

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Essays on Incentives and Measurement of Online Marketing Efforts

Permalink

<https://escholarship.org/uc/item/5rs8m0j0>

Author

Berman, Ron

Publication Date

2014

Peer reviewed|Thesis/dissertation

Essays on Incentives and Measurement of Online Marketing Efforts

by

Ron Berman

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Business Administration

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Zsolt Katona, Chair
Professor Ganesh Iyer
Professor Shachar Kariv

Spring 2014

Essays on Incentives and Measurement of Online Marketing Efforts

Copyright 2014
by
Ron Berman

Abstract

Essays on Incentives and Measurement of Online Marketing Efforts

by

Ron Berman

Doctor of Philosophy in Business Administration

University of California, Berkeley

Assistant Professor Zsolt Katona, Chair

This dissertation contains three essays that examine different aspects of online marketing activities, the ability of marketers to measure the effectiveness of such activities, and the design of experiments to aid in this measurement.

Chapter 2 examines the impact of search engine optimization (SEO) on the competition between advertisers for organic and sponsored search results. The results show that a positive level of search engine optimization may improve the search engine's ranking quality and thus the satisfaction of its visitors. In the absence of sponsored links, the organic ranking is improved by SEO if and only if the quality provided by a website is sufficiently positively correlated with its valuation for consumers. In the presence of sponsored links, the results are accentuated and hold regardless of the correlation.

Chapter 3 examines the attribution problem faced by advertisers utilizing multiple advertising channels. In these campaigns advertisers predominantly compensate publishers based on effort (CPM) or performance (CPA) and a process known as Last-Touch attribution. Using an analytical model of an online campaign we show that CPA schemes cause moral-hazard while existence of a baseline conversion rate by consumers may create adverse selection. The analysis identifies two strategies publishers may use in equilibrium – free-riding on other publishers and exploitation of the baseline conversion rate of consumers.

Our results show that when no attribution is being used CPM compensation is more beneficial to the advertiser than CPA payment as a result of free-riding on other's efforts. When an attribution process is added to the campaign, it creates a contest between the publishers and as a result has potential to improve the advertiser's profits when no baseline exists. Specifically, we show that last-touch attribution can be beneficial for CPA campaigns when the process is not too accurate or when advertising exhibits concavity in its effects on consumers. As the process breaks down for lower noise, however, we develop an attribution method based on the Shapley value that can be beneficial under flexible campaign specifications. To resolve adverse selection created by the baseline we propose that the advertiser will require publishers to run an experiment as proof of effectiveness.

Chapter 4 discusses the type of experiments an advertiser can run online and their required sample sizes. We identify several shortcomings of the current prevailing experimental design that may result in longer experiments due to overestimation of the required sample sizes.

We discuss the use of sequential analysis in online experiments and the different goals of the experiments to make experiments more efficient. Using these techniques we show that a significant lowering of required sample sizes is achievable online.

To Racheli

Who stood by me through thick and through thin.

and

To Zsolt

Who supported and guided me with excellent ideas and careful attention.

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Introduction	1
2 The Role of Search Engine Optimization in Search Marketing	4
2.1 Overview	4
2.2 Model	7
2.3 SEO Equilibrium	9
2.4 Conclusion	13
3 Attribution in Online Advertising	14
3.1 Overview	14
3.2 Industry Description and Related Work	19
3.3 Model of Advertiser and Publishers	20
3.4 CPM vs. CPA and the Role of Attribution	23
3.5 Last-Touch and Shapley Value Attribution	26
3.6 Baselines and Experiments	30
3.7 An Application to Online Campaigns	34
3.8 Conclusion	37
4 Reducing Sample Sizes in Large Scale Online Experiments	39
4.1 Overview	39
4.2 Tests of Equality, Superiority and Selection	42
4.3 Sequential Analysis for Tests of Superiority	45
4.4 Sequential Analysis for Tests of Selection	51
4.5 Extensions	52
4.6 Application	53
4.7 Conclusion	55

Bibliography	57
A Appendix for Chapter 2	61
A.1 Proofs	61
A.2 SEO Contest with a General Error Distribution	64
A.3 SEO with Errors Observed by Players - Relation to All-Pay Auctions	64
A.4 SEO with Multiple Organic Links	70
A.5 Simultaneous SEO and Sponsored Auction	74
A.6 Heterogeneous Search Costs	75
A.7 Heterogeneous Preferences	80
B Appendix for Chapter 3	82
B.1 Proofs	82
B.2 Asymmetric Publishers	87
B.3 Estimation of Publisher Effectiveness	90

List of Figures

3.1	Converters and Conversion Rates by Publisher Exposure	15
3.2	Converters and Conversion Rates of Visitors by Publisher Exposure	18
3.3	Timing of the Campaign	23
3.4	Best Response of Player 1 Under Last-Touch Attribution	28
3.5	Profitability of Each Compensation Scheme	30
3.6	Last Touch Attribution vs. Shapley Value	37
4.1	Minimum Required Sample Sizes	45
4.2	Simulated result of SPRT testing $H_0 : p = 0.1$ vs. $H_1 : p = 0.11$	46
4.3	eqMaxSPRT Log likelihood-ratio for the Reactful experiment	55
A.1	Mixed strategy equilibrium of an all-pay auction as a function of the headstart of player 1.	67

List of Tables

3.1	Performance of Car Rental Campaign in the UK	34
3.2	Last Touch Attribution for the Car Rental Campaign	35
3.3	Logit Estimates of Publisher Effectiveness	36
4.1	Simulation results of probability of rejecting H_0 and of not stopping eqMaxSPRT	48
4.2	Simulation results of expected sample sizes $E[N]$ of eqMaxSPRT	49
4.3	Inequality MaxSPRT probability of rejecting H_0	50
4.4	Inequality MaxSPRT expected sample size	51
4.5	Simulation results for eqMaxSPRT test of Selection.	52
4.6	Results of Reactful.com experiment on CPG trial website	54

Acknowledgments

I want to deeply thank my advisor Prof. Zsolt Katona for advising me through the development of the essays in this thesis, undivided attention and many helpful discussions. I would also like to acknowledge and thank the crucial and helpful discussions with Prof. Ganesh Iyer and Prof. Shachar Kariv.

The marketing group at UC Berkeley, predominantly Professors Miguel Villas-Boas and Przemek Jeziorski, also deserve special thanks for their commitment and support.

This work has been partially supported by the California Management Review, the Joe Shoong Foundation, Benton C. Coit and Lam Research.

Chapter 2 is reprinted by permission, Ron Berman and Zsolt Katona, The Role of Search Engine Optimization in Search Marketing, Marketing Science. Forthcoming. Copyright 2013, the Institute for Operations Research and the Management Sciences, 5521 Research Park Drive, Suite 200, Catonsville, Maryland 21228 USA.

Chapter 1

Introduction

During the past 20 years the Internet has changed the way marketers interact with consumers and how consumers shop and consume content online. If at the beginning online behavior mimicked offline traditions in terms of advertising and shopping experience, the past 10 years have seen a dramatic shift in these trends towards mass customization, individual targeting, and usage of experimentation and mechanism design instead of traditional market research.

Although there are many contributing factors to this shift, a few select trends can be categorized as having substantial influence and include:

- **Computing Power** - The increase in computing power and ability to dynamically allocate resources for solving complex problems just-in-time has made previously untackled problems solvable. For example, firms today can estimate very large empirical models using high dimensional data in an efficient manner. These models are used in estimating click through rates on keyword ads, estimating consumer preferences for products and more.
- **Data Availability** - There is an increase in both the breadth and depth of information collected on each consumer today. Information such as location, purchase history, individual characteristics and more help firms react to consumer behavior in a more nuanced manner than before.
- **Use of Mechanism Design** - Mechanism design, especially through ad auctions, has allowed the creation of efficient platforms for displaying advertising and promoting products. While in the past price discrimination of products and services may have required complex allocation rules and designing menus of prices, today it is possible to run multiple auctions and let market players bid for items. The result is that the complex problem of computing optimal prices has been replaced with the easier task of creating a market and letting players reach an equilibrium. Recent analytical and empirical results have allowed the analysis of these markets to predict their behavior and improve their efficiency.

- Individual Targeting - The ability to collect and track consumer data and to compute an appropriate dynamic response is reinforced by the ability to follow consumers over multiple sites and devices. As such there is a unique one-to-one matching between collected data and an individual consumer.
- Information Exchanges - Given the distributed nature of data collection, information exchange platforms allow downstream sites that interact with consumers to either sell their data or acquire data that helps with better targeting and analysis.
- Experimentation - Because access to consumers and computing power have allowed finer and finer measurement of effects, a leading recent trend is to replace traditional market research and product design processes with faster experimentation stages where multiple product and service versions are tested and improved gradually.

This dissertation touches on three aspects related to online marketing activities - the impact of incentives and market design on agents running marketing campaigns and competing for profit, the ability to measure the performance of these agents when multiple activities take place concurrently and the design of large scale experiments that aid this measurement process. The linking theme among the essays is that standard analysis to date has typically been done in isolation, ignoring the multitude of stakeholders taking part in the process and using intuition as a guide to the interpretation of results.

Chapter 2 uses a game theoretical model to analyze the competition between websites to achieve a higher ranking on organic search engine results. This phenomenon, known as Search Engine Optimization (SEO), constitutes a substantial effort in terms of time and financial resources invested by websites today. The intuition of consumers and the search engine, however, leads to the conclusion that this type of activity may degrade search engine results and lower consumer welfare. We show that this intuition is misguided and that search results can improve with some level of SEO. Since search engines cannot exactly infer the quality of each website and its matching for its search query, the search results will be noisy and SEO can serve as a mechanism to remedy the errors. When sponsored ads are added to the mix and the websites can choose whether to compete for organic links or sponsored links, however, we show that the search engine's profit may decrease, although consumers and websites may benefit. As a result, there is a tradeoff between allowing more SEO to increase consumer welfare and the volume of visitors and the decrease in profit. Our analysis identifies this set of conditions and can serve as a guide for the design of search environments.

Chapter 3 examines the measurement and compensation problem an advertiser faces when contracting with multiple online agents to display ads. Examples of these agents may be a firm to buy sponsored ads on a search engine, a firm to perform SEO and a firm to run a display ad campaign on another platform. The chapter focuses on display ad campaigns with multiple channels that can autonomously decide on the number of ads to show and on which consumers to target. I first build a game theoretical model that allows the analysis of varied compensation schemes, and show that the current approaches for compensation and measurement of campaign performance may result in moral hazard and adverse selection.

The conclusion of the analysis is that although performance metrics of campaigns may be maximized by certain agents, the metrics themselves do not properly measure performance and can be gamed by profit maximizing agents. Using concepts from cooperative game theory, I then proceed to show how experimentation can be used to generate data that can be used to estimate the true effectiveness of different advertising channels. An application on real campaign data compares the current standard practice by firms to the proposed method and identifies substantial discrepancies in current estimates of campaign effectiveness.

