WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

EFMD
EQUIS
ACCREDITED

ePub^WU Institutional Repository

Nadia Abou Nabout

A Novel Approach for Bidding on Newly Set-Up Search Engine Advertising Campaigns

Article (Submitted)
(Refereed)

http://epub.wu.ac.at/

A Novel Approach for Bidding on Keywords in Newly Set-up Search Advertising Campaigns

Nadia Abou Nabout

Forthcoming in European Journal of Marketing

Prof. Dr. Nadia Abou Nabout, Department of Marketing, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria, email: nadia.abounabout@wu.ac.at, phone: +43 1 31 336 4900.

A Novel Approach for Bidding on Keywords in Newly Set-up Search Advertising Campaigns

Abstract

Purpose: Advertisers setting up search engine advertising campaigns for the first time need to place bids on keywords, but typically lack experience and data to determine ranks that maximize a keyword's profit (generally referred to as a cold-start problem). This article aims at solving the problem of bidding on keywords in newly set-up search engine advertising campaigns.

Approach: We suggest that advertisers collect data from the Google Keyword Planner to obtain precise estimates of the percentage increases in prices per click and clickthrough rates, which are needed to calculate optimal bids (exact approach). Together with the profit contribution per conversion and the conversion rate, the advertiser might then set bids that maximize profit. In case advertisers cannot afford to collect the required data, we suggest two proxy approaches and evaluate their performance using the exact approach as a benchmark.

Findings: The empirical study shows that both proxy approaches perform reasonably well—the easier approach to implement (proxy 2) sometimes performs even better than the more sophisticated one (proxy 1). As a consequence, advertisers might just use this very simple proxy when bidding on keywords in newly set-up SEA campaigns.

Originality/value: This research extends the stream of literature on how to determine optimal bids, which so far focuses on campaigns that are already running and where the required data to calculate bids is already available. This research offers a novel approach of determining bids when advertisers lack the aforementioned information.

Keywords: Electronic Commerce, Online Marketing, Search Engine Advertising, Campaign Set-Up, Bidding Decision, Cold-Start Problem

Introduction

Search engine advertising (SEA) has grown into a multibillion-dollar business that attracts about $1 of every $2 spent on online advertising (IAB 2014). The mechanism supporting SEA works as follows (Abou Nabout et al. 2012; Yao and Mela 2008): A consumer types a keyword, such as "cruise vacation," into a search engine (e.g., Google, Bing) and receives two types of results. The lower, left-hand portion of the page shows unsponsored search results, whose ranking reflects the relevance assigned to these different results by a search algorithm. On the top and right-hand side, sponsored search results appear. Whereas the display of unsponsored search results is free of charge, advertisers pay for each click on their ads that appear among the sponsored search results.

The rankings and prices paid per click depend on keyword auctions, which are generalized, second-price, sealed-bid auctions (Edelman, Ostrovsky, and Schwarz 2007; Varian 2007). In the auctions, advertisers submit bids for a specific keyword by stating their maximum willingness to pay for each click. The search engine provider then weighs the submitted bids according to the ad's quality, which it measures using a proprietary quality score (QS), and displays the sponsored search results in decreasing order of weighted bids (Abou Nabout and Skiera 2012; Jerath et al. 2011; Katona and Sarvary 2010). If a consumer clicks on an ad, the advertiser pays the search engine provider an amount equal to the next highest weighted bid divided by its own QS. The consumer who clicked on the ad gets redirected to the advertiser's website to place an order or request a sales quote—both cases of potential conversions.

Previous research has devised solutions for optimal bidding in SEA in those cases where the advertiser is able to obtain estimates of the following metrics (Abou Nabout et al. 2014; Abou Nabout et al. 2012; Skiera and Abou Nabout 2013)[1]:

- profit contribution per keyword (i.e., the revenue per keyword times the profit margin);
- conversion rate per keyword (i.e., how many of those who click on an ad finally convert into a customer);
- percentage increase in prices per click (i.e., how strongly prices per click increase within better ranks);
- percentage increase in clickthrough rates (i.e., how strongly clickthrough rates (= number of clicks / number of searches) increase within better ranks).

---

[1] We acknowledge that the bidding decision model was already presented in two dissertations at Goethe University Frankfurt, namely Stepanchuk (2010) and Gerstmeier (2011).

The intuition behind the optimal bid presented in Abou Nabout et al. (2012), Skiera and Abou Nabout (2013), and Abou Nabout et al. (2014) is as follows: The higher the average profit contribution per keyword, the higher the amount that the advertiser will be able to spend on the acquisition of new customers. A higher average conversion rate per keyword indicates a more successful SEA campaign, allowing the advertiser to increase the bids. However, high percentage increases in prices per click mean that prices diminish substantially within ranks, which makes better ranks less attractive. As a consequence, the advertiser should place lower bids for keywords with high percentage increases in prices per click. Finally, if the percentage increase in clickthrough rates is high, then the number of clicks diminishes substantially within ranks, which makes better ranks more attractive. As a result, higher bids should be placed for keywords with high percentage increases in clickthrough rates.

Let's consider the following numerical example to illustrate the percentage increases in prices per click and clickthrough rates (see Table 1).

Insert Table 1 about here

In the numerical example in Table 1, only five ranks are displayed. Rank 5 costs the advertiser 1.00€. With a percentage increase in prices per click of 20%, rank 4 costs 1.20€, rank 3 costs 1.44€, rank 2 costs 1.73€, and rank 1 costs 2.07€. The percentage increase in clickthrough rates is 40% in the numerical example, with a clickthrough rate at rank 5 of 1.00%. As shown in Table 1, the clickthrough rate then increases to 1.96% for rank 3 and to 3.84% for rank 1. Given a profit contribution of 100€ and a conversion rate of 1%, the optimal bid according to Abou Nabout et al. (2012), Skiera and Abou Nabout (2013), and Abou Nabout et al. (2014) is:

$$Bid^* = \frac{100 \cdot 0.01 \cdot \ln(0.40 + 1)}{\ln(0.20 + 1) + \ln(0.40 + 1)} = 2.58€ \qquad (1)$$

However, advertisers setting up SEA campaigns for the first time typically lack data and consequently good estimates of the above metrics (Abhishek and Hosanagar 2013)—often referred to as the cold-start problem (Kim and Srivastava 2007). Thus, it is almost impossible to predict which rank maximizes the keyword's profit after acquisition costs. The aim of this article therefore is to solve the problem of bidding on keywords in newly set-up SEA campaigns.

3

The approach presented in this article thereby overcomes two common problems in SEA: limited variation in bids and endogeneity of bids (Abhishek and Hosanagar 2013). In typical SEA campaigns, advertisers change their bids infrequently, such that it might be difficult to identify the percentage increases from advertisers' individual keyword data. In addition, potential endogeneity of bids might be a problem because bids for a particular keyword might be correlated with random shocks (e.g., a sunny weekend). The use of Google's Keyword Planner (Goldfarb and Tucker 2011) overcomes both problems as it is independent of individual advertiser behavior. Collecting this data, however, is costly and time-consuming. Thus, we additionally suggest two proxy approaches that support advertisers in making good bidding decisions even when they cannot afford to collect the required data.

The remainder of the article is organized as follows: After a review of previous research and an outline of the article's contribution, we present a novel approach for bidding on newly set-up SEA campaigns (exact approach). Next, we present two proxy approaches that provide advertisers with guidance in case they cannot afford to implement the exact approach. We benchmark the two proxy approaches against the exact approach and conclude with a summary of the results and managerial implications.

Previous Research and Contribution

Recent work in SEA focuses on the question of how to determine optimal bids (Abou Nabout et al. 2014; Abou Nabout et al. 2012; Selçuk and Özlük 2013; Skiera and Abou Nabout 2013; Yang and Ghose 2010; Yao and Mela 2011). Selçuk and Özlük (2013) develop a model that minimizes costs for a certain number of impressions and clicks. Skiera and Abou Nabout (2013), on the other hand, implement a model already presented in Abou Nabout et al. (2012). Here, bids aim at maximizing a keyword's profit after acquisition costs and Skiera and Abou Nabout (2013) establish the superiority of such bids by comparing profits from optimal bidding to profits from some unknown bidding behavior by the advertiser using a large-scale field experiment. Abou Nabout et al. (2012) use this bidding decision model to analyze the performance of fee-based compensation plans in SEA and recommend compensation plans that rely on the idea of sharing profit.

