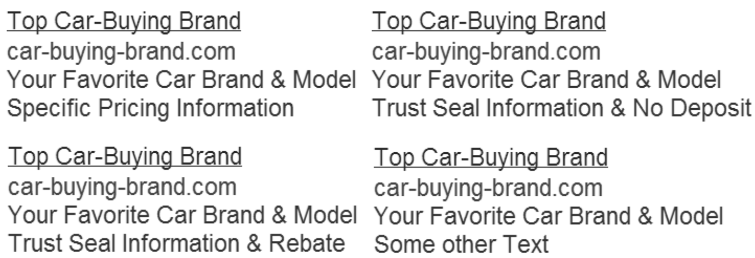

FIGURE 5.1: Variations of Paid Search ads similar the textual ads used for this study

For this study, selected elements have been altered in a number of Paid Search text advertisements for very similar commercial Google search queries. These queries may lead to a business offer from the advertiser and an online product purchase, as can be seen in fig. 5.1. The advertisers' product is a major investment for the average private customer. The data were generated directly by Google Adwords as part of the normal conducted from 2012 to 2013. The resulting data set contains a large number of Paid Search key performance indicators (KPI) for the given period. All advertisements that had less than 100 impressions in the given period were filtered out. To ensure that only the impact of the specific text alteration is analyzed and to exclude other factors that would blur the results (especially the aforementioned strong position effects), we only analyze the advertisements that were displayed above the organic search results and that were part of the described multivariate test. For this study, we take almost **3 million ad impressions** resulting in **more than 300,000 clicks** and a total of **1.976 ads** into account. The occurrences of the examined text features in the ads are shown in table 5.1. The aggregated CTR for each ad over the whole test period is used for the analysis. The resulting mean $CTRs$ are displayed in fig 5.3. The nominal variables **Title**, **Body 1**, **Body 2**, **Display-URL** of the landing page for each ad are also included.

TABLE 5.1: Occurences of text features in the dataset (trust information(x1) and pricing information(x2))

	no trust info($X1A1$)	**test winner($X1A2$**	**trust seal($X1A3$)**
no pricing($X2B1$)	789	99	330
no deposit($X2B2$)	20	35	91
save x %($X2B3$)	91	84	101
real price($X2B4$)	39	306	4

In this study we want to predict the metric variable CTR by using the described nominal textual predictors. As such, ANOVA is a valid method of choice for us. For the given data the model can be formulated as written in equation (5.1) with the predictor variables **presence of trust seal information** and **presence of pricing information** denoted as $\vec{x}_1$ and $\vec{x}_2$. β serves as a deflection parameter. β_0 indicates the baseline value of the prediction. For example, if x_2 is at the value of x_{2k}, a deflection of β_{2k} is added to the baseline. Ultimately the sum of all deflections of β_1 and β_2 have to have a sum of zero for both predictors.

$$
\begin{aligned}
y &= \beta_0 + \vec{\beta_1}\vec{x_1} + \vec{\beta_2}\vec{x_2} + \vec{\beta}_{1\times 2}\vec{x}_{1\times 2} \\
&= \beta_0 + \sum_{j=1}^{J_1} \beta_{1,j}\, x_1 + \sum_{j=1}^{J_2} \beta_{2,j}\, x_2 + \sum_{j=1}^{J_1}\sum_{k=1}^{J_2} \beta_{1\times 2,j,k} x_{1\times 2,j,k} \\
&\text{with the constraints} \\
&\sum_{j=1}^{J_1} \beta_{1j} = 0 \text{ and} \sum_{k=1}^{J_2} \beta_{2k} = 0 \text{ and} \\
&\sum_{j=1}^{J_1} \beta_{1\times 2,j,k} = 0 \forall k \text{ and} \sum_{k=1}^{J_2} \beta_{1\times 2,j,k} = 0 \forall k
\end{aligned}
\tag{5.1}
$$

Even a brief look at the numbers in table 5.1 makes it clear that the research design is not very well balanced. Only four ads contain trust seal and concrete pricing information from the advertiser's database, while 345 contain no trust seal information but do include concrete pricing information. This could lead to serious computational difficulties in traditional ANOVA, which is the reason that we follow Kruschke's approach (Kruschke, 2010). We use a Bayesian estimation to perform the data analysis using a hierarchical prior as illustrated in fig. 5.2 . The goal of this analysis is to estimate the additive and interactive β values for each level of $\vec{x}$.