Chapter 4 expands on experimentation methods used online and focuses on the required sample sizes to detect small effects of different online treatments. The analysis identifies two approaches an experimenter may use to decrease sample sizes in experiments without sacrificing the test's statistical power and validity. First, I show how the majority of current online analyses needlessly collect too much information than required or use flawed methodology. Second, I show that the sequential nature of consumer arrival to websites and thus to the experiment can be exploited to make early decision about the termination of experiments when results exceed expectations to the better or worse. The chapter includes a technical description of the techniques that can be used to achieve these goals and substantially decrease the sample sizes required in experiments.

The structure of the chapters follows that of traditional marketing literature. Each chapter begins with a detailed overview of the problem at hand and the results, followed by a model and detailed analysis. Most technical proofs are relegated to appendices, with additional results and extensions of interest appearing in an appendix as well.

Chapter 2

The Role of Search Engine Optimization in Search Marketing

2.1 Overview

Consumers using a search engine face the option of clicking organic or sponsored links. The organic links are ranked according to their relevance to the search query, while the sponsored links are allocated to advertisers through a competitive auction. Since consumers tend to trust organic links more, advertisers often try to increase their visibility in the organic list by gaming the search engine's ranking algorithm using techniques collectively known as *search engine optimization* (SEO)¹.

A notable example of the dramatic impact an SEO campaign can have is that of JCPenney, an American retailer. This retailer's organic links skyrocketed during the 2010 holiday shopping season and suddenly climbed to the top of the search results for many general keywords such as "dresses", "bedding" and "furniture".² JCPenney eventually fired their SEO contractor after finding out that they used "black hat" techniques that eventually led to a punitive response from Google. Search engine optimization is widespread in the world of online advertising; a 2010 survey of 1500 advertisers and agencies revealed that 90% of them engaged in SEO compared to 81% who purchased sponsored links.³ In the past few years, search engine optimization has grown to become a multi-billion dollar business.⁴

This chapter explores the economics of the SEO process and its effects on consumers, advertisers and search engines. Using a game theoretical model we fully characterize the incentives and tradeoffs of all players in the ecosystem. Our model consists of (i) advertisers with exogenous qualities and potentially correlated valuations for clicks, competing for the attention of consumers, (ii) a search engine that offers both organic and sponsored links

¹We focus only on "black hat" SEO which does not improve the actual relevance of the webpage to the query, but just games the ranking algorithm.

²"The Dirty Little Secrets of Search", The New York Times, Feb 12, 2011.

³"The SEMPO Annual State of Search Survey 2010".

⁴"US Interactive Marketing Forecast, 2009 to 2014", Forrester Research, July 6, 2009.

and can set minimum bids, and (iii) consumers who engage in costly search to find the highest quality site. In order to capture the effect of SEO, we model the imperfections in the algorithms used by search engines, assuming that there is a measurement error that prevents the search engine from perfectly ordering links according to quality. Advertisers can, in turn, manipulate the potentially erroneous quality observations to their advantage through SEO and improve their ranking. A key parameter of our model is the effectiveness of SEO, determining the extent to which SEO efforts by advertisers affect the organic results.

We first ask how SEO changes the organic results and whether these changes are always detrimental to consumers and high quality advertisers. The interest in this question stems from the strong stance that search engines typically take against SEO by emphasizing the potential downside on organic link quality. To justify their position, search engines typically claim that manipulation of search engine results hurts consumer satisfaction and decreases the welfare of “honest” sites. In contrast, search engines also convey the message that the auction mechanism for sponsored links ensures that the best advertisers will obtain the links of highest quality, resulting in higher social and consumer welfare. This reasoning suggests that consumers should trust sponsored links more than organic links in equilibrium, and would prefer to start searching on the sponsored side. A substantial contribution of using a sophisticated model for consumers is that we are able to derive their optimal search behavior. Contrary to claims by search engines, we find that search engines fight SEO because of the trade-off advertisers face between investing in sponsored links and investing in influencing organic rankings. Consequently, search engines may lose revenue if sites spend significant amounts on SEO activities instead of on paid links and content creation.

To approach the issue of diminished welfare from SEO, we first focus on the case where sponsored links are not available to advertisers and consumers. This base model serves as a benchmark and gives us a deeper understanding of the nature of the competition for organic links when using SEO activities. Our first result reveals that SEO can be advantageous by improving the organic ranking. In the absence of sponsored links, this only happens when advertiser quality and valuation are positively correlated. That is, if sites’ valuations for consumers are correlated with their qualities then consumers are better off with some positive level of SEO than without. By contrast, if there are sites that extract high value from visitors yet provide them with low quality then SEO is generally detrimental to consumer welfare. The SEO process essentially allows sites with a high value for consumers to correct the search engine’s imperfect ranking through a contest.

The second question we ask focuses on the full interaction between organic and sponsored links when SEO is possible. The institutional differences between the organic and sponsored lists are critical to the understanding of our model. First, advertisers usually pay for SEO services up front and the effects can take months to materialize. Bids for sponsored links, on the other hand, can be frequently adjusted depending on the ordering of the organic list. Second, SEO typically involves a lump sum payment for initial results and the variable portion of the cost tends to be convex, whereas payment for sponsored links is on a per-click basis with very little or no initial investment. Finally, there is substantial uncertainty as to the outcome of the SEO process depending on the search engine algorithms, whereas

sponsored links are allocated through a deterministic auction.

Interestingly, the presence of sponsored links accentuates the results of the base model and SEO favors the high quality advertiser regardless of the correlation between quality and valuation. The intuition is that sponsored links act as a backup for high quality advertisers in case they do not possess the top organic link. When consumers have low search costs, they will eventually find the high quality advertiser, reducing the value of the organic position for a low quality player. In equilibrium, consumers will start searching on the organic side and high quality sites will have an increased chance of acquiring the organic link as SEO becomes more effective.

Although SEO clearly favors high quality advertisers, we find that there is a strong tension between the interests of consumers and the search engine. As advertisers spend more on SEO and consumers are more likely to find what they are looking for on the organic side, they are less likely to click on revenue generating sponsored links. This tension may explain why search engines take such a strong stance against SEO, even though they favor a similar mechanism on the sponsored side. Furthermore, we obtain an important normative result that could help search engines mitigate the revenue loss due to SEO: we find that there is an optimal minimum bid the search engine can set that is decreasing in the intensity of SEO. Setting the minimum bid too high, however, could drive more advertiser dollars away from the sponsored side towards SEO.

As common the practice of SEO may be, research on the topic is scant. Many papers have focused on sponsored links and some on the interaction between the two lists. In all of these cases, however, the ranking of a website in the organic list is assumed exogenous, and the possibility of investing in SEO is ignored. On the topic of sponsored search, works such as those by Rutz and Bucklin (2011) and Ghose and Yang (2009) focus on consumer response to search advertising and the different characteristics that impact advertising efficiency. Other recent examples, such as those by Chen and He (2011), Athey and Ellison (2012) and Xu et al. (2011) analyze models that include both consumers and advertisers as active players.

A number of recent papers study the interplay between organic and sponsored lists. Katona and Sarvary (2010) show that the top organic sites may not have an incentive to bid for sponsored links. In an empirical piece, Yang and Ghose (2010) show that organic links have a positive effect on the click-through rates of paid links, potentially increasing profits. Taylor (2012), White (2009) and Xu et al. (2012) study how the incentives of the search engine to provide high quality organic results are affected by potential losses on sponsored links. The general notion is that search engines have an incentive to provide lower quality results in order to maximize revenues.

The work of Xing and Lin (2006) is the closest antecedent to our work. It defines “algorithm quality” and “algorithm robustness” to describe the search engine’s ability to accurately identify relevant websites. Their paper shows that when advertisers’ valuations for organic links is high enough, SEO is sustainable and SEO service providers can then free-ride on the search engine due to their “parasitic nature”. The relationship between advertiser qualities and valuations and the strategic nature of consumer search are not taken into account. An earlier work by Sen (2005) develops a theoretical model that examines the

optimal strategy of mixing between investing in SEO and buying ad placements. Surprisingly, the model shows that SEO should not exist as part of an equilibrium strategy.

2.2 Model

We set up a static game in which consumers search for a phrase and advertisers compete for their visits. We assume there is a monopolistic search engine that provides search results to consumers by displaying links to one of two websites. These sites can also buy sponsored links from the search engine. Whenever a consumer enters the search phrase, the search engine ranks the sites according to a scoring mechanism, and presents one organic link and one sponsored link according to the scores and bids of the sites. The incentives and characteristics of the search engine, advertisers, and consumers are described below.

Websites and Consumers

Consumers in our model seek to consume one unit of a good that can have a quality $q_i \in \{q_L, q_H\}$ with $q_H > q_L$. The good is provided by websites and can either be information, content or a physical product. Regardless of its nature, the good provides a utility of q_i to those who consume it (net of price). The two possible quality levels of q_H and q_L are common knowledge, but consumers need to search to discover the particular qualities provided by each website. When visiting the search engine, consumers see an organic link and possibly a sponsored link. In order to discover the quality provided by a site consumers need to click the links. Upon visiting a site they incur a search cost $c \geq 0$ and discover the quality of the good. Then consumers decide whether to continue the search, abandon it, or consume the good they had found. The decision on which link to start with (organic or sponsored) and the decision to continue searching depends on the expected distribution of qualities behind each link. A rational consumer will continue searching only if the expected increase in utility from visiting the next link outweighs the search cost. Once the consumer has decided to stop searching, she will consume the good with the highest net utility, possibly returning to a previously visited link.

As an example, if the consumer started searching with the organic link and found a website providing quality q_H , she has no reason to continue searching. She will consume the good yielding utility of $q_H - c$. If, on the other hand, she started searching with the organic link and found a site providing quality q_L , she will prefer to continue searching when search costs are low. If she also found q_L behind the sponsored link when continuing, she would eventually receive utility $q_L - 2c$.

The website that provides the good chosen receives an exogenously determined revenue valued at $v_i \in \{v_L, v_H\}$ with $v_H > v_L$. The total revenue of site i (net of manufacturing costs) is thus the number of consumers who consume its good multiplied by v_i . For example, in case when the good is a product sold by the advertiser, v_i can be thought of as the per unit margin of the seller. The individual site qualities q_i and valuations v_i are known by

the competing websites, but are unknown to the consumers or the search engine a priori. However, the following distribution is common knowledge: $\Pr(q_i = q_L) = \Pr(q_i = q_H) = \frac{1}{2}$, $\Pr(v_i = v_L) = \Pr(v_i = v_H) = \frac{1}{2}$, and the correlation between q_i and v_i is ρ for each site i . Both qualities and valuations are independent across sites. The sign of the correlation between the quality and valuation of a particular site could be driven by several factors in a market. For example, in a vertically differentiated market firms offering a higher quality product can charge a premium and often make a higher margin, suggesting a possibly positive correlation. However, a negative correlation is also possible between qualities and valuations due to deceptive marketing practices or interaction with other channels.

To influence their organic ranking, websites can invest SEO effort e_i at a quadratic cost of $\frac{e_i^2}{2}$. In order to win the sponsored link, websites submit per-click bids, denoted by b_i . The total payment for the sponsored link is determined in a generalized second price auction with minimum bid r , where bids are corrected for expected click-through rates (CTRs). The final payoff of site i is therefore its revenue minus the SEO investments costs and the sponsored payment.