For keywords in newly set-up SEA campaigns, however, advertisers lack the information needed to calculate optimal bids according to Abou Nabout et al. (2012), Skiera and Abou Nabout (2013), and Abou Nabout et al. (2014). We thus suggest that advertisers collect data from the Google Keyword Planner to obtain rather precise estimates of the percentage increases in prices per click and clickthrough rates, which are

needed to calculate these optimal bids (see Figure 1). We call this approach the **exact approach** of calculating bids. Together with the profit contribution per conversion and the conversion rate, the advertiser might then set bids that maximize profit. The exact approach additionally overcomes two common problems in SEA, limited variation in bids and endogeneity of bids (Abhishek and Hosanagar 2013), because it does not use individual advertiser data.

However, collecting the Keyword Planner data is costly and time-consuming. In case advertisers cannot afford to collect the required data, we suggest two proxy approaches (see Figure 1) and evaluate their performance in a simulation study:

1. **Proxy 1** is based on percentage increases calculated per keyword type.
2. **Proxy 2** assumes that the ratio of the percentage increases equals 50%.

Insert Figure 1 about here

Description of Exact Approach

The Google Keyword Planner (https://adwords.google.com/KeywordPlanner)[2] provides traffic and cost estimates for new keywords before advertisers add them to their campaign. When advertisers enter a keyword into the Keyword Planner, they obtain estimates for the keyword's search volume, the expected price per click (depending on the provided maximum bid)[3], the expected daily SEA costs alongside the expected average rank, and the number of resulting clicks.[4]

Step 1: Data Collection

In order to calculate the percentage increases in prices per click and clickthrough rates, advertisers will need to collect information about the prices per click and clickthrough rates for different ranks using the Google Keyword Planner (Goldfarb and Tucker 2011).

This data is specifically suited for estimating percentage increases in prices per click and clickthrough rates because it overcomes two common problems that would result from the use of historical advertiser data (when campaigns are already running and not newly set-up): limited variation in bids and endogeneity of bids (Abhishek and Hosanagar

---

[2] The service was previously called traffic estimator and the data in this article was collected during the time the service was known as such. The same data might today be collected using the Keyword Planner.
[3] The prices per click are typically displayed in the advertiser's currency provided when first creating their Google AdWords account.
[4] For details about these metrics see: https://support.google.com/adwords/answer/3022575?hl=en.

2013). In typical SEA campaigns, advertisers change their bids infrequently, such that it becomes difficult to identify the percentage increases from advertisers' individual keyword data. In addition, potential endogeneity of bids might be a problem. Imagine, for instance, an advertiser who increases the bid for a keyword in response to a random increase in demand, e.g., on a sunny weekend. It is very likely that the bids for a particular keyword are correlated with these random shocks, resulting in potential endogeneity. The use of Google's Keyword Planner (Goldfarb and Tucker 2011) overcomes both problems as it is independent of individual advertiser behavior.

Now, imagine a UK-based advertiser who wants to know how strongly prices per click and clickthrough rates increase within better ranks for the keyword "cruise vacation." The advertiser goes to https://adwords.google.com/KeywordPlanner, types in "cruise vacation" and provides a maximum bid of 5€. Figure 2 shows the output that the advertiser is provided with by Google. With a maximum bid of 5€, the advertiser will need to pay 2.17€ (Avg. CPC in Figure 2) and is assigned rank 1.83 (Avg. Pos. in Figure 2) in the sponsored search results.[5] The clickthrough rate is estimated to be 1.4% (CTR in Figure 2) for rank 1.83, which will result in about seven clicks per day (daily number of searches equals 510). The advertiser will learn about the prices per click and clickthrough rates at different ranks while repeating this task for different maximum bids (e.g., 4€, 3€, 2€, etc.).

Insert Figure 2 about here

Unfortunately, SEA campaigns often contain thousands of keywords (Abou Nabout et al. 2014). These might differ substantially in their prices per click for rank 1. As a consequence, starting with a maximum bid of 5€ and decreasing it gradually by 1€ might be suitable for some keywords, which may lead to enough variation to learn about prices per click and clickthrough rates for different ranks. For very expensive or inexpensive keywords, however, this approach will not generate enough variation in ranks to estimate percentage increases in prices per click and clickthrough rates. Evaluating each keyword manually would be extremely cumbersome and time-consuming.

In order to automate the task of retrieving information for different ranks, advertisers might use the developer service for the Google Keyword Planner

---

[5] Google only provides advertisers with information about the daily average rank (e.g., rank is 2.8) so that the advertiser will not know exactly whether their advertisement will be displayed at the top of the screen or at the right-hand side.

(https://developers.google.com/adwords/api/docs/reference/v201302/TrafficEstimatorServi ce)[6]. In order to ensure that enough variation in ranks is obtained, we suggest proceeding as follows:

- **Step i:** submit maximum bid (here 5€was selected)[7];
- **Step ii:** check whether submitted bid results in rank better than rank 2;
- **Step iii:**
    - **No:** increase maximum bid by 1€and continue with Step ii;
    - **Yes:** continue with Step iv;
- **Step iv:** decrease bid gradually by 0.10€until a bid of 0.10€is reached.

Based on the collected data, the advertiser might then calculate the corresponding percentage increases in prices per click and clickthrough rates.

Step 2: Calculation of Percentage Increases

According to Skiera and Abou Nabout (2013), these percentage increase in prices per click might be estimated as follows:

$$\ln(Bid_k) = \alpha_k + \beta_k \cdot Rank_k + \varepsilon_k, \tag{2}$$

where *Bid$_k$* denotes the bid of keyword *k* and *Rank$_k$* is the rank of keyword *k*. The multiplier reflecting the increase in prices per click for keyword *k*, δ_k, then equals $1/\exp(\beta_k)$ and the percentage increase in prices per click is $\delta_k - 1$.

Similarly, the advertiser might estimate the percentage increase in clickthrough rates using:

$$\ln(CTR_k) = \gamma_k + \phi_k \cdot Rank_k + \upsilon_k, \tag{3}$$

where *CTR$_k$* corresponds to the clickthrough rate of keyword *k*. The multiplier reflecting the increase in clickthrough rates for keyword *k*, ξ_k, then equals $1/\exp(\phi_k)$ and the percentage increase in clickthrough rates is $\xi_k - 1$.

Please note that both multipliers reflecting the increase in prices per click and CTR are typically larger than 1. For example, a percentage increase $\xi_k - 1$ of 50% indicates a

---

[6] According to previous conventions, this service is still called the traffic estimator service.
[7] Research shows that prices per click differ substantially with respect to keyword type (Ghose and Yang 2009). Generic keywords are typically very competitive and more expensive than branded or retailer keywords. The advertiser might thus categorize keywords according to their keyword type (usually performed when creating ad groups anyway). Next, the advertiser might manually obtain the prices per click at rank 1 for three representatives of each category through the Keyword Planner tool. The average price per click should give the advertiser a good indication of which maximum bid to select in step i for each of the three categories.

multiplier of 150%, used to calculate the optimal bid. The same is true for the percentage increase $\delta_k - 1$.

Step 3: Determination of Optimal Bid

The optimal bid according to Abou Nabout et al. (2012), Skiera and Abou Nabout (2013), and Abou Nabout et al. (2014) maximizes the profit after acquisition costs π_k for a specific keyword k:

$$\underset{Bid_k}{\text{Maximize}} \; \pi_k \left(Bid_k \right) = \left(PC_k - g_k \left(Bid_k \right) \right) \cdot S_k \left(Bid_k \right), \qquad (4)$$

$$\text{subject to } 0 \leq Bid_k \leq Bid_k^1. \qquad (5)$$

The advertiser's profit is the difference between the profit contribution per conversion, PC_k, and the acquisition costs per conversion, g_k, multiplied by the number of conversions, S_k. The closed-form solution for the optimal bid Bid_k^* then is:

$$Bid_k^* = \begin{cases} \dfrac{PC_k \cdot CR_k \cdot \xi_k'}{\delta_k' + \xi_k'}, & \text{if } Bid_k^* < Bid_k^1, \\[2mm] Bid_k^1, & \text{if } Bid_k^* \geq Bid_k^1, \end{cases} \qquad (6)$$

where PC_k denotes the average profit contribution per conversion, CR_k corresponds to the average conversion rate per conversion, and δ_k' and ξ_k' are logarithms of the multipliers in prices per click, δ_k, and clickthrough rates, ξ_k, such that $\delta_k' = \ln(\delta_k)$, and $\xi_k' = \ln(\xi_k)$. In case the optimal bid, Bid_k^*, is higher than the bid required for rank 1, Bid_k^1, the optimal bid is set to the bid at rank 1.