The following assumptions are made regarding the hierarchical prior: The observed data y_i is assumed to be normally distributed around the predicted value or central

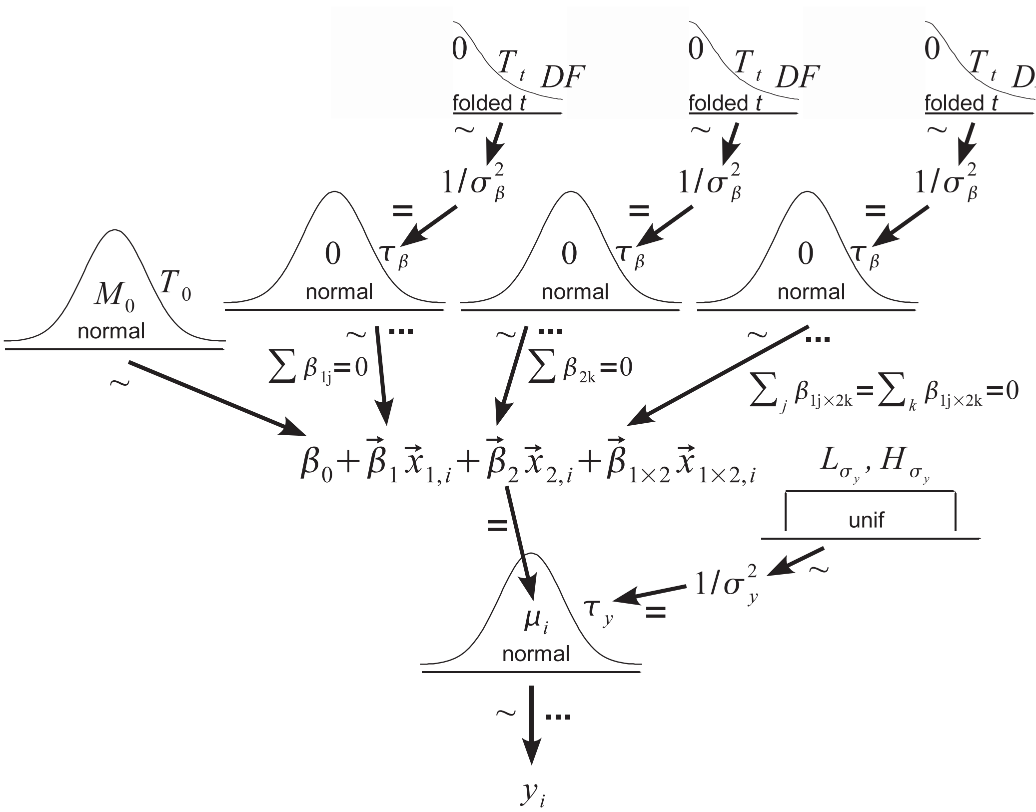

FIGURE 5.2: Hierarchical dependencies for two-way BANOVA (Kr-uschke,2010)

tendency μ_i. As defined and illustrated by Kruschke (Kruschke, 2010), the equation above the observed data distribution in fig 5.2 illustrates that the predicted value is composed from the baseline (β_0) plus additive deflection ($\beta_0 + \vec{\beta_1}\vec{x_1} + \vec{\beta_2}\vec{x_2}$) caused by each predictor individually, and interactive deflection ($\vec{\beta}_{1\times2}\vec{x}_{1\times2}$) caused by the combination of the given predictors. The basic assumptions about the respective β_i can be found in the distributions in the top level of fig. 5.2. Here, we indicate that all β_i are normally distributed around zero. The variances of all β_i are estimated from the given data. The hyperdistributions are applied separately to the various predictors and interactions. This is due to the assumption that the magnitude of the effect of the predictive variable $\vec{x_1}$ is probably not informative on the magnitude of the effect of $\vec{x_2}$ (Kruschke, 2010).

As previously mentioned, we want to predict CTR using variations of the nominal predictors **"trust seal information"** and **"pricing information"**, denoted as $\vec{x_1}$ and $\vec{x_2}$. Both variables have several levels. $\vec{x_1}$ includes **"no trust information given"** (x_{1A1}), **"unspecified test winner information given" (**x_{1A2}**)** and **"concrete trust seal and test winner information given"(**x_{1A3}**)**. $\vec{x_2}$ includes **"no pricing information"**(x_{2B1}), **"no deposit"**(x_{2B2}), **"percentage of savings given"**(x_{2B3}) and **"concrete pricing information from database given"**(x_{2B4}). Although that we have an unbalanced study design and very few observations for one of the cases, we can see from fig 5.3 that ads with different contents gain significantly varying $CTRs$.