The Search Engine

The search engine acts as an intermediary between consumers and websites. Its goal is to provide consumers with links to the highest quality websites on the organic side while making a profit through the auctioning of sponsored links. In order to rank websites, the search engine scores each website on its estimated quality using information gathered from the Internet using crawling algorithms and data mining methods. The search engine can therefore only measure quality with an error and cannot observe it directly. We model the score of each site as

$$s_i = q_i + \alpha e_i + \varepsilon_i, \quad (2.1)$$

where α is a parameter denoting the effectiveness of SEO, and ε_i is the measurement noise, distributed according to a distribution with c.d.f. F_ε and mean 0. The parameter α measures how easy it is to change one's ranking using SEO methods. That is, $1/\alpha$ influences the cost of SEO which can be controlled by several factors including the search engine. Indeed, if the search engine ignores the possibility of SEO activities, α presumably increases.

Sponsored links are awarded by the search engine in a standard click-through rate corrected second price auction with a reserve minimum bid of r . If website i has an expected click-through rate ctr_i , the search engine awards the links in order of the ranking of the scores $ctr_i \cdot b_i$, as long as they are higher than the minimum bid. When a consumer clicks on a sponsored link, the website who owns it pays the bid of the next highest bidder corrected for the click-through-rate differences. The click-through rates are a result of the endogenous consumer search process in equilibrium. They determine the payoff of the search engine, as well as influence the incentives of the advertisers to invest in SEO. Our model takes these click-through rates into account when considering the bids of advertisers for sponsored links.

Timing

At the beginning of the game the search engine publishes the minimum bid for sponsored links r . In parallel, Nature determines the quality q_i and the valuation v_i for each website given the correlation parameter ρ , but independently across sites. Then, websites decide on the amount of effort e_i to invest in SEO. The search engine then determines the scores s_i of each site, and publishes their score ranking. Following the organic ranking, sites bid for the sponsored links which are then awarded according to a CTR-corrected generalized second price auction with minimum bid r . Once both rankings have been finalized consumers initiate a search process.

Before visiting the search engine, consumers decide which link gives them the highest expected utility and start their search with that link.⁵ The consumers then decide whether to consume the good encountered or continue their search. Once the consumer has searched through all of the links, decided to stop searching and consume, payoffs are realized.

2.3 SEO Equilibrium

Organic Links Only

When the minimum bid is higher than the profit websites expect from a visitor, advertisers cannot afford sponsored links. This scenario is very common when sites provide free content to consumers and make a profit by selling advertising. It also serves as a benchmark case before analyzing the impact of sponsored links on the SEO process. The expected payoff of site i is then

$$\pi_i = v_i \cdot Pr(s_i > s_j) - \frac{e_i^2}{2} \quad (2.2)$$

To illustrate our results, we assume that the measurement error has a uniform distribution $\varepsilon_i \sim U[-\frac{\sigma}{2}, \frac{\sigma}{2}]$ with a large enough support⁶. To show the impact of SEO on consumers and the overall ranking, we use $P(\alpha) = P(\alpha; \sigma, v_1, v_2, q_1, q_2)$ to denote the efficiency of the ranking process, which is the probability of the website with the highest quality winning the organic link. Since the utility of the consumer is the quality of the consumed good, consumer welfare increases with efficiency.

Simple analysis shows that when search engine optimization is *not* possible, *i.e.*, when $\alpha = 0$, we get $P(0) < 1$ as long as $q_1 \neq q_2$ due to the noise in the ranking process. Furthermore, $P(0, \sigma)$ is decreasing in σ as higher levels of noise make the ranking less efficient. When search engine optimization becomes effective, *i.e.*, when $\alpha > 0$, websites can actively

⁵Since there might be a case with no sponsored links, we assume that consumers incur the cost c of the first search even if their favorite link does not exist. This is a technical assumption that makes the analysis cleaner. Alternative, and perhaps more realistic, assumptions lead to similar results.

⁶We need to assume $\sigma > q_1 - q_2$ for the error to have any effect. The Online Appendix illustrates equivalent results for general distribution of the errors.

influence the order of results. The following proposition summarizes how SEO affects the ranking, consumer welfare and firm profits.

Proposition 1.

1. *When $\rho = 1$, any $\alpha > 0$ which is not too large improves the efficiency of the ranking and consumer satisfaction. However, when $\rho = -1$, SEO is detrimental to consumer satisfaction. For intermediate $-1 < \rho < 1$ values, SEO can improve consumer satisfaction for some α values.*
2. *Suppose α is small. When $\rho = -1$, both sites' profits are decreasing in α . When $\rho = 1$, sites' profits are decreasing in α , except for the higher quality site, whose profits are increasing iff $v_H > 2v_L$.*

The first part demonstrates the main effect of equilibrium SEO investments on the ranking. The SEO mechanism gives both sites incentives to invest in trying to improve their ranking, but favors bidders with high valuations. Since the search engine cannot measure site qualities perfectly, this mechanism corrects some of the error when valuations are positively correlated with qualities. On the flip side, when lower quality sites have high valuations for traffic, SEO creates incentives that are not compatible with the utilities of consumers. In this latter case, the high valuation sites that are not relevant can get ahead by investing in SEO. Examples are cases of “spammer” sites that intentionally mislead consumers. Consumers gain little utility from visiting such sites, but these sites may profit from consumer visits.

Closer examination of the proof suggests that $\frac{\partial P(\alpha, \sigma)}{\partial \alpha \partial \sigma}$ is positive for small α 's. This suggests, somewhat counter-intuitively, that investments against SEO on the search engine's part complement investments in better search algorithms rather than substitute them. That is, only search engines that are already very good at estimating true qualities should fight hard against SEO. Nevertheless, as measurement error can depend on exogenous factors and can vary from keyword to keyword, it may make sense to allow higher levels of SEO in areas where the quality measurement is very noisy.

To analyze the relationship between α and advertiser profits we focus on small levels⁷ of α . As the second part of the proposition shows, the player with the lower valuation is always worse off with higher SEO effectiveness regardless of its quality. The only site that benefits from SEO is the one with a quality advantage, and only if its valuation is substantially higher than its competitor's. The intuition follows from the fact that higher levels of SEO emphasize the differences in valuations; the higher the difference the more likely that the higher valuation will win. Importantly, an advantage in valuation only helps when the site also has a higher quality, that is, spammer sites with low quality and high valuation will not benefit from SEO due to the intense competition with better sites.

⁷This relationship can be quite complex in the general case.

The Role of Sponsored Links

We now examine how the availability of sponsored advertising changes the incentive of investing in SEO and the resulting link order. Since the search engine's main source of revenue comes from sponsored links, this analysis is crucial to understanding how SEO affects the search engine's revenue. We solve the model outlined in Section 2.2 with $r < v_H$. That is, at the minimum, sites with a high valuation will be able to pay for sponsored links. When describing the intuition, we focus on the case of $r < v_L$ so that any site can afford sponsored links.

In order to determine advertisers' SEO efforts and sponsored bids, we also need to uncover where consumers start their search process. We assume that consumers always incur a small, but positive search cost. They have rational expectations and start with the link that gives them the highest probability of finding a high quality result without searching further. The following proposition summarizes our main results.

Proposition 2. *There exists a $\bar{c} > 0$, such that if $c < \bar{c}$ then*

1. *In the unique equilibrium consumers begin their search on the organic side.*
2. *If $r < v_L$ the likelihood of a high quality organic link is increasing in α for any $-1 \leq \rho \leq 1$.*
3. *If $v_L \leq r$, the likelihood of a high quality organic link is increasing in α iff ρ is high enough.*
4. *The search engine's revenue increases in α iff the likelihood of a high quality organic link decreases.*

In short, we prove that the presence of sponsored links accentuates the potential benefits of SEO on increasing the quality of the organic link. As α increases and SEO becomes more effective, the probability that the higher quality site acquires the organic link increases even if advertisers' qualities and their valuations for consumers are negatively correlated. Contrary to the commonly held view that SEO often helps low quality sites climb to the top of the organic list if they have enough resources, we find that in the presence of sponsored links, low quality sites cannot take advantage of SEO. The intuition relies on the notion that sponsored links serve as a second chance to acquire clicks from the search engine for the site that does not possess the organic link. However, as a result of exhaustive consumer search, high quality sites enjoy a distinct advantage as they are likely to be found no matter what position they are in. Low quality advertisers, on the other hand, suffer if a higher quality competitor is also on the search page. Thus a low quality site's incentive to obtain the organic link will be reduced, while high quality sites will face less competition in the SEO game and will be more likely to win it. For high quality sites, the main value of acquiring the top organic link is not merely the access to consumers. Instead, the high quality site benefits from the organic link because it does not have to pay for the access to consumers, as it would have to on the sponsored side.

In the ensuing equilibrium, high quality advertisers always spend more on SEO than their low quality competitors. Since this increases the chances of high quality organic links, we find that rational consumers start their search on the organic side. Consumers benefit from finding a high quality link as early as possible, and thus more effective SEO increases their welfare by increasing the likelihood of a high quality organic link. This fact, however, hurts the search engine whose revenues decrease when the high quality advertiser competes less for the sponsored link. The misalignment between consumer welfare and search engine profits has already been recognized by White (2009) and Taylor (2012). Our results reconfirm this tension and shed light on an interesting fact: The main danger of SEO for search engines is not the disruption of the organic list which has long-term impact on reputation and visitors, but rather decreased revenues on the sponsored side which are of a short-term nature. Often advertisers pay third parties to conduct SEO services instead of paying the search engine for sponsored links. The result from the advertiser's perspective is not much different, but the search engine is stripped of significant revenues.

The search engine has an important tool on the sponsored side – setting the minimum bid that affects what the winning advertiser pays. In the absence of SEO, an increased minimum bid directly increases the revenue from advertisers who have a valuation above the minimum bid. When SEO is possible the situation is different:

Corollary 1. *There exists an $\hat{r}(\alpha) > 0$ such that the search engine's revenue is increasing in r for $r < \hat{r}(\alpha)$ and decreasing for $\hat{r}(\alpha) < r < v_L$. When v_L is high enough then $\hat{r}(\alpha)$ is the unique optimal minimum bid which is decreasing in α .*

The inverse U-shape of the effect is a result of two opposing forces. An increasing minimum bid increases revenue directly. However, in the presence of SEO, a higher minimum bid makes sites invest more in SEO, which makes the high quality site more likely to acquire the organic link. This, in turn, will lower sponsored revenues as most of these revenues come from the case when the low quality site possesses the organic link. The combination of these two forces will make the search engine's revenue initially increase with an increased minimum bid, but begin to decrease when sites invest more in SEO. The maximal profit is reached at a lower minimum bid as SEO becomes more effective (α increases). Finally, we examine how a site's revenues are affected by SEO.

Corollary 2. *If $r < v_L$ and the two sites have different qualities, the profit of the higher quality site increases, while the profit of the lower quality site decreases in α .*

As we explained above, the possibility of using sponsored links as a backup gives an advantage to the higher quality site. The more effective search engine optimization is, the less the site has to spend to secure the top organic link. The lower quality site faces the exact opposite situation. When the two sites have the same qualities SEO only makes a difference when those qualities are low. In this case a higher α benefits the site with the higher valuation.