Using the Google Keyword Planner data, the advertiser will thus be able to calculate the percentage increases in prices per click and clickthrough rates. In order to estimate the optimal bid, the advertiser will additionally need to retrieve estimates for the profit contribution per conversion, which might be the average profit contribution earned from a single purchase or a customer lifetime value. For the average conversion rate per keyword, the advertiser might rely on the conversion rate for other online activities (e.g., affiliate marketing) or assume that the average conversion rate is 1%.[8]

---

[8]This suggested conversion rate of 1% is based on a Google Analytics Benchmarking Newsletter from July 2nd, 2011 and applies to many Western European countries as well as the U.S.

Drawback of Exact Approach

While the exact approach allows the advertiser to calculate optimal bids for each single keyword in the campaign, a major drawback of the approach is that the advertiser will either need to collect keyword information manually, which is very time-consuming and cumbersome, or invest in the development of software that allows the advertiser to connect with Google's developer service. Thus, an important question is how advertisers who are not able to afford one of the two solutions can still make good decisions when bidding on newly set-up SEA campaigns.

Description of Proxy Approaches

Aim

The aim of the two proxy approaches is to obtain reliable estimates for the percentage increases in prices per click and clickthrough rates for different types of keywords such that advertisers who cannot afford to collect the required data can still make good bidding decisions.

Proxy 1

The first proxy is based on an analysis of the percentage increases of 321 keywords that differ in keyword type, number of characters, number of words, degree of competition[9], industry, etc. The aim of the analysis is to establish logarithms of "typical" multipliers reflecting the increase in prices per click and clickthrough rates for different types of keywords, ξ'_{type} and δ'_{type}:

$$Bid_k^{Proxy1} = \begin{cases} \dfrac{PC_k \cdot CR_k \cdot \xi'_{type}}{\delta'_{type} + \xi'_{type}}, & \text{if } Bid_k^{Proxy1} < Bid_k^1, \\ Bid_k^1, & \text{if } Bid_k^{Proxy1} \geq Bid_k^1, \end{cases} \tag{7}$$

In the past, researchers have come up with different categorizations of keywords. In this article, we compare two different categorizations, the first of which is the one by Ghose and Yang (2009) and the second is the one by Broder (2002). While Ghose and Yang (2009) categorize keywords into being branded, generic or retailer keywords, Broder (2002) thinks of keywords as either being transactional, navigational or informational. We

---

[9] Google denotes the degree of competition as the number of advertisers that showed on each keyword relative to all keywords across Google. For more details, please see: https://support.google.com/adwords/answer/3022575?hl=en.

compare both categorizations in order to find the one that is more suitable for predicting logarithms of "typical" multipliers.

Derivation of Proxy 1

The data contain keyword information on prices per click and clickthrough rates at different ranks in the sponsored search results at one point in time for 321 keywords from nine different industries (airlines, automotive, banking, direct banking, drugstores, energy, insurance, telecommunications, and travel). The keywords included in the data were selected from a German price index published by explido Web Marketing.[10] They either belong to an industry's most important sponsored search keywords (according to their number of searches), its Top 20 websites in the organic search results, or they indicate the brand names of the industry's market leaders. Figure 3 shows a histogram of the keywords in each industry.

Following Ghose and Yang (2009), the above keywords can be categorized into three keyword types: generic keywords (e.g., cruise vacation), retailer keywords (e.g., Travelocity), and branded keywords (e.g., Royal Caribbean). As shown by Ghose and Yang (2009), these keyword types typically exhibit significant differences in prices per click and clickthrough rates. As a consequence, differences are likely to occur in the percentage increases in prices per click and clickthrough rates across keyword types.

Broder (2002) categorizes keywords as being either navigational, transactional or informational. Navigational searches aim at finding one particular website, which requires the searcher to already be aware of the website and its existence. Navigational searches, thus, have only one "right" result. Transactional searches, on the other hand, aim at performing some web-mediated activity such as shopping online or downloading music or software. A transactional search is very specific and frequently contains keywords such as "book," "buy," and "download." Finally, informational searches aim at acquiring information that is assumed to be present on the Internet. Informational searches might be very broad and unspecific.

The information about prices per click and clickthrough rates at different ranks was collected from the Google Keyword Planner (Goldfarb and Tucker 2011), which provides

---

[10] For details see: http://www.explido.de/news/downloads/.

potential advertisers with estimates of the prices per click that they would expect to pay for different keywords. The Keyword Planner allows advertisers to state their maximum willingness to pay for a click by entering a corresponding bid for a specific keyword. It then returns the expected number of global and local searches for a keyword, its expected competition index, the expected rank resulting from the provided bid, the corresponding expected price per click and the expected number of clicks, and finally, the expected SEA costs.[11]

To automate the task of repeatedly entering 0.10€lower bids to attain as many different ranks as possible, a software was developed that connects to the Google developer service and collects the corresponding Keyword Planner data for bids that range from 5€to 0.10€(see description of approach in Section "Step 1: Data Collection").

Table 2 provides summary statistics of the final data set. We start with the summary statistics according to the categorization by Ghose and Yang (2009): There are 213 generic keywords in the dataset, 85 branded and 23 retailer keywords. The average number of global searches is highest for branded keywords (8,681,148). On average, retailer keywords receive the lowest number of global (1,127,515) and local (194,844) searches (in Germany), which also confirms the results from the explido price index.

Insert Table 2 about here

The degree of competition, scaled between 0 (low competition) and 1 (high competition), is highest for generic keywords (0.79), which confirms industry reports of high competition for generic keywords (CyberWyre 2011; Kim 2011). Consequently, they garner the highest prices per click (1.66€for top ranks) in this study. As expected and in line with previous research (Ganchev et al. 2007; Ghose and Yang 2009; Skiera and Abou Nabout 2013), prices per click at rank 1 to 3 (i.e., top ranks) are generally higher than prices per click at rank 4 to 6 for all three keyword types.

The average number of clicks decreases within worse ranks due to primacy effects and is highest for generic keywords (670.00 clicks for top ranks). Therefore, SEA costs are far higher for generic keywords (9,427.37€for ranks 1 to 3, and 1,030.18€for ranks 4 to 6) than for other keywords and lowest for retailer keywords (1,493.84€for ranks 1 to 3, and 4.31€for ranks 4 to 6), which receive very few clicks (304.00 clicks for top ranks) at rather low prices per click (1.19€for top ranks).

---

[11] Google denotes SEA costs as the average amount that the advertiser might spend per day for this keyword.

Finally, the percentage increases in prices per click range between 43% for generic keywords and 175% for branded keywords. The percentage increases in clickthrough rates are lowest for generic keywords (52%) and highest for branded keywords (129%).

When using the categorization by Broder (2002), we recorded 106 keywords with informational intent, 122 keywords with transactional intent and 93 keywords with navigational intent. While the number of global searches is highest for navigational keywords (7,054,256), the number of local searches is highest for informational keywords (1,099,118), which reflects that navigational searches often include URLs that are relevant worldwide and not only in Germany (often English keywords).

The degree of competition is highest for keywords with transactional intent (0.86) as they are often located at the very end of the purchase funnel. The degree of competition is lowest for navigational keywords (0.34), which suggests that advertisers rarely bid on URLs of competitor brands.

Prices per click are highest for informational and transactional keywords (1.70€ and 1.57€ for top ranks, respectively). The number of clicks for top ranks is surprisingly low for navigational keywords (298.79 clicks) compared to informational (1,085.96 clicks) and navigational keywords (358.74 clicks). The high number of clicks combined with high prices per click then results in very high SEA costs for top ranks for informational keywords (18,467.61€). But given that the user obviously knows the destination URL already when using navigational keywords, SEA costs are still pretty high for these keywords (2,695.22€).

Finally, due to their specific nature, percentage increases in prices per click and clickthrough rates are very high for navigational keywords—183% and 130%, respectively. As a consequence, prices per click and clickthrough rates decrease dramatically within worse ranks. Percentage increases in prices per click and clickthrough rates are fairly low for informational and transactional keywords, which suggests that prices per click and clickthrough rates remain fairly stable across ranks (40% and 54% for informational keywords and 52% and 50% for transactional keywords).