The results of the Bayesian analysis concerning the effects of the text features are shown in fig 5.4. The top left histogram shows that the baseline (β_0) for the given combinations of text features is at 0.114. This is quite high in terms of average Click-Through Rates in Paid Search Advertising in general, at least for such a great number of queries as we observed within the test period in the context of the given campaigns. This could be a good hint for the fact that the campaigns have already been very well organized and optimized in terms of relevance to the specific queries that lead to the display of ads of the given advertiser on Search Engine Results Pages. Which influences selected text features have in such an environment can be seen in the remaining histograms in fig. 5.4. Each histogram illustrates deflections from the baseline for any given feature combination. The third histogram in the first row for example indicates that 95% of the most credible values that have to be subtracted from the baseline to describe the effect of the text feature "unspecified test winner information" fall in the area between 0.0282 and 0.0164 ($\beta1_2$) with the most probable value at 0.0213. What is really interesting and makes this kind of analysis so helpful in the case of unbalanced study design is the additional information concerning

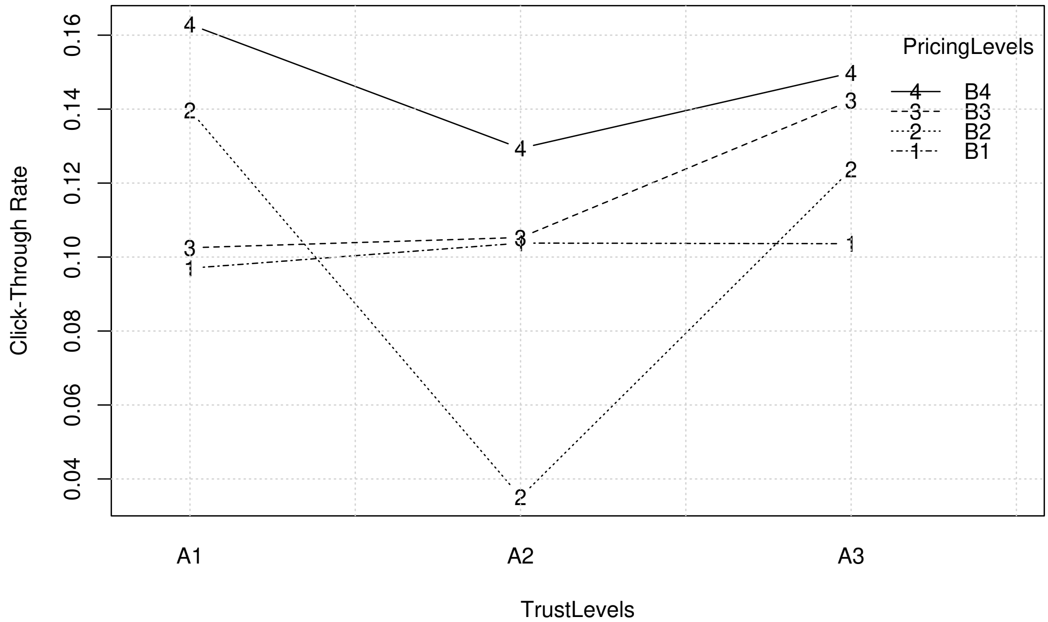

FIGURE 5.3: Mean CTR values for ads with different text feature combinations

the 95% HDI (Highest Density Interval) for each estimation. This interval illustrates the area in which 95% of all credible parameter values for the respective level of a variable are situated. This becomes especially important when it comes to levels of a variable with only very few observed data points where the in-group variance is estimated with the help of the prior knowledge from other levels within the same predictive variable.

We estimate the effects for each of the groups but we are also interested in the answer to the question whether the groups are credibly different from each other. In typical A/B test scenarios for example this can be examined by applying an $NHST$ t-test or a Bayesian Parameter Estimation of t-distributions for the comparison of two groups (Kruschke, 2012; Blask, 2013). For the case of more than two predictive variables this can also be done via contrast analysis in Multifactor Analysis of Variance (MANOVA).