2.4 Conclusion

The options facing consumers when using an online search engine are highly affected by search engine marketing decisions made by website owners and the policy of the search engine. Site owners can choose to invest in SEO effort to promote their site in organic listings as well as bid for sponsored links. Search engines can choose to handicap SEO activities or to impose a minimum bid requirement. We find that, contrary to popular belief, SEO can sometimes be beneficial to consumers by giving an advantage to high quality sites, especially when the search engine's crawling algorithms do not provide an accurate ranking. Such improvement in the quality of search results will attract more consumers, yet will hurt the revenues of search engines.

Our results also provide important recommendations to advertisers. When organic links are the only option, SEO is an important tool to increase a site's visibility for advertisers who can afford to pay more. The majority of online advertisers invest in both SEO and sponsored links, and face an important dilemma as to how to allocate their budget between the two activities. Our results imply that high quality sites have an advantage as they can always use sponsored links as a backup option if their organic link does not place well. Consequently, the main value of SEO for them is to avoid the potentially hefty payments for sponsored clicks.

We believe that the economics of search engine optimization is a topic of high importance for both academics and practitioners. In this chapter we examine the basic forces of this intriguing, complex ecosystem. Given the complexity of the problem, our model has a number of limitations that could be explored by future research. First, we model SEO as a static game, whereas in reality sites invest in SEO dynamically, reacting to each other's and the search engine's actions. Our static approach limits our ability to explore how the search engine's reputation is affected in the long run. Second, we focus our attention on a single keyword with one organic and one sponsored link throughout the chapter. In reality, advertisers bid for millions of keywords to obtain sponsored links. Conducting SEO is less a fine-grained activity and may affect the ranking of a site for several different keywords. Third, we use the term SEO exclusively for black-hat type optimization, and do not model white-hat methods that directly increase quality. Finally, we assume that consumers search rationally and stick to their objectives. In reality, consumers might make mistakes or get distracted by different types of links leading to clicks that our model does not predict. Despite these limitations, we believe that this is an important step in the direction of understanding the role of search engine optimization in marketing.

Chapter 3

Attribution in Online Advertising

3.1 Overview

Digital advertising campaigns in the U.S. commanded US \$36.6 Billion in revenues during 2012 with an annual growth rate of 19.7% in the past 10 years,¹ surpassing all other media spending except broadcast TV. In many of these online campaigns advertisers choose to deliver ads through multiple publishers with different media technologies (*e.g.* Banners, Videos, etc.) that can reach overlapping target populations.

This chapter analyzes the *attribution* process that online advertisers perform to compensate publishers following a campaign in order to elicit efficient advertising. Although this process is commonly used to benchmark publisher performance, when asked about how the publishers compare, advertisers' responses range from "We don't know" to "It looks like publisher X is best, but our intuition says this is wrong." In a recent survey², for example, only 26% of advertisers claimed they were able to measure their social media advertising effectiveness while only 37% of advertisers agreed that their facebook advertising is effective. In a time when consumers shift their online attention towards social media, it is surprising to witness such low approval of its effectiveness.

To illustrate the potential difficulties in attribution from multiple publisher usage, Figure 3.1 depicts the performance of a car rental campaign exposed to more than 13 million online consumers in the UK, when the number of converters³ and conversion rates are broken down by the number of advertising publishers that consumers were exposed to. As can be seen, a large number of converters were exposed to ads by more than one publisher; it also appears that the conversion rate of consumers increases with the number of publishers they were exposed to.

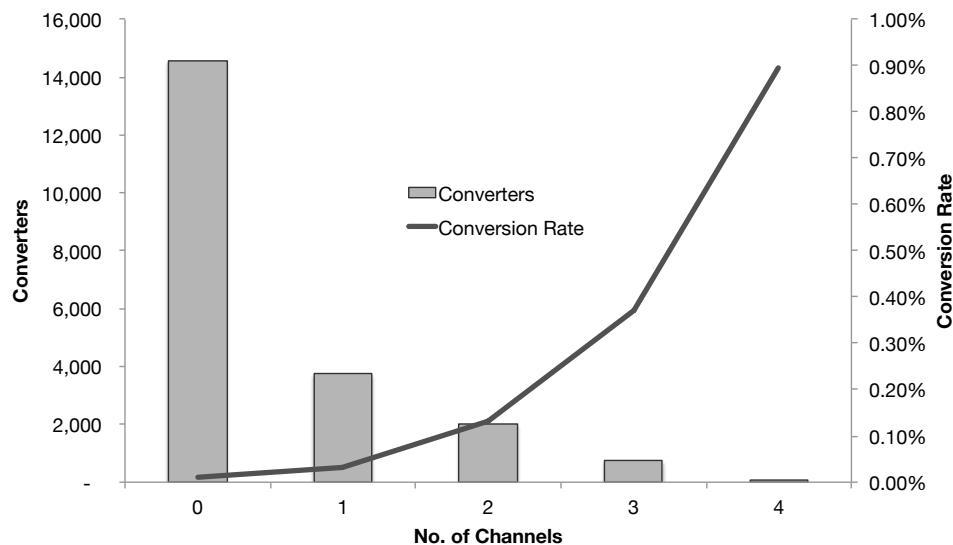
An important characteristic of such multi-publisher campaigns is that the advertisers do not know a-priori how effective each publisher may be. Such uncertainty may arise, *e.g.*, when

¹Source: 2012 IAB internet advertising revenue report.

²Source: "2013 Social Media Marketing Industry Report", www.socialmediaexaminer.com

³Converters are car renters in this campaign. Conversion rate is the rate of buyers to total consumers.

Figure 3.1: Converters and Conversion Rates by Publisher Exposure



publishers can target consumers based on prior information, when using new untested ads or because consumer visit patterns shift over time. Given that online campaigns collect detailed browsing and ad-exposure history from consumers, we ask what obstacles this uncertainty may create to the advertiser's ability to properly mount a campaign.

The first obstacle that the advertiser faces during multi-publisher campaigns is that the ads interact in a non-trivial manner to influence consumers. From the point of view of the advertiser, getting consumers to respond to advertising constitutes a team effort by the publishers. In such situations a classic result in the economics literature is that publishers can piggyback on the efforts of other publishers, thus creating moral hazard (Holmstrom, 1982). If the advertiser tries to base its decisions solely on the measured performance of the campaign, such free-riding may prevent it from correctly compensating publishers to elicit efficient advertising.

A second obstacle an advertiser may face is lack of information about the impact of advertising on different consumers. Since the decision to show ads to consumers is delegated to publishers, the advertiser does not know what factors contributed to the decision to display ads nor does it know the impact of individual ads on consumers. The publishers, on the other hand, have more information about the behavior of consumers and their past actions, especially on targeted websites with which consumers actively interact such as search-engines and social-media networks. Such asymmetry in information about ad effectiveness may create adverse selection – publishers who are ineffective will be able to display ads and claim their effectiveness is high, with the advertiser being unable to measure their true effectiveness.

To address these issues advertisers use contracts that compensate the publishers based on the data collected during a campaign. We commonly observe two types of contracts in

the industry: effort based and performance based contracts. In an effort based contract, publishers receive payment based on the number of ads they showed during a campaign. These schemes, commonly known as cost per mille (CPM) are popular for display (banner) advertising, yet their popularity is declining in favor of performance based payments.

Performance based contracts, in contrast, compensate publishers by promising them a share of the observed output of the campaign, *e.g.*, number of clicks, website visits or purchases. The popularity of these contracts, called Cost Per Action (CPA), has been on the rise, prompting the need for an attribution process whose results are used to allocate compensation. Among these methods, the popular *last-touch* method credits conversions to the publisher that was last to show an ad (“touch the consumer”) prior to conversion. The rationale behind this method follows traditional sales compensation schemes – the salesperson who “closes the deal” receives the commission.

This chapter uses analytical modeling to focus on the impact of different incentive schemes and attribution processes on the decision of publishers to show ads and the resulting profits of the advertisers. Our goal is to develop payment schemes that alleviate the effects of moral-hazard and asymmetric information and yield improved results to the advertiser. To this end Section 3.3 introduces a model of consumers, two publishers and an advertiser engaged in an advertising campaign. Consumers in our model belong to one of two segments: a baseline and a non-baseline segment. Baseline consumers are not impacted by ads yet purchase products regardless. In contrast, exposure to ads from multiple publishers has a positive impact on the purchase probabilities of non-baseline consumers. Our model allows for a flexible specification of advertising impact, including increasing returns (convex effects) and decreasing returns (concave effects) of multiple ad exposures. The publishers in our model may have private information about whether consumers belong to the baseline and make a choice regarding the number of ads to show to every consumer in each segment. The advertiser, in its turn, designs the payment scheme to be used after the campaign as well as the measurement process that will determine publisher effectiveness.

Section 3.4 uses a benchmark fixed share compensation scheme to show that moral-hazard is more detrimental to advertiser profits than using effort based compensation. We find that CPM campaigns outperform CPA campaigns for every type of conversion function and under quite general conditions. As ads from multiple publishers affect the same consumer, each publisher experiences an externality from actions by other publishers and can reduce its advertising effort, raising a question about the industry’s preference for this method. We give a possible explanation for this behavior by focusing on single publisher campaigns in which CPA may outperform CPM for convex conversion functions.

Since CPA campaigns suffer from under-provision of effort by publishers, we observe that advertisers try to make these campaigns more efficient by employing an attribution process such as last-touch. By adding this process advertisers effectively create a contest among the publishers to receive a commission, and can counteract the effects of free-riding by incentivizing publishers to increase their advertising efforts closer to efficient amounts. We include attribution in our model through a function that allocates the commission among publishers based on the publishers’ efforts and performance and has the following four requirements:

Efficiency, Symmetry, Pay-to-play and Marginality. To model Last-Touch attribution with these requirements, we notice that publishers are unable to exactly predict whether they will receive attribution for a conversion because of uncertainty about the consumer's behavior in the future. As a result, our model admits last-touch attribution as a noisy contest between the publishers that has these four properties. The magnitude of the noise serves as a measurement of the publisher's ability to predict the impact of showing an additional ad on receiving attribution and depends on the technology employed by the publisher. Our analysis of this noisy process shows that in CPA campaigns with last-touch attribution, publishers increase their equilibrium efforts and yield higher profits to the advertiser when the noise is not too small. When the attribution process is too discriminating or the conversion function too convex, however, no pure strategy equilibrium exists, and publishers are driven to overexert effort. Cases of low noise level can occur, for example, when publishers are sophisticated and can predict future consumer behavior with high accuracy.