To predict the size of the multipliers, δ_k and ξ_k, for different types of keywords (calculated according to Section "Step 2: Calculation of Percentage Increases"), we estimate the following four regressions using ordinary least squares[12]:

Categorization according to Ghose and Yang (2009):

$$\ln\left(\delta_k\right) = a_1 + \sum_{i=1}^{8} b_{1i} \cdot Ind_{ki} + c_1 \cdot Branded_k + d_1 \cdot Retailer_k + f_1 \cdot Comp_k$$
$$+ g_1 \cdot WordCount_k + h_1 \cdot CharCount_k + i_1 \cdot Searches_k + j_1 \cdot German_k + e_{1k}, \tag{8}$$

$$\ln\left(\xi_k\right) = a_2 + \sum_{i=1}^{8} b_{2i} \cdot Ind_{ki} + c_2 \cdot Branded_k + d_2 \cdot Retailer_k + f_2 \cdot Comp_k$$
$$+ g_2 \cdot WordCount_k + h_2 \cdot CharCount_k + i_2 \cdot Searches_k + j_2 \cdot German_k + e_{2k}, \tag{9}$$

Categorization according to Broder 2002:

$$\ln\left(\delta_k\right) = a_3 + \sum_{i=1}^{8} b_{3i} \cdot Ind_{ki} + c_3 \cdot Informational_k + d_3 \cdot Transactional_k + f_3 \cdot Comp_k$$
$$+ g_3 \cdot WordCount_k + h_3 \cdot CharCount_k + i_3 \cdot Searches_k + j_3 \cdot German_k + e_{3k}, \tag{10}$$

$$\ln\left(\xi_k\right) = a_4 + \sum_{i=1}^{8} b_{4i} \cdot Ind_{ki} + c_4 \cdot Informational_k + d_4 \cdot Transactional_k + f_4 \cdot Comp_k$$
$$+ g_4 \cdot WordCount_k + h_4 \cdot CharCount_k + i_4 \cdot Searches_k + j_4 \cdot German_k + e_{4k}, \tag{11}$$

where:

$\ln(\delta_k)$:	logarithm of the multiplier reflecting the increase in prices per click for keyword k,
$\ln(\xi_k)$:	logarithm of the multiplier reflecting the increase in clickthrough rates for keyword k,
Ind_i:	dummy variable indicating industry i,
Branded:	dummy variable indicating branded (vs. generic) keywords,
Retailer:	dummy variable indicating retailer (vs. generic) keywords,
Informational:	dummy variable indicating informational (vs. navigational) keywords,
Transactional:	dummy variable indicating transactional (vs. navigational) keywords,
Comp:	degree of competition,
WordCount:	number of words in a keyword,
CharCount:	number of characters in a keyword,

---

[12] Please note that SUR estimation will be equivalent to OLS as the independent variables are identical across equations.

Searches: number of searches for a keyword,

Please note that we decided to run a regression on the logarithms of the multipliers that reflect the increase in prices per click and clickthrough rates, δ_k and ξ_k, instead of the percentage increases directly, in order to make the dependent variable more suitable to the assumption of linear regression. Table 3 shows their determinants, which enable advertisers to calculate the logarithms of "typical" multipliers for different types of keywords. Both categorizations share the fact that the keyword type (whether branded/ retailer or informational/transactional) has a significant impact on the size of the multiplier reflecting the increase in prices per click. For instance, the multiplier reflecting the increase in prices per click for branded keywords is, on average, 0.36 percentage points higher than the one for generic keywords. For informational keywords, the multiplier reflecting the increase in prices per click is, on average, 0.43 percentage points lower than the one for navigational keywords.

Insert Table 3 about here

Interestingly, multipliers reflecting the increase in prices per click do not differ across industries, which is also a stable result across categorizations. But the degree of competition (-0.60 percentage points) has a significant impact on the size of the multiplier reflecting the increase in prices per click. Finally, the fact that a keyword is German significantly increases the size of the aforementioned multiplier (0.06 to 0.07 percentage points higher for German keywords).

The main difference between both categorizations is that the regression that uses the categorization by Ghose and Yang (2009) explains 62% of the variance in the multiplier reflecting the increase in prices per click, while the one by Broder (2002) explains 64% of it.

Regarding the multiplier reflecting the increase in clickthrough rates, we find that generic keywords are no different from branded or retailer keywords. However, multipliers reflecting the increase in clickthrough rates are somewhat lower for informational compared to navigational keywords (-0.12 percentage points); navigational keywords do not differ from transactional keywords.

In contrast to the multipliers reflecting the increase in prices per click, multipliers reflecting the increase in clickthrough rates differ across industries: They are lower in the

airline (-0.12 to -0.16 percentage points), online banking (-0.12 to -0.15 percentage points), and travel industry (-0.10 to -0.12 percentage points) compared to the insurance industry. In addition, the degree of competition (i.e., how many ads are shown) has a large and significant impact on the multiplier reflecting the increase in clickthrough rates (-0.62 to -0.64 percentage points) following both categorizations. The number of searches also has a significant, but very small effect on the multiplier reflecting the increase in clickthrough rates, but only according to the categorization by Broder (2002).

Again, the main difference between both categorizations is that the regression that uses the categorization by Ghose and Yang (2009) explains 57% of the variance in multipliers reflecting the increase in clickthrough rates and the one by Broder (2002) explains 59% of it.

In order to make sure that no multicollinearity problem is present, we report the correlations between all variables in Appendix A. As expected, the two categorizations (Ghose and Yang 2009 and Broder 2002) are rather highly correlated (between -0.91 and 0.68), but not identical. In addition, the VIF values for all variables are below 4.10, which suggests that we do not have a multicollinearity problem. Finally, we include a plot of the residuals in Appendix B as an additional indicator that the residuals seem to be well behaved.

Drawback of Proxy 1

Even though proxy 1 allows the advertiser to treat keywords differently, the major drawback of proxy 1 is that the coefficients for predicting the multipliers (Table 3) might change over time. In addition, we only looked at 321 keywords from nine industries, so it remains unclear whether these coefficients really generalize to other settings (particularly, for the multipliers reflecting the increase in clickthrough rates).

Proxy 2

Because of the above drawbacks, we suggest using proxy 2, which supports advertisers in submitting good bids even when they do not want to calculate percentage increases for specific keywords (exact approach) or keyword types (proxy 1).

Derivation of Proxy 2

Skiera and Abou Nabout (2013) suggest a heuristic that is based on the assumption that the percentage increases in prices per click and clickthrough rates are often very closely aligned. If this assumption holds true, then a good guess for the optimal bid is 50% times

the profit contribution per conversion times the average conversion rate per keyword. The authors justify the 50% heuristic by looking at the data of four different advertisers (mobile phones, fashion, industrial goods, and travel) and the average percentage increase in prices per click and clickthrough rates. They find that these average percentage increases are indeed very close together.

In contrast to their analysis, we use percentage increases that are calculated based on macro-level data from the Google Keyword Planner. The advantage of this data is that it overcomes the two common problems in SEA, which individual advertiser data might suffer from: limited variation in bids and endogeneity of bids. In addition, we extend their analysis to more industries (airlines, automotive, banking, direct banking, drugstores, energy, insurance, telecommunications, and travel) and cover 321 keywords. Our analysis confirms that the two percentage increases are indeed fairly close together (see Figure 4).

Insert Figure 4 about here

To further verify this finding, the intercept of the regression $\delta_k - 1 = a + \beta \cdot (\xi_k - 1)$ needs to be equal to zero and the coefficient for the percentage increase in clickthrough rates needs to be equal to one. By estimating a regression using our data with the intercept restricted to zero, we cannot reject the hypothesis that the coefficient for the percentage increase in clickthrough rates is equal to one, thus providing further support for the assumption that a good guess for the optimal bid is indeed 50% times the profit contribution per conversion times the average conversion rate per keyword. Based on this empirical finding, proxy 2 equals:

$$Bid_k^{proxy2} = PC_k \cdot CR_k \cdot 50\%. \tag{12}$$

Drawback of Proxy 2

The major drawback of proxy 2 is that it might be overly simplistic, with keywords differing only in their bid because of the profit contribution and the conversion rate that they generate. Thus, we compare all three approaches to calculate bids for keywords in newly set-up SEA campaigns (exact approach, proxy 1 and proxy 2) in a large-scale simulation study with the exact approach being the benchmark for the two proxy approaches.

Evaluation of Approach

In order to evaluate the performance of the proxy approaches compared to the exact approach, we set up a large-scale simulation study, in which we vary the profit contribution per conversion as well as the conversion rate to calculate optimal and proxy bids. Simulating different levels of these metrics is necessary as they are not available in our data. The advantage of a simulation study is that it covers many different situations; the drawback is that not all scenarios are equally likely to occur in reality.