What we additionally want to investigate in this study is the overall effect of having information about a trust seal in the Paid Search Ads of the advertiser and the effects of various levels of pricing information. One answer to this question comes from the analysis whether there is a credible difference in click probabilities between

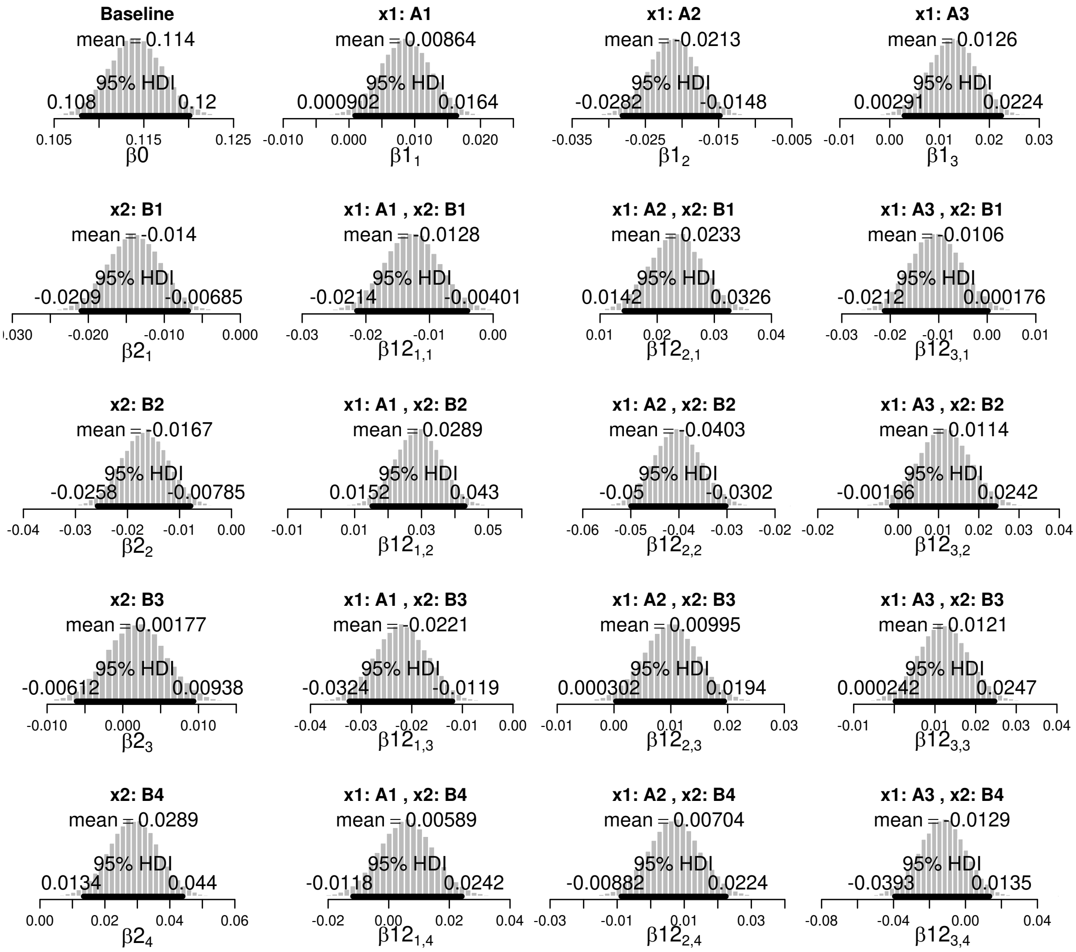

FIGURE 5.4: beta value posterior distributions for each variation given the data (**x1 = trust seal informaton** (x1:A1 = no trust info, x1:A2 = unspecified test winner information, x1:A3 = concrete trust seal and test winner information), **x2 = pricing information** (x2:B1 = no pricing, x2:B2 = no deposit, x2:B3 = save x%, x2:B4 = concrete pricing information from database))

ads that do not contain any specific trust information (x_{1A1}) and and those that include the text feature (x_{1A2} and x_{1A3}). The histograms in fig 5.5 that the effect of having unspecified test winner information in the ad does not help the advertiser to gain a higher CTR. In fact the analysis does reveal that having no trust information is credibly better than the announcement of an unproven test winner statement. Including a trust seal information into the ad does not have such a negative effect. In fact about two thirds of the credible values for the effect, including the most probable value, indicate that this feature may slightly help the advertiser. What is also true is that the zero value is included in the 95% HDI what makes it very probable that there is no credible improvement in the advertisers performance by including this feature. What we can derive from the analysis is the fact that it makes credibly more sense to include the proven trust seal into the ad compared to unproven test winner statement as can be seen in the right histogram in fig 5.5. So, in terms of trust seal information it becomes quite clear what the better choice might be for the given company. It makes no mistake by taking concrete and proven information on trust seals into their ads. What they should not expect is a significant boost in terms of click probability.