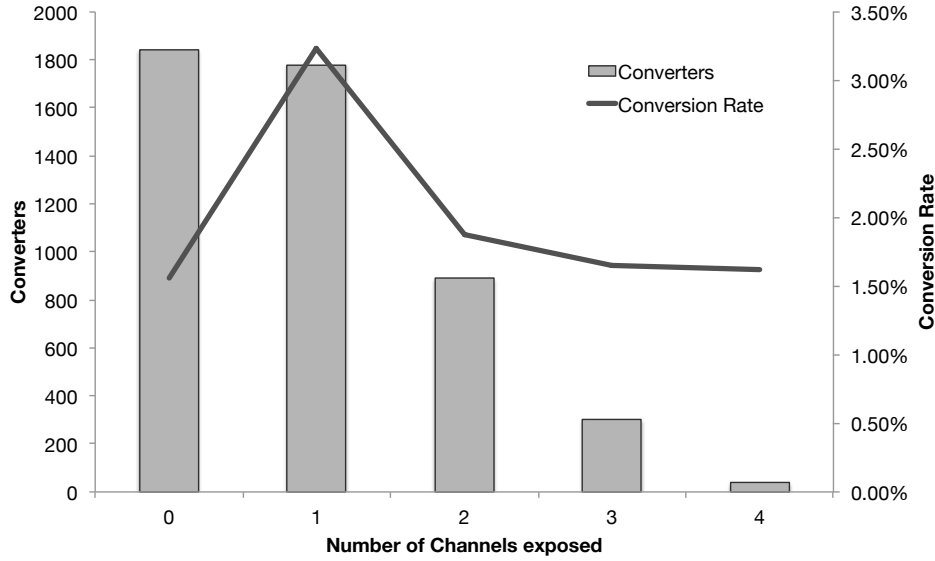
The negative properties of last-touch attribution under low noise levels as well as adverse selection⁴ has motivated us to search for an alternative attribution method that resolves these issues. The Shapley value is a cooperative game theory solution concept that allocates value among players in a cooperative game, and has the advantage of admitting the four requirements mentioned above along with uniqueness over the space of all conversion functions with the addition of an additivity property. Intuitively, the Shapley value (Shapley, 1952) has the economic impact of allocating the average marginal contribution of each publisher as a commission, and this chapter proposes its use as an improved attribution scheme. In equilibrium we find that the Shapley attribution scheme increases profit for the advertiser compared to regular CPA schemes regardless of the structure of the conversion function, while it improves over last-touch attribution for small noise ranges. Since the calculation of the Shapley value is computationally hard and requires data about subsets of publishers, a question arises whether generating this data by experimentation may be profitable for the advertiser.

Section 3.6 analyzes the impact of asymmetric information the publisher may have about the baseline conversion rate of consumers and running experiments on consumers. We first show that running an experiment to measure the baseline may control for the uncertainty in the information. The experiment uses a control group which is not exposed to ads to estimate the magnitude of the baseline. Since not showing ads may reduce the revenues of the publisher, we search for conditions under which the optimal sample size is small enough to merit this action. We find that when the population of the campaign is large enough, experimentation is always profitable, and armed with this result, we analyze the strategies publishers choose to use when they can target consumers with high probability of conversion. In equilibrium, we show that publishers in a CPA campaign with last-touch attribution will target baseline consumers in a non-efficient manner yielding less profit than CPM campaigns. Using the Shapley value with the results of the experiment, however, alleviates this problem completely as the value controls for the baseline.

⁴These results are presented in Section 3.6.

In Section 3.7 we investigate whether evidence exists for baseline exploitation or publisher free-riding in real campaign data. The data we analyze comes from a car rental campaign in the UK that was exposed to more than 13.4 million consumers. We observe that the budgets allocated to publishers exhibit significant heterogeneity and their estimates of effectiveness are highly varied when using last-touch methods. An estimate of publisher effectiveness when interacting with other publishers, however, gives an indication for baseline exploitation as predicted by our model, and lends credibility to the focus on the baseline in our analysis. Evidence for such exploitation can be gleaned from Figure 3.2, which describes the conversion behavior of consumers who were exposed to advertising only after visiting the car rental website without purchasing. If we compare the conversion rate of consumers who were

Figure 3.2: Converters and Conversion Rates of Visitors by Publisher Exposure



exposed to two or more publishers post-visit, it would appear that the advertising had little effect compared to no exposure post-visit.

We posit that the publishers target consumers with high probability of buying in order to be credited with the sale which is a by-product of the attribution method used by advertisers. To try and identify publishers who free-ride on others, we calculate an estimate of average marginal contributions of publishers based on the Shapley value, and use these estimates to compare the performance of publishers to last-touch methods. Calculating this value poses a significant computational burden and part of our contribution is a method to calculate this value that takes into account specific structure of campaign data. The results, which were communicated to the advertiser, show that a few publishers operate at efficient levels, while others target high baseline consumers to game the compensation scheme. We are currently in the process of collecting the information about the changes in behavior of publishers as a result of employing the Shapley value, and the results of this investigation is currently the

focus of research. To the best of our knowledge, this is the first large scale application of this theoretical concept appearing in the literature.

The discussion in Section 3.8 examines the impact of heterogeneity in consumer behavior on publisher behavior and the experimentation mechanism. We conclude with consideration of the managerial implications of proper attribution.

3.2 Industry Description and Related Work

Online advertisers have a choice of multiple ad formats including Search, Display/Banners, Classifieds, Mobile, Digital Video, Lead Generation, Rich Media, Sponsorships and Email. Among these formats, search advertising commands 46% of the online advertising expenditures in the U.S. followed by 21% of spending going to display/banner ads. Mobile advertising, which had virtually no budgets allocated to it in 2009, has grown to 9% of total ad expenditures in 2012. The market is concentrated with the top 10 providers commanding more than 70% of the entire industry revenue.

Although the majority of platforms allow fine-grained information collection during campaigns, the efficacy of these ads remains an open question. Academic work focusing on specific advertising formats has thus grown rapidly with examples including Sherman and Deighton (2001), Dreze and Hussherr (2003) and Manchanda et al. (2006) on banner advertising and Yao and Mela (2011), Rutz and Bucklin (2011) and Ghose and Yang (2009) on search advertising among others. Recent work that employed large scale field experiments by Lambrecht and Tucker (2011) on retargeting advertising, Blake et al. (2013) on search advertising and Lewis and Rao (2012a) on banner advertising have found little effectiveness for these campaigns when measured on a broad population. The main finding of these works is that the effects of advertising are moderate at best and require large sample sizes to properly identify. The studies by Lambrecht and Tucker (2011) and Blake et al. (2013) also find heterogeneous response to advertising by different customer segments.

When contracting with publishers, advertisers make decisions on the compensation mechanism that will be used to pay the publishers. The two major forms of compensation are performance based payment, sometimes known as Cost Per Action (CPA) and impression based payment known as Cost Per Mille (CPM). Click based pricing, known as Cost Per Click (CPC), is a performance based scheme for the purpose of our discussion. In 2012 performance based pricing took 66% of industry revenue compared to 41% in 2005. The growth has overshadowed impression based models that have declined from 46% to 32% of industry revenue. Part of this shift can be attributed to auction based click pricing pioneered by Google for its search ads. This shift resulted in significant research attention given to ad auction mechanisms from both an empirical and theoretical perspective which is not covered in this study. It is interesting to note that hybrid models based on both performance and impressions commanded only 2% of ad revenues in 2012.

In the past few years, the advertising industry has shown increased interest in improved

attribution methods. In a recent survey⁵ 54% of advertisers indicated they used a last-touch method, while 42% indicated that being “unsure of how to choose the appropriate method/model of attribution” is an impediment to adopting an attribution method. Research focusing on the advertiser’s problem of measuring and compensating multiple publishers is quite recent, however, with the majority focusing on empirical applications to specific campaign formats. Tucker (2012) analyzes the impact of better attribution technology on campaign decisions by advertisers. The paper finds that improved attribution technology lowered the cost per attributed converter. The paper also overviews theoretical predictions about the impact of refined measurement technology on advertising prices and makes an attempt to verify these claims using the campaign data. Kireyev et al. (2013) and Li and Kannan (2013) build specific attribution models for online campaign data using a conversion model of consumers and interaction between publishers. They find that publishers have strong interaction effects between one another which are typically not picked up by traditional measurements.

On the theory side, classic mechanism design research on team compensation closely resembles the problem an advertiser faces. Among the voluminous literature on cooperative production and team compensation the classic work by Holmstrom (1982) analyzes team compensation under moral hazard when team members have no private information. Our contribution is in the fact that the advertiser is a profit maximizing and not a welfare maximizing principal, yet we find similar effects and design mechanisms to solve these issues.

3.3 Model of Advertiser and Publishers

Consider a market with three types of players: an advertiser, two publishers and N homogenous consumers. Our interest is in the analysis of the interplay between the advertiser and publishers through the number of ads shown to consumers and allocation of payment to publishers. We assume advertisers do not have direct access to online consumers, rather they have to invest money and show ads through publishers in order to encourage consumers to purchase their products.

Consumers

Consumers in the model visit both publishers’ sites and are exposed to advertising, resulting in a probabilistic decision to “convert”. A conversion is any target action designated by the advertiser as the goal of the campaign that can also be monitored by the advertiser directly. Such goals can be the purchase of a product, a visit to the advertiser’s site or a click on an ad.

The response of consumers to advertising depends on the effectiveness of advertising as well as on the propensity of consumers to convert without seeing any ads which we call the baseline conversion rate. The baseline captures the impact of various states of consumers

⁵Source: “Marketing Attribution: Valuing the Customer Journey” by EConsultancy and Google.

resulting from exogenous factors such as brand preference, frequency of purchase in steady state and effects of offline advertising prior to the campaign. When each publisher $i \in \{1, 2\}$ shows q_i ads, we let $(q_1 + q_2)^\rho$ denote the conversion rate of consumers who have a zero baseline.⁶ By denoting the baseline probability of conversion as s , the advertiser expects to observe the following conversion rate after the campaign:

$$x(q_1, q_2) = s + (q_1 + q_2)^\rho(1 - s) \quad (3.1)$$

The values of ρ and s are determined by nature prior to the campaign and are exogenous. To focus on pure strategies of advertising, we assume that $0 < \rho < 2$.⁷ The assumption implies that additional advertising has a positive effect on the probability of buying of a consumer, yet allows both increasing and decreasing returns. When $\rho < 1$ the response of consumers to additional advertising has decreasing returns and publishers' ads are substitutes. When $\rho > 1$ publishers' ads are complements.

Finally, we let the baseline s be distributed $s \sim \text{Beta}(\alpha, \beta)$ with parameters $\alpha > 0, \beta > 0$. The flexible structure will let us understand the impact of various campaign environments on the incentives of advertisers and publishers.

Publishers

Publishers in the model make a simultaneous choice about the number of ads q_i to show to each consumer and try to maximize their individual profits. When showing these ads publishers incur a cost resulting from their efforts to attract consumers to their websites. We define the cost of showing q_i ads as $\frac{q_i^2}{2}$. Both publishers have complete information about the values of ρ and s , as well as the conversion function x and the cost functions.

At the end of the campaign, each publisher receives a payment b_i from the advertiser that may depend on the amount of ads that were shown and the conversion rate observed by the advertiser. The profit of each publisher i is therefore:

$$u_i = b_i(q_1, q_2, x) - \frac{q_i^2}{2} \quad (3.2)$$

The Advertiser

The advertiser's goal is to maximize its own profit by choosing the payment contract b_i to use with each publisher prior to the campaign. The structure of the conversion function x , as well as the value of ρ are known to the advertiser. Initially, we assume as a benchmark that the baseline s is known to the advertiser, which we normalize to zero without loss of generality. The goal of this assumption, to be relaxed later, is to distinguish the effects

⁶The additivity of advertising effects is not required but simplifies exposition. Asymmetric publisher effectiveness is discussed in Appendix B.2.