Design of Study

Table 4 details the design of the simulation study. For each pair of percentage increases in prices per click and clickthrough rates (N=321), we randomly draw 10 values from the uniform distributions for all six factor levels (3 x 2). For the profit contribution per conversion, we use three factor levels (high, medium, low) that are based on the article by Abou Nabout et al. (2012) and range between 10€ and 500€ For the conversion rate, we differentiate between two factor levels (high and low) that range between 0.05% and 5% and are also based on the simulation study conducted by Abou Nabout et al. (2012). In total, we simulate 19,260 different keyword scenarios to evaluate how close proxy bids are to the ones generated by the exact approach.

Insert Table 4 about here

Results

In the previous section, we derived two possible categorizations (Ghose and Yang 2009 and Broder 2002) that might be used to calculate proxy 1. In our evaluation of the different proxy approaches, we use both categorizations as well as proxy 2 to test which of the proxy approaches is best suited to calculate bids for keywords in newly set-up SEA campaigns. Table 5 shows the descriptive results for the bids calculated under the different approaches, as well as the difference between these bids.

Insert Table 5 about here

Surprisingly, the average bids are all very similar despite the very different approaches used to calculate them. The average bid under the exact approach is at 1.29€ with a minimum bid of 0.01€ and a maximum bid of 14.35€ Proxy 1, according to Ghose

and Yang (2009), generates an average bid of 1.30€ with a minimum bid of 0.02€ and a maximum bid of 13.07€ According to Border (2002), proxy 1 generates an average bid of 1.30€ with a minimum bid of 0.02€ and a maximum bid of 14.22€ The range of bids under proxy 2 is somewhat smaller, with an average bid of 1.30€, a minimum bid of 0.03€ and a maximum bid of 11.98€

The largest difference between the exact bid and the proxy bid is generated by proxy 1 (Ghose and Yang 2009) with a deviation from -4.45€ to 2.99€ Surprisingly, the smallest difference between the exact bid and the proxy bid is generated by proxy 2 with deviations from -3.61€ to 2.85€ The deviations for proxy 1 (Broder 2002) range between -4.03€ and 2.94€

To better assess the difference in bids under the different proxy approaches compared to the benchmark (i.e., the exact approach), we calculate four different accuracy measures: mean absolute error, median absolute error, mean absolute percentage error, and median absolute percentage error. The reason for not only calculating the means of these accuracy measures, but also their medians, is that the means might be overly influenced by extreme situations. Table 6 reports the results for the different proxy approaches.

Insert Table 6 about here

Surprisingly, the mean absolute error and the median absolute percentage error of proxy 2 (0.0881€ and 0.0393%) are lower than both instances of proxy 1, which have slightly higher errors (Ghose and Yang 2009: 0.1004€ and 0.0479%; Broder 2002: 0.0972€ and 0.0403%). However, the differences between all proxies are rather small. The mean (median) absolute error is never larger than 10 (2) cents and the mean (median) absolute percentage error is always below 12% (5%).

According to the median absolute error and the mean absolute percentage error, the Broder (2002) version of proxy 1 performs best. As a consequence, advertisers are advised to use proxy 1 according to Broder (2002), but not proxy 1 according to Ghose and Yang (2009). However, because proxy 2 also performs very well and is much easier to implement without requiring keyword classification (which is time-consuming and cumbersome), we suggest that advertisers use proxy 2 if they cannot afford to collect the required Google Keyword Planner data to implement the exact approach.

Summary, Conclusions, and Managerial Implications

SEA is today's most popular online advertising instrument, as evidenced by the steady increase in SEA expenditures in recent years. Previous research has created solutions for optimal bidding in SEA, but these require that advertisers obtain good estimates of several metrics (average profit contribution per conversion, average conversion rate per keyword, percentage increase in prices per click, and percentage increase in clickthrough rates) to calculate these optimal bids. The percentage increase in prices per click thereby reflects how strongly prices per click increase within better ranks and the percentage increase in clickthrough rates captures how strongly clickthrough rates (= number of clicks / number of searches) increase within better ranks.

Unfortunately, advertisers who set up SEA campaigns for the first time need to place bids on keywords, but typically lack good estimates of the above metrics. It is particularly challenging to find good estimates for the percentage increases in prices per click and clickthrough rates, as advertisers need to know how prices per click and clickthrough rates change across different ranks. As a consequence, it is fairly hard to predict which rank maximizes the keyword's profit after acquisition costs.

This article reports on a novel approach for solving the problem of bidding on keywords in newly set-up SEA campaigns. Our approach overcomes two common problems in SEA—limited variation in bids and endogeneity of bids—by using macro-level data from Google's Keyword Planner tool rather than individual advertiser data. In typical SEA campaigns, advertisers change their bids infrequently, such that it might be difficult to identify the percentage increases from advertisers' individual keyword data. In addition, potential endogeneity of bids might be a problem because bids for a particular keyword might be correlated with random shocks (e.g., a sunny weekend).

Our exact approach, however, requires advertisers to manually collect additional data from the Google Keyword Planner or to develop software in order to retrieve the corresponding data from Google's developer service. Both solutions are costly and time-consuming. Thus, we additionally present two proxy approaches that are aimed at supporting advertisers in making good bidding decisions even in cases in which they cannot afford to collect the required data.

Proxy 1 is thereby based on "typical" percentage increases for different types of keywords across different industries. It allows the advertiser to not only take differences in the profit contribution per conversion as well as the conversion rate into account, but also to consider differences in the percentage increases across different types of keywords. The

drawback of such an approach is that the coefficients established in this article to predict the percentage increases in prices per click and clickthrough rates might be subject to changes over time. In addition, we only cover nine different industries. "Typical" percentage increases in other industries might thus be very different from the ones found in this empirical study.

Thus, we also present proxy 2, which uses the empirical finding that both percentage increases are often very close together and assumes that the ratio of the percentage increases is equal to 50% (Skiera and Abou Nabout 2013). The beauty of proxy 2 lies in its simplicity: The advertiser only needs to know the average profit contribution per conversion for a keyword as well as its average conversion rate. However, its simplicity might also be its major drawback.

To understand how well the proxy approaches perform in contrast to the exact approach, we used a large-scale data set covering nine different industries as well as different keyword types (Ghose and Yang 2009: branded/retailer/generic; Broder 2002: informational/navigational/transactional) of varying competition levels. Through this method, we benchmark the two proxy approaches against the exact approach in a simulation study. The surprising result is that all approaches seem to perform reasonably well and that the easiest approach to implement (proxy 2) even performs best in terms of the mean absolute error and the median absolute percentage error.

Based on the results of the simulation study, advertisers are advised to use proxy 1 where feasible, categorizing keywords according to Broder (2002) rather than Ghose and Yang (2009). But because proxy 2 is much easier to implement without requiring keyword classification (which is time-consuming and cumbersome), and even performs best in terms of the mean absolute error and the median absolute percentage error, we suggest that advertisers use proxy 2 if they cannot afford to collect the required Google Keyword Planner data to implement the exact approach.

References

Abhishek, V. and K. Hosanagar (2013), "Optimal Bidding in Multi-Item Multi-Slot Sponsored Search Auctions," *Operations Research*, 61 (4), 855-873.

Abou Nabout, N. and B. Skiera (2012), "Return on Quality Improvements in Search Engine Marketing," *Journal of Interactive Marketing*, 26 (3), 141-154.

Abou Nabout, N., B. Skiera, E. Gerstmeier, and T. Stepanchuk (2014), "Profitability of Different Bidding Heuristics in Search Engine Marketing," Goethe University Frankfurt, *Working Paper*.

Abou Nabout, N., B. Skiera, T. Stepanchuk, and E. Gerstmeier (2012), "An Analysis of the Profitability of Fee-Based Compensation Plans for Search Engine Marketing," *International Journal of Research in Marketing*, 29 (1), 68-80.

Broder, A. (2002), "A Taxonomy of Web Search," *SIGIR Forum*, 36 (2).

CyberWyre (2011), "Highest Paying Search Terms," http://www.cwire.org/highest-paying-search-terms/ (17 July, 2011).

Edelman, B., M. Ostrovsky, and M. Schwarz (2007), "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords," *American Economic Review*, 97 (1), 242-259.

Ganchev, K., A. Kulesza, J. Tan, R. Gabbard, Q. Liu, and M. Kearns (2007), "Empirical Price Modeling for Sponsored Search," In *Internet and Network Economics*, (eds) D. Xiaotie & G. Fan Chung, Lecture Notes in Computer Science, Berlin / Heidelberg: Springer, 541-548.

Ghose, A. and S. Yang (2009), "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets," *Management Science*, 55 (10), 1605-1622.