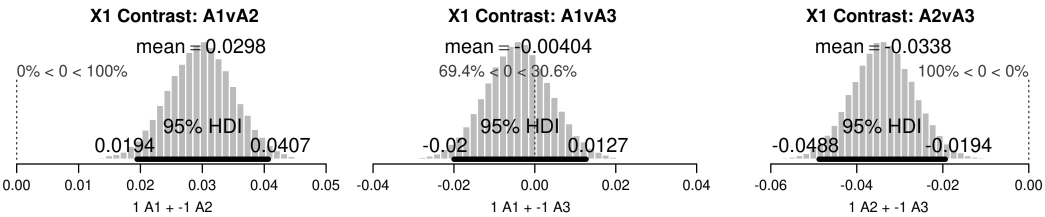

FIGURE 5.5: contrasts for various levels of trust seal information ($X1$)

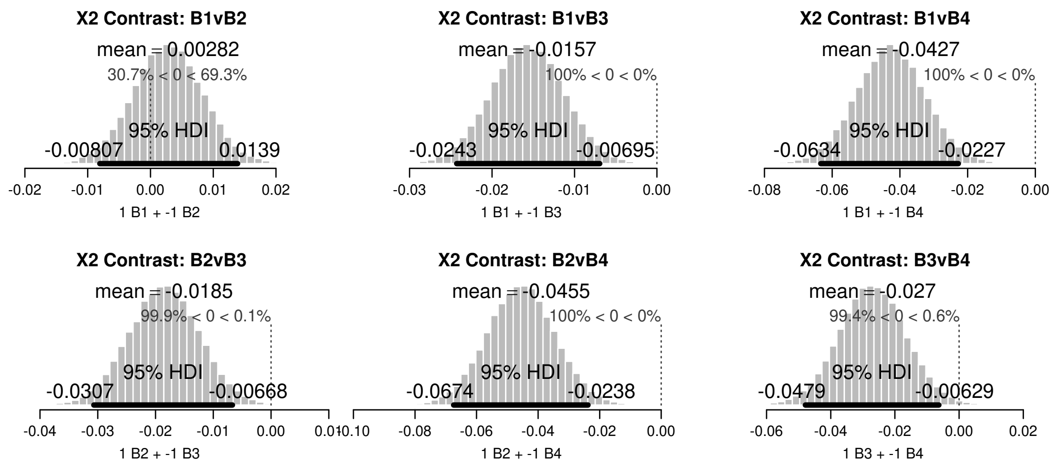

FIGURE 5.6: contrasts for various levels of pricing information ($X2$)

What is not so clear until now is the question which level of pricing information finally leads to the best click probability. What we can see is that there seem to exist two differently performing clusters. On the one hand there are the ads that contain a concrete pricing information from the database of the given advertiser (x_{2B4}) and ads that contain a specific discount in percent as text feature (x_{2B3}) and on the other hand there are the ads with no pricing information at all (x_{2B1}) and those advertising that no deposit has to be made (x_{2B2}). These two groups are credibly different from each other as can be seen in the histograms in fig 5.6. In detail, everything, including no pricing information at all (x_{2B1}), seems to be be better than advertising "no deposit" (x_{2B2}). The next best text feature in terms of pricing is to give an exact value for the percentage that a user can save on the advertiser's website (x_{2B3}). This feature is performing credibly better than those mentioned above. What is the best way to communicate pricing in Paid Search Ads - given the data - is to provide exact pricing information from the advertiser's database (x_{2B4}). In fact it is credibly superior to any other feature in terms of pricing communication, given the observed data.

5.4 Conclusions and Outlook

What we applied in this paper offers a valid way to evaluate text features and other nominal predictive variables where tests are an essential part of the daily business. In terms of substantive issues it is the hard facts that the potential customers are looking for when they research in a search engine. **The more specific information on pricing is provided in an ad - the better is the chance of winning the customers click.** Building up trust is one good feature for an advertiser to support this effect or even substitute parts of this positive effect if they do not have competitive prices or special rebates available. In this specific case this has been achieved by communicating the existence of a credible trust seal in the ad-copy. What we did not asses in this research but would find interesting for an ongoing investigation is the question whether these findings have additional impact on the conversion probability on the advertiser's landing-page as well. Applying Bayesian ANOVA to multivariate tests in Online Advertising, especially Paid Search Advertising, has various advantages compared to applying conventional Analysis of Variance. This is especially true for unbalanced data like the present one. One obvious limitation to the results is that they should probably only be true for advertisers with competitive prices. Additionally this test should be repeated for a number of other advertisers from various industries to answer the question whether these observations can be generalized.