⁷Restricting $\rho < 2$ is sufficient for the existence of profitable pure strategies when costs are quadratic.

of strategic publisher interaction on the advertiser's profit from the effects of additional information the publishers may have about consumers.

Normalizing the revenue from each consumer to 1, the profit of the advertiser is then:

$$\pi = x(q_1, q_2) - b_1(q_1, q_2, x) - b_2(q_1, q_2, x) \quad (3.3)$$

Types of Contracts - CPM and CPA

The advertising industry primarily uses two types of contracts - performance based contracts (CPA) in which publishers are compensated on the outcome of a campaign, and effort based contracts (CPM) in which publishers receive payment based on the amount of ads they show. As noted in the introduction, hybrid contracts that make use of both types of payments are uncommon. As shown by Zhu and Wilbur (2011), in environments that allow hybrid campaigns, rational publishers expectations will rule out hybrid strategies by advertisers.

CPM contracts (cost per mille or cost per thousand impressions) are effort based contracts in which the advertiser promises each publisher a flat rate payment p_i^M for each ad displayed to the consumers. The resulting payment function $b_i^M(q_i; p_i^M) = q_i p_i^M$ depends only on the number of ads shown by each publisher. The profit of the publisher becomes:

$$u_i = q_i p_i^M - \frac{q_i^2}{2} \quad (3.4)$$

CPA contracts (cost per action) are performance based contracts. In these contracts the advertiser designates a target action to be carried out by a consumer, upon which time a price p_i^A will be paid to the publishers involved in causing the action. The prices are defined as a share of the revenue x , yielding the following publisher profit:

$$u_i = (q_1 + q_2)^\rho p_i^A - \frac{q_i^2}{2} \quad (3.5)$$

The timing of the game is illustrated in Figure 3.3. The advertiser first decides on a compensation scheme based on the observed efforts q_i , performance x or both. The publishers in turn learn the value of the baseline s and make a decision about how many ads q_i to show to the consumers. Consumers respond to ads and convert according to $x(q_1, q_2)$. Finally, the advertiser observes q_i and x , compensates each publisher with b_i and payouts are realized.

Several features of the model make the analysis interesting and are considered in the next sections. The first is that the interaction among the publishers is essentially of a team generating conversions. A well known result by Holmstrom (1982) shows that no fixed allocation of output among team members can generate efficient outcomes without breaking the budget. In our model, however, a principal is able to break the budget, yet its goal is profit maximization rather than efficiency. Nonetheless, the externality that one publisher causes on another by showing ads will create moral hazard under a CPA model as will be presented in the next section.

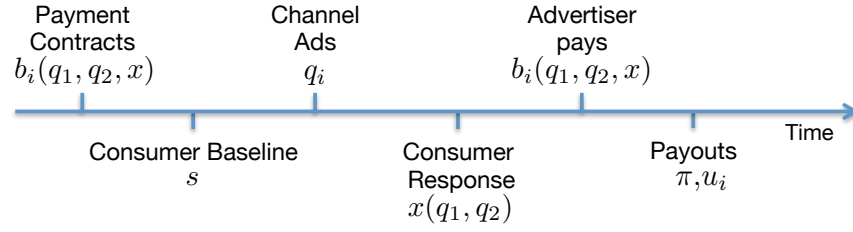


Figure 3.3: Timing of the Campaign

The second feature is that under CPM payment neither the performance of the campaign nor the effect of the baseline enter the utility function of the publishers directly and therefore do not impact a publisher's decision regarding the number of ads to show. Consequently, if the advertiser does not use the performance of the campaign as part of the compensation scheme, adverse selection will arise.

Finally, we note that both the effort of the publishers as well as the output of the campaign are observed by the advertiser. Traditional analysis of team production problems typically assumed one of these is unobservable by the advertiser and cannot be contracted upon. Essentially, CPA campaigns ignore the observable effort while CPM campaigns ignore the observable performance. As we will show, a primary effect of an attribution process is to tie the two together into one compensation scheme.

We now proceed to analyze the symmetric publisher model under CPM and CPA payments. The analysis builds towards the inclusion of an attribution mechanism with a goal of making multi-publisher campaigns more profitable for the advertiser.

3.4 CPM vs. CPA and the Role of Attribution

We start by developing a benchmark that assumes the advertiser is integrated with the publishers. The optimal allocation of ads is found by solving $\max_{q_1, q_2} (q_1 + q_2)^\rho - \frac{q_1^2}{2} - \frac{q_2^2}{2}$ yielding

$$q_1^* = q_2^* = (\rho \cdot 2^{\rho-1})^{\frac{1}{2-\rho}} \quad (3.6)$$

which is strictly increasing in ρ .

When using CPM based payments, the publisher will choose to show $q_i^M = p_i^M$ ads. Because of symmetry, in equilibrium $q^M = p^M = p_1^M = p_2^M$ and the number of ads displayed is:

$$q^M = p^M = \arg \max_p (2p)^\rho - 2p^2 = \frac{\rho^{\frac{1}{2-\rho}}}{2} \quad (3.7)$$

In contrast, under a CPA contract, publisher i will choose q_i to solve the first order condition $q_i = \rho(q_i + q_{-i})^{\rho-1} p_i^A$. Invoking symmetry again, we expect $p_1^A = p_2^A$ and $q_1^A = q_2^A$,

as a result yielding:

$$q^A = (\rho 2^{\rho-1} p^A)^{\frac{1}{2-\rho}} \quad (3.8)$$

We notice that the number of ads displayed in a CPA campaign increases with the price p^A offered to the publishers.

By performing the full analysis and solving for the equilibrium prices p^M and p^A offered by the advertiser we find the following:

Proposition 3. *When $0 < \rho < 2$:*

- $q^A < q^M < q^*$ - *the level of advertising under CPA is lower than the level under CPM. Both of these are lower than the efficient level of advertising.*
- $\pi^M > \pi^A$ - *the profit of the advertiser is higher when using CPM contracts.*
- *There exists a critical value ρ^c with $0 < \rho^c < 1$ s.t. for $\rho < \rho^c$, $u^A > u^M$ and CPA is more profitable for the publishers. When $\rho > \rho^c$, $u^M > u^A$ and CPM is more profitable for the publishers.*

Proposition 3 shows that using CPA causes the publishers to free-ride and not provide enough effort to generate sales in the campaign. The intuition is that the externality each publisher receives from the other publisher gives an incentive to lower efforts, which consequently lowers total output of the campaign. Under CPM payment, however, publishers do not experience this externality and cannot piggyback on efforts by other publishers. By properly choosing a price for an impression, the advertiser can then incentivize the publishers to show a higher number of ads.

In terms of profits, we observe that advertisers should always prefer to use CPM contracts when multiple publishers are involved in a campaign. This counter-intuitive result stems from the fact that the resulting under-provision of effort overcomes the gains from cooperation by the publishers even when complementarities exist.

The final part of Proposition 3 gives one explanation to the market observation that campaigns predominantly use CPA schemes. When the publishers have market power to determine the payment scheme, *e.g.* the case of Google in the search market, the publishers should prefer a CPA based payment when ρ is small, *i.e.*, when publishers are extreme substitutes. In this case, the possibility for free-riding is at its extreme, and even minute changes in efforts by competing publishers increase the profits of each publisher significantly. For example, if consumers are extremely prone to advertising and a single ad is enough to influence them to convert, any publisher that shows an ad following the first one immediately receives “free” commission. If a search engine which typically arrives later in the buying process of a consumer, is aware of that, it will prefer to use CPA payment to free-ride on previous publisher advertising.

A question that arises is about the motivation of advertisers, in contrast to publishers, to prefer CPA campaigns over CPM ones. The following corollary shows that when advertisers

do not take into account the interaction between the publishers, CPA campaigns are also profitable for the advertiser.

Corollary 3. *When there is one publisher in a campaign and $0 < \rho < 2$:*

- $q^A > q^M$ iff $\rho > 1$: the publisher shows more ads under CPA payment.
- $\pi^A > \pi^M$ iff $\rho > 1$: more revenue and more profit is generated for the advertiser when using CPA payment and advertising has increasing returns ($\rho > 1$).

Corollary 3 reverses some of the results of Proposition 3 for the case of one publisher campaigns. Since free-riding is not possible in these campaigns, we find that CPA campaigns better coordinate the publisher and the advertiser when ads have increasing marginal returns, while CPM campaigns are more efficient for decreasing marginal returns.

The Role of Attribution

An attribution process in a CPA campaign allocates the price p^A among the participating publishers in a non-fixed method. We model the attribution process as a two-dimensional function $f(q_1, q_2, x) = (f_1, f_2)$ that allocates a share of a conversion to each of the players respectively. When publishers are symmetric and the baseline is zero, candidates for effective attribution functions will exhibit the following properties:

- Efficiency - The process will attribute all conversions to the two publishers: $f_1 + f_2 = 1$.
- Symmetry - If both publishers exhibit the same effort ($q_1 = q_2$) then they will receive equal attribution: $f_1(q, q, x) = f_2(q, q, x) = \frac{1}{2}$.
- Pay to play (Null Player) - Publishers have to invest to get credit. When a publisher does not show any ads, it will receive zero attribution: $f_i(q_i = 0, q_{-i}, x) = 0$.
- Marginality - Publishers who contribute more to the conversion process should receive higher attribution: if $q_1 > q_2$ then $f_1 \geq f_2$.

Although these properties are straightforward, they limit the set of possible functions that can be used for attribution. We also assume that $f(\cdot)$ is continuously differentiable on each of its variables.

The profit of each publisher in a CPA campaign can now be written as:

$$u_i^A = f_i(q_i, q_{-i}, x)x(q_1, q_2)p^A - \frac{q_i^2}{2} \quad (3.9)$$

An initial observation is that the process creates a contest between the two publishers for credit. Once ads have been shown, the investment has been sunk yet credit depends on delayed attribution. It is well known (see, *e.g.*, Sisak (2009) and Konrad (2007)) that contests will elicit the agents to overexert effort in equilibrium compared to a non-contest

situation. As a result the attribution process can be used to incentivize the publishers to increase their efforts and show a number of ads closer to the integrated market levels.

In the next section we analyze the impact of the commonly used last-touch attribution method, and compare it to a new method based on the Shapley value we developed to attribute performance in online campaigns.

3.5 Last-Touch and Shapley Value Attribution

Advertiser surveys report that last-touch attribution is the most widely used process in the industry. This process gives 100% of the credit for conversion to the last ad displayed to a consumer before conversion. From the point of view of the publisher, if the consumer visits both publisher sites, last-touch attribution creates a noisy contest in which the publisher cannot fully predict whether it will receive credit by showing a specific impression. Even if the publisher can predict the equilibrium behavior of the other publisher and expect the number of ads shown by the other publisher, it has little knowledge of the timing of these ads, and in addition it cannot fully predict the timing of a consumer purchase.