Goldfarb, A. and C.E. Tucker (2011), "Search Engine Advertising: Channel Substitution When Pricing Ads to Context," *Management Science*, 57 (3), 458-470.

IAB (2014), "Iab Internet Advertising Revenue Report: 2013 Full-Year Results," http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2013.pdf (10 April, 2014).

Jerath, K., L. Ma, Y. Park, and K. Srinivasan (2011), "A "Position Paradox" in Sponsored Search Auction," *Marketing Science*, 30 (4), 612-627.

Katona, Z. and M. Sarvary (2010), "The Race for Sponsored Links: Bidding Patterns for Search Advertising," *Marketing Science*, 29 (2), 199-215.

Kim, L. (2011), "The Most Expensive Keywords in Google Adwords," http://www.wordstream.com/blog/ws/2011/07/18/most-expensive-google-adwords-keywords (17 July, 2011).

Kim, Y.A. and J. Srivastava (2007), "Impact of Social Influence in E-Commerce Decision Making," In *Proceedings of the ninth international conference on Electronic commerce*, Minneapolis, MN, USA: ACM.

Selçuk, B. and Ö. Özlük (2013), "Interfaces with Other Disciplines: Optimal Keyword Bidding in Search-Based Advertising with Target Exposure Levels," *European Journal of Operational Research*, 226 (1), 163-172.

Skiera, B. and N. Abou Nabout (2013), "Prosad: A Bidding Decision Support System for Profit Optimizing Search Engine Advertising," *Marketing Science*, 32 (2), 213-220.

Varian, H.R. (2007), "Position Auctions," *International Journal of Industrial Organization*, 25 (6), 1163-1178.

Yang, S. and A. Ghose (2010), "Analyzing the Relationship between Organic and Sponsored Search Advertising: Positive, Negative or Zero Interdependence?," *Marketing Science*, 29 (4), 602-623.

Yao, S. and C.F. Mela (2008), *Sponsored Search Auctions: Research Opportunities in Marketing*, 3: Foundations and Trends in Marketing.

— (2011), "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30 (3), 447-468.

Figures

Figure 1
Overview of Bidding Approaches
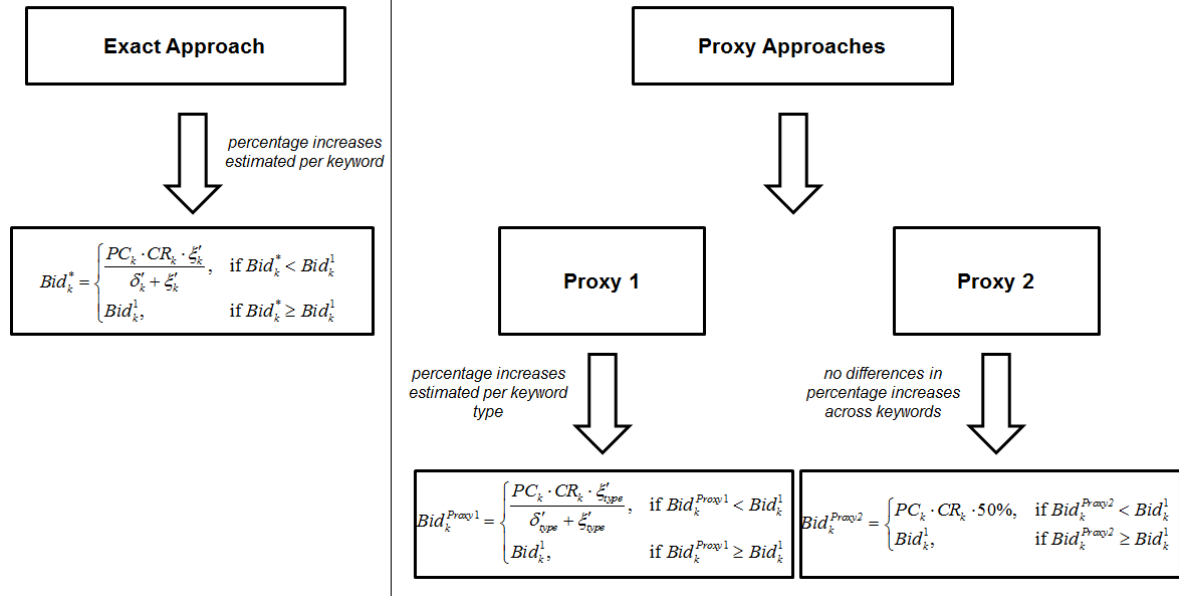
Figure 2
Google Keyword Planner Interface
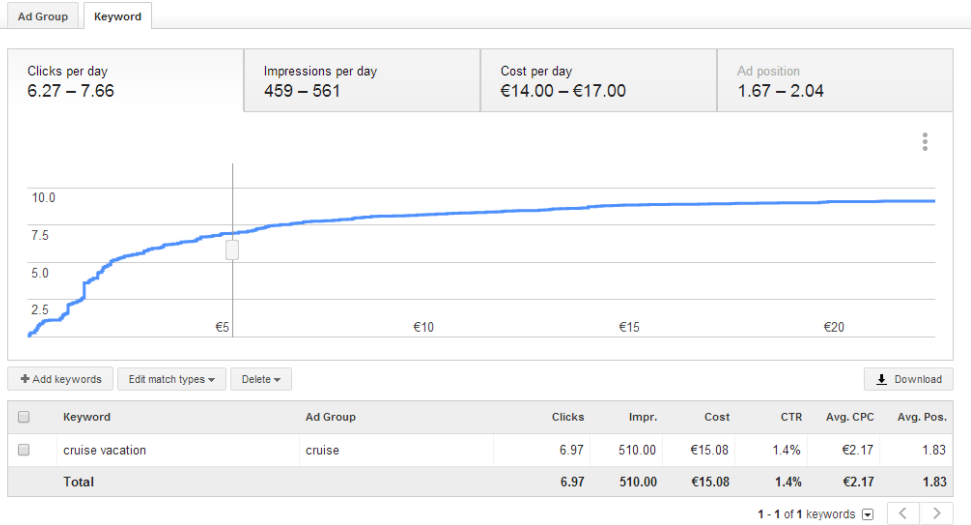
Figure 3
Number of Keywords per Industry
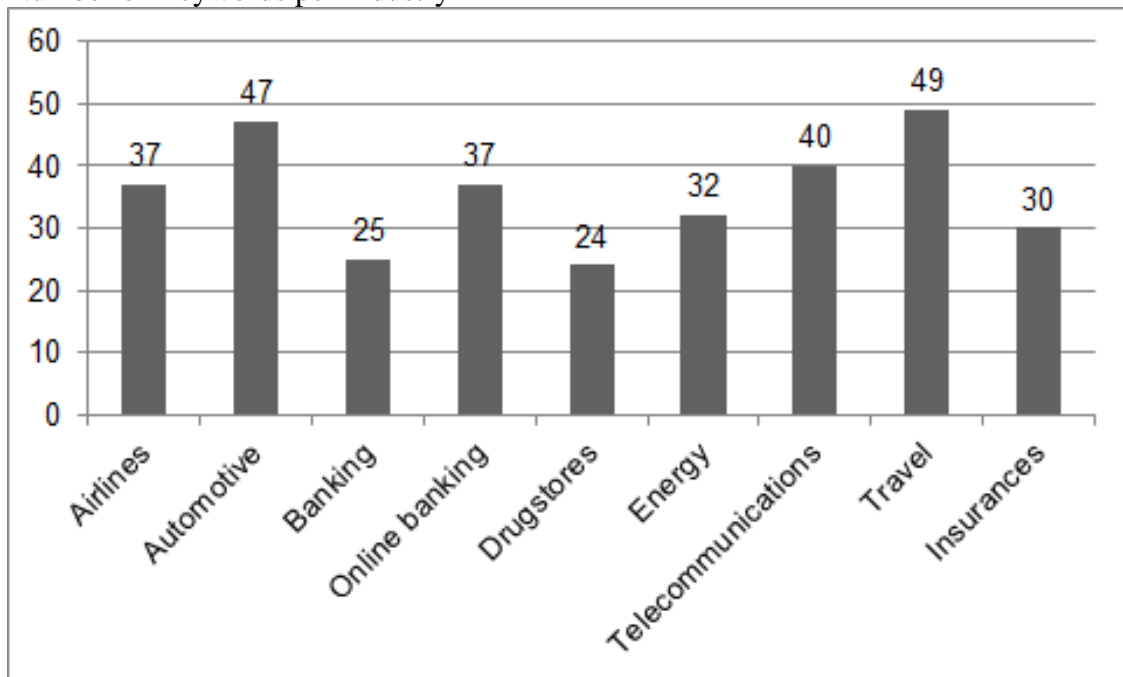
Figure 4
Comparison of Percentage Increases in Prices per Click and Clickthrough Rates