References

Agarwal, Ashish, Kartik Hosanagar, and Michael D Smith (2011). "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets". In: *Journal of Marketing Research* 48.6, p. 1057.1073.

Alby, T. and B. Funk (2011). "Search Engine Marketing in Small and Medium Companies: Status Quo and Perspectives." In: *E-Business Managerial Aspects, Solutions and Case Studies*. Ed. by M. Cruz-Cunha and J. Varajão. igi-global, pp. 206–221.

Blask, T., B. Funk, and R. Schulte (2011). "Should Companies Bid on their Own Brand in Sponsored Search?" In: *Proceedings of the International Conference on e-Business (ICE-B 2011)*. Seville, Spain, pp. 14–21.

Blask, Tobias (2013). "Applying Bayesian Parameter Estimation to A / B Tests in E-Business Applications : Examining the Impact of Green Marketing Signals in Sponsored Search Advertising". In: *Proceedings of the International Conference on e-Business (ICE-B 2013)*.

Chapelle, Olivier and Ya Zhang (2009). "A dynamic bayesian network click model for web search ranking". In: *Proceedings of the 18th international conference on World wide web - WWW '09*. New York, New York, USA: ACM Press, p. 10.

Craswell, Nick et al. (2008). "An experimental comparison of click position-bias models". In: *Proceedings of the international conference on Web search and web data mining - WSDM '08*. New York, New York, USA: ACM Press, p. 8.

Dave, Kushal S. and Vasudeva Varma (2010). "Learning the click-through rate for rare/new ads from similar ads". In: *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10*. July. New York, New York, USA: ACM Press, p. 897.

Gauzente, Claire (2009). "Information search and paid results proposition and test of a hierarchy-of-effect model". In: *Electronic Markets* 19.2-3, pp. 163–177.

Gluhovsky, Ilya (2010). "Forecasting Click-Through Rates Based on Sponsored Search Advertiser Bids and Intermediate Variable Regression". In: *October* 10.3, pp. 1–28.

Jansen, Bernard J. et al. (2009). "The Components and Impact of Sponsored Search". In: *Computer* 42.5, pp. 98–101.

Jansen, B.J. and T. Mullen (2008). "Sponsored search: an overview of the concept, history, and technology". In: *International Journal of Electronic Business* 6.2, pp. 114–131.

Jansen, Jim (2011). "The Serious Game of Bidding on Keywords". In: *Understanding Sponsored Search Core Elements of Keyword Advertising*. Cambridge University Press. Chap. 8, pp. 176–201.

Joachims, Thorsten et al. (2005). “Accurately interpreting clickthrough data as implicit feedback”. In: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '05*. New York, New York, USA: ACM Press, p. 154.

Joachims, Thorsten et al. (2007). “Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search”. In: *ACM Transactions on Information Systems* 25.2, pp. 1–26.

Kruschke, John K (2010). *Doing Bayesian Data Analysis : A Tutorial with R and BUGS*. Vol. 1. Academic Press, p. 672.

– (2012). “Bayesian Estimation Supersedes the t Test.” In: *Journal of experimental psychology. General*.

Richardson, M., E. Dominowska, and R. Ragno (2007). “Predicting clicks: estimating the click-through rate for new ads”. In: *Proceedings of the 16th international conference on World Wide Web*. ACM New York, NY, USA, pp. 521–530.

Varian, H.R. (2009). “Online ad auctions”. In: *American Economic Review* 99.2, pp. 430–434.

Yang, S. and A. Ghose (2010). “Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?” In: *Marketing Science* 30.1, pp. 1–22.

Zhong, Feimin et al. (2010). “Incorporating post-click behaviors into a click model”. In: *Annual ACM Conference on Research and Development in Information Retrieval*, pp. 355–362.

Zhu, Z.A. et al. (2010). “A novel click model and its applications to online advertising”. In: *Proceedings of the third ACM international conference on Web search and data mining*. ACM, pp. 321–330.