Consequently, we model the process as a noisy contest. The noise in the contest models the uncertainty the publisher has about whether a consumer is about to purchase the product or not, and whether they will visit the site again in the future. We let ε_i denote the uncertainty of publisher i with respect to its ability to win the attribution process. When publisher i shows q_i ads it will receive credit only if $Pr(q_1\varepsilon_1 > q_2\varepsilon_2)$. In a static model this captures the effect of showing an additional ad by the publisher. By assuming that ε_i are uniformly i.i.d on $[1, d]$ for $d > 1$, we can define the last-touch attribution function as following:

$$f_i^{LT}(q_i, q_{-i}) = Pr(q_i\varepsilon_i > q_{-i}\varepsilon_{-i}) = \int_1^d G\left(\frac{q_i}{q_{-i}}\varepsilon\right) g(\varepsilon) d\varepsilon \quad (3.10)$$

when $G(\cdot)$ is the CDF of the uniform distribution on $[1, d]$ and $g(\cdot)$ its PDF.

The value of d measures the amount of uncertainty the publishers have about the consumer's behavior in terms of future visits and purchases, and will be the focus of our analysis of Last-Touch attribution. Higher values of d , for example, can model consumers who visit both publishers with very high frequency, allowing both of them to show many ads to the consumer. Lower values of d make the contest extremely discriminating, having a “winner-take-all” effect on the process. In such cases, the publishers can time their ads exactly to be the last ones to be shown, and as a result compete fiercely for attribution. A natural extension which is left for future work is to allow asymmetric values of d among the publishers. This will allow modeling of publishers who have an advantage in timing their advertising to receive credit, although their ads may have the same effectiveness.

Two noticeable properties of last-touch attribution are due discussion. The first is that the more ads a publisher will show, the higher probability it has of being the last one to show an ad before a consumer's purchase. Last-touch attribution therefore has the Marginality

property described above. It also trivially has the 3 other properties. The second property is that last-touch attribution makes use of the conversion rate only in a trivial manner. The credit given to the publisher only depends on the number of ads shown to a consumer and whether the consumer had converted. It does not depend on the actual conversion rate of the consumer and therefore ignores the value of x .

It is useful to examine the equilibrium best response of the publishers in a CPA campaign in order to understand the impact of last-touch attribution on the quantities of ads being displayed. Recall that when no attribution is used, the publisher will display q ads according to the solution of:

$$(2q)^{\rho-1} \rho p^A = q \quad (3.11)$$

When using last-touch attribution, a publisher faces a winner-take-all contest which increases its marginal revenue when receiving credit for the conversion, even if the conversion rate remains the same. In a CPA campaign the first order condition in a symmetric equilibrium becomes:

$$(2q)^{\rho-1} \left(2f'_1(1) + \frac{1}{2}\rho \right) p^A = q \quad (3.12)$$

where $f'(1)$ is the marginal increase in the share of attribution when showing an additional ad when $q_1 = q_2$. Comparing equations (3.11) and (3.12) we see that if $(2f'_1(1) + \frac{1}{2}\rho) > \rho$, then the publisher faces a higher marginal revenue for the same amount of effort. As a result it will have an incentive to increase its effort in equilibrium when the conversion function is concave compared to the case when no attribution was used. Gershkov et al. (2009) show conditions under which such a tournament can achieve Pareto-optimal allocation when symmetric team members use a contest to allocate the revenue among themselves. Whether this contest is sufficient to compensate for free-riding in online campaigns remains yet to be seen.

To answer this question we are required to perform the full analysis that considers the price p^A offered by the advertiser in equilibrium. In addition, the accuracy of the attribution process which depends on the magnitude of the noise d has an impact and may yield exaggerated effort by each publisher. Finally, the curvature of the conversion function x that depends on the parameter ρ may also influence the efficiency of last-touch attribution.

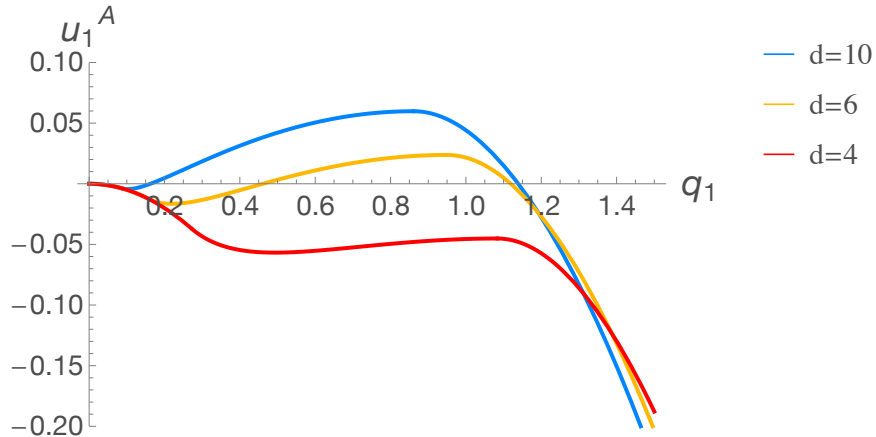
When performing the complete analysis for both CPA and CPM campaigns, we find the following:

Proposition 4. *When $0 < \rho < 2$ and last-touch attribution is being used:*

- *In a CPA campaign a symmetric pure strategy equilibrium exists for $0 < \rho < 2 - \frac{4}{d-1}$. In this equilibrium $q^{A-LT} = \left(\frac{\rho}{2} 2^{\rho-1} \left(\frac{d+1}{d-1} + \frac{\rho}{2} \right) \right)^{\frac{1}{2-\rho}}$.*
- *For any noise level d , $q^A < q^M < q^{A-LT}$.*

Proposition 4 shows surprising findings about the impact of last-touch attribution on different campaign types. The contest among the publishers has a symmetric pure strategy equilibrium in a CPA campaign when ρ is low enough or when the noise d is high enough. In these cases, more advertising is being shown in equilibrium compared to regular CPM and CPA campaigns, and more revenue will be generated by the campaign. As a result, the advertiser may make higher profit compared to the case of no attribution as well as for the case of CPM campaigns with no attribution. To understand the impact of low noise, we focus on the case of $d < 3$. In this case, the contest is too discriminating and the effort required from the publishers in equilibrium is too high to make positive profit, and publishers would prefer not to participate. Figure 3.4 illustrates the best-response of publisher 1 to publisher's 2 equilibrium strategy to give intuition for this result. When the noise becomes small and the contest too discriminating, the best-response function loses the property of having a maximum point which yields positive profit as a result of too strong competition for attribution.

Figure 3.4: Best Response of Player 1 Under Last-Touch Attribution



Publisher's 1 best response to publisher's 2 strategy of showing q^{A-LT} ads when $\rho = 1$.

Finally, a comparison of the profits the advertiser makes with and without last-touch attribution yields the following result:

Corollary 4. *When $0 < \rho < 2 - \frac{4}{d-1}$, $\pi^{A-LT} > \pi^M > \pi^A$ and the advertiser makes higher profit under last-touch attribution.*

The Shapley Value as an Attribution Scheme

The Shapley value (Shapley, 1952) is a cooperative game theory solution concept that allocates value among players in a cooperative game. A cooperative game is defined by a characteristic function $x(q_1, \dots, q_M)$ that assigns for each coalition of players and their contribution q_i the value they created. For a set of M publishers, the Shapley value is defined

as following:⁸

$$\phi_i(x) = \sum_{S \subset (M \setminus i)} \frac{|S|!(|M| - |S| - 1)!}{|M|!} (x_{S \cup i} - x_S) \quad (3.13)$$

where M is the set of publishers and x is the set of conversion rates for different subsets of publishers.

The value has the four properties mentioned in the previous section: Efficiency, Symmetry, Null Player and Marginality.⁹ In addition, it is the unique allocation function that has these properties with the addition of an additivity property over the space of cooperative games defined by the conversion function $x(\cdot)$. For the case of two publishers $M = 2$ the Shapley value reduces to:

$$\phi_1 = \frac{x(q_1+q_2)-x(q_2)+x(q_1)-0}{2} \quad \phi_2 = \frac{x(q_1+q_2)-x(q_1)+x(q_2)-0}{2} \quad (3.14)$$

Using the Shapley value has the benefit of directly using the marginal contribution of the publishers to compensate them. In addition, the process's accuracy does not depend on exogenous noise and yields a pure strategy equilibrium for all values of ρ .

In a CPA campaign, the profit of a publisher will become: $u_i^{A-S} = \phi_i p^{A-S} - \frac{q_i^2}{2}$.

Solving for the symmetric equilibrium strategies and profits of the advertiser and publishers yield the following result:

Proposition 5.

When $0 < \rho < 2$, using the Shapley value for attribution yields $q^{A-S} = \left(\frac{\rho^2}{4} (2^{\rho-1} + 1) \right)^{\frac{1}{2-\rho}}$.

For $\rho < 2 - \frac{4}{d-1}$, $q^A < q^{A-S} < q^{A-LT}$.

The profit of the advertiser is higher under Shapley value than under Last-Touch attribution iff $q^{A-S} > q^{A-LT}$, i.e. $d < \frac{4}{2-\rho} + 1$.

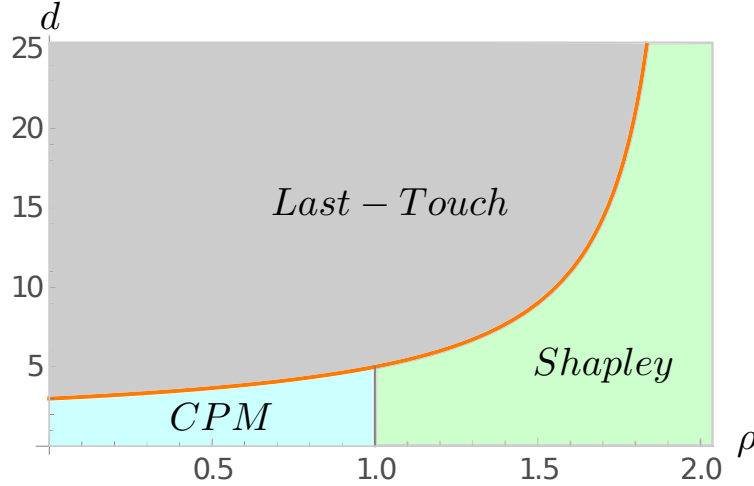
The profit of the publisher is higher under the Shapley value attribution than under regular CPM pricing iff $\rho > 1$.

Proposition 5 is a major result of this chapter, showing that the Shapley value can be more profitable when publishers are complements. Contrary to Last-touch attribution, a symmetric pure strategy equilibrium exists for any value of ρ , including very convex functions. When considering lower values of ρ for which Last-Touch attribution improves the efficiency of the campaign, we see that when the noise level d is low enough, the Shapley value will yield better results for the advertiser if $\rho > 1$, while CPM will be better when $\rho < 1$. Figure 3.5 depicts for which values of ρ and d is each attribution and compensation scheme more profitable.

⁸This is a continuous version of the value.

⁹Some of these properties can be shown to be derived from others.

Figure 3.5: Profitability of Each Compensation Scheme



Values of ρ and d for which each compensation scheme is more profitable for the advertiser.

The intuition behind this result can be illustrated best for extreme values of ρ . When $\rho < 1$ and is extremely low, the initial ads have the most impact on the consumer. As a result, there will be significant free-riding which Last-touch is best suited to solve, while the marginal increase that the Shapley value allocates is not too high. When $\rho > 1$, however, if the noise is low enough, the publishers will be inclined to show too many ads because of the low uncertainty about their success of being the last one to show an ad. In essence, the competition is too strong and overcompensates for free-riding. The Shapley value in this case is better suited to incentivize the players as the marginal increase between two symmetric publishers to one is highest with a convex function.