Estimated regression:
$(\delta_k\text{-}1) = 0.95 \cdot (\xi_k\text{-}1)$
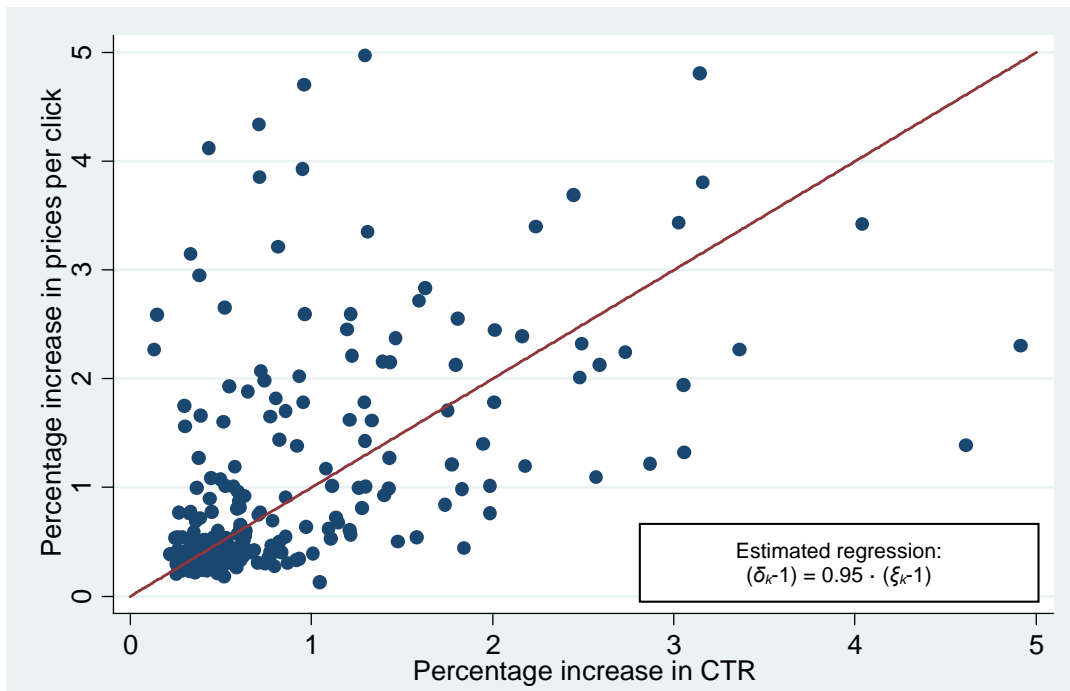
Tables

Table 1

Illustration of Percentage Increases in Prices per Click and Clickthrough Rates

	Prices per click (in €) Percentage increase in prices per click: 20%	**Clickthrough rates** (in %) Percentage increase in clickthrough rates: 40%
Rank 1	2.07	3.84
Rank 2	1.73	2.74
Rank 3	1.44	1.96
Rank 4	1.20	1.40
Rank 5	1.00	1.00

Table 2

Descriptive Statistics by Keyword Type

Average values per keyword (N=321)		Ghose & Yang (2009)			Broder (2002)		
		Generic	Branded	Retailer	Informational	Transactional	Navigational
Number of keywords		213	85	23	106	122	93
Number of observations		4,658	1,775	534	2121	2,792	2,054
Number of global searches		1,268,959	8,681,148	1,127,515	2,149,635	1,029,055	7,054,256
Number of local searches		598,740	889,871	194,844	1,099,118	206,141	762,281
Degree of competition		0.79	0.35	0.56	0.71	0.86	0.34
Price per click (in €)	*Rank 1-3*	1.66	1.14	1.19	1.70	1.57	1.16
	Rank 4-6	1.49	0.79	0.56	1.23	1.67	0.68
Number of clicks	*Rank 1-3*	670.00	355.00	304.00	1,085.96	298.79	358.74
	Rank 4-6	161.00	58.00	4.00	261.95	76.06	51.77
Search engine advertising costs (in €)	*Rank 1-3*	9,427.37	2,725.09	1,493.84	18,467.61	1,454.55	2695.22
	Rank 4-6	1,030.18	15.93	4.31	1,185.43	868.56	9.30
Percentage increase in prices per click (in %)		43	175	158	40	52	183
Percentage increase in clickthrough rates (in %)		52	129	86	54	50	130

Table 3

Determinants of the Logarithms of the Multiplier Reflecting the Increase in Prices per Click and Clickthrough Rates for Different Types of Keywords

	(1) Ghose & Yang (2009)				(2) Broder (2002)			
	$\ln(\delta_k)$		$\ln(\xi_k)$		$\ln(\delta_k)$		$\ln(\xi_k)$	
Intercept	0.67 (0.08)	***	0.88 (0.07)	***	1.08 (0.07)	***	0.99 (0.06)	***
Branded	0.36 (0.04)	***	0.06 (0.04)	n.s.	.		.	
Retailer	0.37 (0.06)	***	0.04 (0.05)	n.s.	.		.	
Informational	.		.		-0.43 (0.04)	***	-0.12 (0.04)	***
Transactional	.		.		-0.31 (0.05)	***	0.03 (0.05)	n.s.
Airlines	0.09 (0.07)	n.s.	-0.12 (0.06)	**	0.05 (0.06)	n.s.	-0.16 (0.06)	***
Automotive	0.06 (0.06)	n.s.	0.07 (0.05)	n.s.	0.06 (0.06)	n.s.	0.05 (0.05)	n.s.
Banking	0.07 (0.07)	n.s.	0.01 (0.06)	n.s.	0.06 (0.07)	n.s.	-0.00 (0.06)	n.s.
Online banking	0.03 (0.06)	n.s.	-0.12 (0.05)	**	-0.02 (0.06)	n.s.	-0.15 (0.05)	***
Drug stores	0.03 (0.07)	n.s.	-0.01 (0.06)	n.s.	0.04 (0.07)	n.s.	-0.01 (0.06)	n.s.
Energy	-0.01 (0.06)	n.s.	-0.05 (0.06)	n.s.	-0.03 (0.06)	n.s.	-0.06 (0.06)	n.s.
Telecommunications	0.03 (0.6)	n.s.	-0.07 (0.05)	n.s.	0.05 (0.06)	n.s.	-0.09 (0.06)	n.s.
Travel	0.08 (0.06)	n.s.	-0.10 (0.05)	**	0.06 (0.06)	n.s.	-0.12 (0.05)	**
Degree of competition	-0.60 (0.07)	***	-0.64 (0.06)	***	-0.60 (0.07)	***	-0.62 (0.06)	***
Number of words	0.02 (0.03)	n.s.	0.04 (0.03)	n.s.	0.01 (0.03)	n.s.	0.01 (0.03)	n.s.
Number of characters	0.00 (0.00)	n.s.	-0.00 (0.00)	n.s.	0.00 (0.00)	n.s.	-0.00 (0.00)	n.s.
Number of searches	0.00 (0.00)	n.s.	0.00 (0.00)	n.s.	0.00 (0.00)	n.s.	0.00 (0.00)	*
German	0.06 (0.03)	*	0.00 (0.03)	n.s.	0.07 (0.03)	**	0.00 (0.03)	n.s.
R^2	0.62		0.57		0.64		0.59	
F-value	33.73	***	26.96	***	36.29	***	28.84	***
Number of observations	321		321		321		321	

Note: Base keyword type is generic for (1) and navigational for (2). Base industry is the insurance industry for both categorizations. Standard errors are given in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$, n.s. $p > 0.1$.

Table 4
Simulation Study Design

Factors	Number of factor levels	Value for factor levels	
Profit contribution per customer	3	• **High:** • **Medium:** • **Low:**	$PC_k = [251€;500€]$ $PC_k = [51€;250€]$ $PC_k = [10€;50€]$
Conversion rate	2	• **High:** • **Low:**	$CR_k = [.026;.05]$ $CR_k = [.005;.025]$
Percentage increase in prices per click		*Available through data*	
Percentage increase in clickthrough rates		*(N = 321 keywords)*	
Number of keywords in one replication		321 x 3 x 2 = 1,926	
Number of replications		10	
Total number of keywords		**19,260**	

Note: The factor levels are based on Abou Nabout et al. (2012).

Table 5

Different Bids according to Exact Approach, Proxy 1, and Proxy 2

N=19,260	Mean	Standard deviation	Median	Min	Max
Exact bid in €	1.29	1.23	1.05	0.01	14.35
Proxy 1 in €(Ghose & Yang 2009)	1.30	1.27	1.04	0.02	13.07
Proxy 1 in €(Broder 2002)	1.30	1.26	1.04	0.02	14.22
Proxy 2 in €	1.30	1.24	1.04	0.03	11.98
Difference proxy 1 (Ghose & Yang 2009) and exact bid in €	0.01	0.27	-0.01	-4.45	2.99
Difference proxy 1 (Broder 2002) and exact bid in €	0.01	0.27	-0.01	-4.03	2.94
Difference proxy 2 and exact bid in €	0.01	0.22	-0.01	-3.61	2.85

Table 6
Simulation Study Results

N=19,260	Proxy 1 (Ghose & Yang 2009)	Proxy 1 (Broder 2002)	Proxy 2
Mean absolute error in €	0.1004	0.0972	**0.0881**
Median absolute error in €	0.0181	**0.0162**	0.0170
Mean absolute percentage error in %	0.1092	**0.1062**	0.1120
Median absolute percentage error in %	0.0479	0.0403	**0.0393**

Table A1
Correlation Table of All Variables

	ξ_k-1	δ_k-1	branded	retailer	generic	airlines	automotive	banking	online banking	drug stores	energy	telecomm-unications	travel	insurances	competition	word count	character count	searches	german	naviga-tional	informa-tional	transac-tional
ξ_k-1	1.00																					
δ_k-1	0.00	1.00																				
branded	0.07	-0.04	1.00																			
retailer	-0.02	0.17	-0.21	1.00																		
generic	-0.05	-0.06	-0.79	-0.31	1.00																	
airlines	-0.02	-0.02	-0.09	0.15	0.02	1.00																
automotive	0.02	-0.02	0.12	-0.09	-0.04	-0.13	1.00															
banking	0.03	-0.02	-0.09	0.02	0.04	-0.10	-0.12	1.00														
online banking	-0.03	-0.02	-0.02	-0.10	0.09	-0.12	-0.14	-0.10	1.00													
drug stores	-0.02	0.18	-0.14	0.28	0.01	-0.09	-0.11	-0.08	-0.10	1.00												
energy	-0.03	0.00	-0.05	-0.10	0.06	-0.12	-0.14	-0.10	-0.12	-0.10	1.00											
telecomm-unications	-0.03	-0.02	0.16	-0.06	-0.10	-0.13	-0.16	-0.12	-0.14	-0.11	-0.14	1.00										
travel	-0.03	-0.02	-0.08	0.07	0.05	-0.13	-0.16	-0.12	-0.14	-0.11	-0.14	-0.16	1.00									
insurances	0.11	-0.02	0.13	-0.11	-0.11	-0.13	-0.15	-0.11	-0.13	-0.10	-0.13	-0.15	-0.15	1.00								
competition	-0.11	-0.06	-0.66	-0.09	0.68	0.15	-0.04	0.05	0.01	-0.11	-0.05	-0.08	0.16	-0.10	1.00							
word count	-0.05	0.03	-0.16	-0.16	0.29	0.11	-0.11	0.02	0.07	0.02	-0.13	0.06	0.06	-0.10	0.44	1.00						
character count	-0.05	0.04	-0.44	-0.03	0.42	0.03	-0.17	0.08	0.08	0.00	-0.10	-0.10	0.08	0.12	0.57	0.64	1.00					
searches	0.30	-0.01	0.11	-0.08	-0.03	0.00	0.13	0.01	-0.12	-0.06	-0.03	0.06	0.04	-0.05	-0.20	-0.22	-0.28	1.00				
german	-0.02	-0.02	-0.06	-0.02	0.02	-0.12	-0.13	0.11	0.08	-0.07	0.08	0.04	-0.03	0.05	0.17	0.00	0.10	-0.13	1.00			
naviga-tional	0.09	0.07	0.68	0.34	-0.91	-0.01	0.02	-0.04	-0.05	-0.01	-0.04	0.02	-0.04	0.12	-0.73	-0.39	-0.47	0.07	-0.03	1.00		
informa-tional	-0.05	-0.04	-0.44	-0.18	0.56	-0.15	0.00	0.04	-0.03	0.07	0.10	-0.09	0.05	0.03	0.21	-0.28	-0.04	0.09	0.01	-0.52	1.00	
transact-tional	-0.05	-0.04	-0.30	-0.19	0.44	0.15	-0.02	0.00	0.09	-0.06	-0.05	0.07	-0.01	-0.16	0.57	0.69	0.55	-0.17	0.02	-0.57	-0.40	1.00

Appendix B

Figure B1

Histograms of the Residuals for Regressions (8)-(11)

| Ghose & Yang (2009) | Broder (2002) |

Percentage increase in prices per click

Percentage increase in clickthrough rates
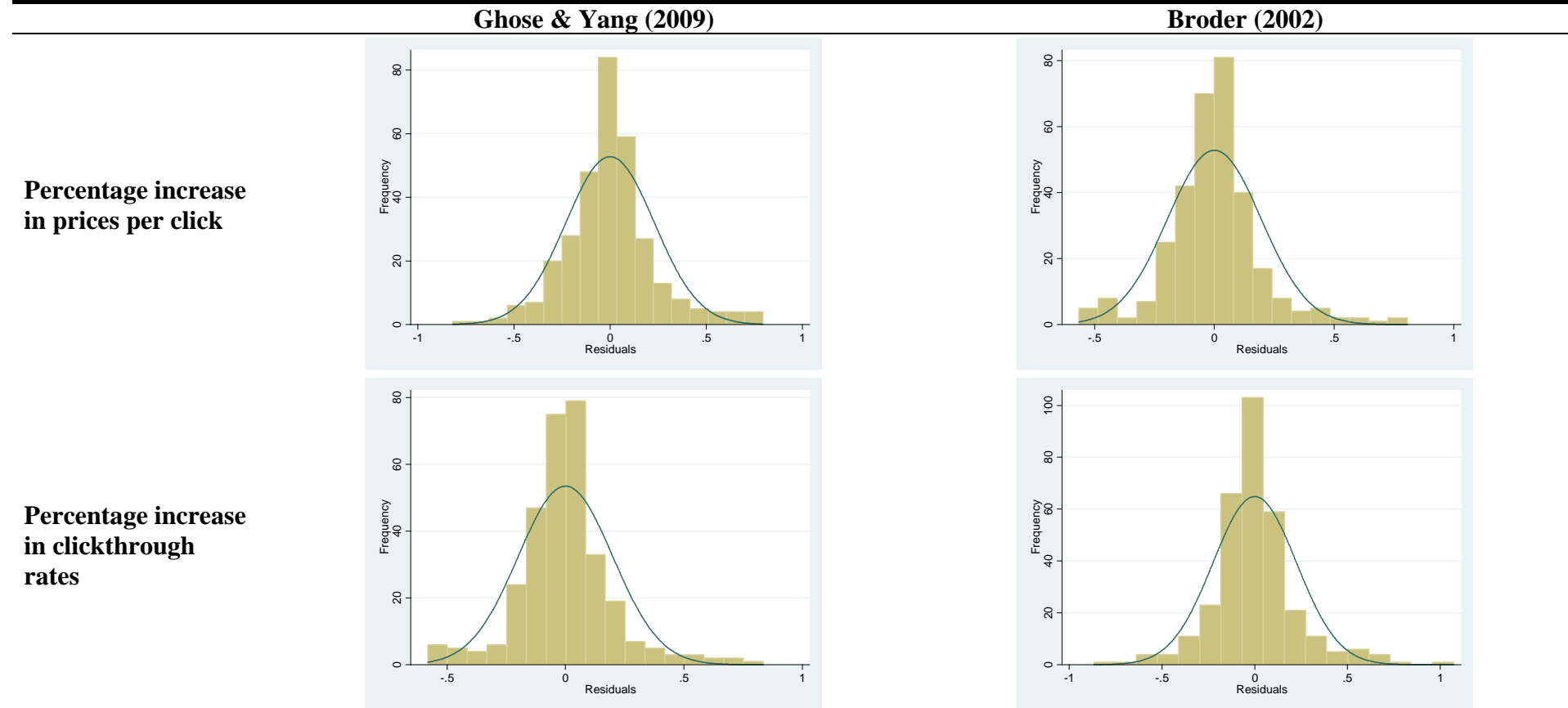
Note: The above histograms show the distribution of the residuals for each of the four regressions (8)-(11). The residuals are approximately normally distributed and thus well-behaved.