To make use of the Shapley value in an empirical application, it is required that the advertiser can observe the conversion rates of consumers who were exposed to publisher 1 solely, publisher 2 solely and to both of them together. In addition, when a baseline is present, it cannot be assumed that not being exposed to ads yields no conversions.

The next section discusses the baseline and the use of experimentation to generate the data required to calculate the Shapley value.

3.6 Baselines and Experiments

In this Section we relax the assumption that the baseline $s = 0$ and examine its impact on the performance of the attribution schemes, and methods to fix this impact. When the baseline is non-zero, the advertiser cannot discern from conversions whether they were caused by advertising effects or simply because consumers had other reasons for converting. As publishers have more information about consumers reaching their sites, this private information

may cause adverse selection - publishers can target consumers with high baselines to receive credit for those conversions.

Specifically, if we consider again equation (3.12) the first order condition of an advertiser showing q ads to all consumers now becomes:

$$\left[(2q)^{\rho-1} \left(2f'_1(1) + \frac{1}{2}\rho \right) (1-s) + f'(1)s \right] p^A = q \quad (3.15)$$

In the extreme case of $s = 1$, the publishers will elect to show advertising to baseline consumers and be attributed credit.

To understand how experimentation may be beneficial for the advertiser in light of this problem, we analyze a model with a single publisher, but now assume the baseline is non-zero and known to the publisher. We also assume $\rho = 1$, and recall that s is distributed $Beta(\alpha, \beta)$. Thus, if all consumers are exposed to q ads, the expected observed number of converters will be $N(s + q(1-s))$. We note however that if non-baseline consumers are not exposed to ads at all, the advertiser would still expect to observe $N(s + q(1-s))$ converters.

When the advertiser is integrated with the publisher and can target specific consumers, it can choose to show q_b ads to baseline consumers and q ads to the non-baseline consumers. If the cost of showing q ads to a consumer is $\frac{q^2}{2}$ the firm's profit from advertising is:

$$\pi(q, q_b; s) = N \left(s + q(1-s) - s \frac{q_b^2}{2} - (1-s) \frac{q^2}{2} \right) \quad (3.16)$$

The insight gained from this specification is that when consumers have a high baseline, the advertiser has a smaller population to affect with its ads, as consumers in the baseline would convert anyway.

It is obvious that when the advertiser can target consumers exactly, it has no reason to show ads to baseline consumers, and therefore will set $q_b = 0$. The allocation of ads that maximizes the advertiser's profit under full information is then $q^* = 1$ and $q_b^* = 0$, while the total number of ads shown will be $N(1-s)$. We call this strategy the *optimal* strategy and note that the number of ads to show decreases in the magnitude of the baseline. The profit achieved under the optimal strategy is $\pi^{max} = N \frac{\mu+1}{2}$ when μ is the expectation of s .¹⁰ This profit increases with α , and decreases with β . This means that when higher baselines are more probable in terms of mass above the expectation, a higher profit is expected.

Turning to the case of a firm with uncertainty about s , one approach the firm may choose is to maximize the expected profit over s by showing a number of ads q to all consumers independent of the baseline. This *expected* strategy solves:

$$\max_q \mathbb{E}_s[\pi(q, q; s)] \quad (3.17)$$

The achieved profit in this case can serve as a lower bound π^{min} on profit the firm can achieve in the worst case. Any additional information is expected to increase this profit; if it does not, the firm can opt to choose the expected strategy.

¹⁰ $\mu = \frac{\alpha}{\alpha+\beta}$

The following result compares the expected strategy with the optimal one:

Lemma 1. *Let $q^E = \arg \max_q \pi(q, q; s)$. Then:*

- *The firm will choose to show $q^E = 1 - \mu$ ads when using the expected strategy.*
- *The firms profit, π^{\min} is lower than π^{\max} by $\frac{N}{2}(\mu - \mu^2)$.*

Lemma 1 posits that the number of ads displayed using this strategy treats the market as if s equals its expected value. As a result, the achieved profit increases with the expected value of s . When this strategy is the only one available the value of full information to the firm is highest when the expected baseline is close to $1/2$.

The most common strategy that firms employ in practice, however, is to learn the value of s through experimentation. The firm can decide to not show ads to $n < N$ consumers and observe the number of converters in the sample. This information is then used to update the firm's belief about s and maximize q . We call this strategy the *learning* strategy.

When the firm observes k converters in the sample it will base the number of ads to show on this updated belief (DeGroot, 1970). The expected profit of the firm in this case is:

$$n\mathbb{E}_s[x(q = 0; s)] + (N - n)\mathbb{E}_s\left[\mathbb{E}_{k|s}\left[\max_q \pi(q; s)|s\right]\right] \quad (3.18)$$

The caveat here is that by designating consumers as the sample set, the firm forfeits potential added profit from showing ads to these consumers. We are interested to know when this strategy is profitable, and also how much can be gained from using it and under what conditions.

Let n^* denote the optimal sample size that maximizes (3.18) given the distribution of s . As the distribution of the observed converters k is $Bin(n, s)$, the posterior $s|k$ is distributed $Beta(\alpha + k, \beta + n - k)$. Using Lemma 1, the optimal number of ads to show when observing k converters becomes $q^*(\mu(k))$ when $\mu(k) = \mathbb{E}_s[s|k] = \frac{\alpha + k}{\alpha + \beta + n}$. A comparative statics analysis of the optimal sample size n^* shows the following behavior:

Lemma 2. *The optimal sample size n^* :*

- *Is positive when the population N is larger than $\frac{\beta(\alpha + \beta)(1 + \alpha + \beta)}{\alpha} = \beta \frac{1 - \mu}{\sigma^2}$.*
- *Increases with N and decreases with β .*
- *Decreases in α when α is large.*

Lemma 2 shows that unless the distribution of s is heavily skewed towards 0 by having a large β parameter, even with small populations some experimentation can be useful. On the flip side, when the distribution is heavily skewed towards 1 with very large α , the high probability baseline makes it less valuable to experiment, and the optimal sample size decreases.

Having set conditions for the optimal size of the sample during experiments, we now revisit our question: when is it profitable for the firm to learn compared to choosing an expected strategy. Our finding is that for a large enough population N , it is always more profitable to learn than to use an expected strategy:

Proposition 6. *When $N > \beta \frac{1-\mu}{\sigma^2}$, learning yields more profit than the expected strategy.*

To exemplify this result, if $\alpha = \beta = 1$, then the baseline s is distributed uniformly over $[0, 1]$. In this case, it is enough for the population to be larger than 6 consumers for experimentation to be profitable.¹¹

Baseline Exploitation

When the advertiser is not integrated with the publisher, the publisher has a choice of which consumers to target and how many ads to show to each segment. We can solve for the behavior of an advertiser under CPM and CPA pricing in this special case without attribution to get the following result:

Proposition 7. • *Under CPM the publisher will show $q^M = \frac{1-\mu}{2}$ ads to each consumer in both segments.*

- *Under CPA, the publisher will show $q^A = \frac{2\mu-1}{2\mu-2}$ when $\mu < \frac{1}{2}$. The ads will be shown to consumers only in the non-baseline segment. When $\mu > \frac{1}{2}$, the advertiser will opt to not use CPA at all.*
- *Under CPA the publisher will show a total number of ads which is higher than the efficient number q^* , as well as higher than q^M , for every value of s .*
- *The profit of the advertiser under CPM is higher than under CPA for any value of μ .*

Proposition 7 exposes two seemingly contradicting results. Since under CPM payment the publisher is paid for the amount of ads it shows, it will opt to show both $q > 0$ and $q_b > 0$ ads. Given the same price and cost for each ad displayed, it will show exactly the same amount to both segments, which will be lower than the efficient amount of ads to show. Specifically, when μ is high, *i.e.*, the expectation of the baseline is high, the publisher will lower its effort as the advertiser would have wanted. Under CPA, however, the publisher will use an efficient allocation of ads in terms of targeting and will not show ads to the baseline population. Since the publisher gets a commission from the baseline as well, however, it experiences lower effective cost for each commission payment, and as a result will show too many ads compared to the optimal amount. The apparent contradiction may be that although the publisher now allocates its ads correctly under CPA compared to CPM, the profit of the advertiser is still higher under CPM payment for low baseline values. The

¹¹This result assumes n is continuous. As n is discrete, the actual n^* is slightly larger than this bound to allow for discrete sizes of samples.

intuition is that CPM allows the advertiser to internalize the strategy of the publisher and control it through the price, while in CPA the advertiser will need to trade-off effective ads for ineffective exploitation of the baseline if it lowers the price paid per conversion.

Adding Last-Touch attribution to the CPA process will only exacerbate the issue. If the publisher will show a different number of ads in each segments, the advertiser can infer which segment may be the baseline one and not compensate the publisher for it. The publisher, as a result, will opt to show the same number of ads to all consumers, and the number of ads shown will now depend on the size of the baseline population s . The result will be too many ads shown by a CPA publisher to the entire population, and reduced profit to the advertiser.

Using the Shapley value, in contrast, will allocate revenue to the publisher only for non-baseline consumers, as the Shapley value will control for the observed baseline through experimentation. When solving for the total profit of the advertiser including the cost of experimentation, it can be shown that Shapley value attribution in a CPA campaign reaches a higher profit than CPM campaigns.

We thus advocate moving to an attribution process based on the Shapley value considering the adverse effects of the baseline. The next section discusses a preliminary analysis of data from an online campaign using Last-Touch attribution to detect whether baseline exploitation is indeed occurring.

3.7 An Application to Online Campaigns

This section applies the insights from Sections 3.4, 3.5 and 3.6 to data from a large scale advertising campaign for car rental in UK.

The campaign was run during April and May 2013 and its total budget exceeded US \$65,000 while utilizing 8 different online publishers. These publishers include two online magazines, two display (banner) ad networks, two travel search websites, an online travel agency and a media exchange network. During the campaign more than 13.4 million online consumers¹² were exposed to more than 40.4 million ads.

The summary of the campaign results in Table 3.1 shows that the campaign more than quadrupled conversion rates for the exposed population.

Ad Exposure	Population	Converters	Conversion Rate
Exposed	13,448,433	6,030	0.045%
Not Exposed	144,745,194	15,087	0.010%
Total	158,193,627	21,117	0.013%

Table 3.1: Performance of Car Rental Campaign in the UK

To associate the return of the campaign the advertiser computed last-touch attribution for the publishers based on the last ad they displayed to consumers. Table 3.2 shows the

¹²An online consumer is measured by a unique cookie file on a computer.

attributed performance alongside the average cost per attributed conversion. We see that the allocation of budgets correlates with the attributed performance of the publishers, while the cost per conversion can be explained by different average sales through each publisher and quantity discounts.¹³

Publisher No.	Type	Attribution	Budget (\$)	Cost per Converter (\$)
1	Online Magazine	386	8,300	21.50
2	Travel Agency	218	8,000.02	36.69
3	Travel Magazine	40	6,000	150
4	Display Network	168		
5	Travel Search	50		
6	Display Network	1,330	13,200	9.92
7	Travel Search	69		
8	Media Exchange/Retargeting	3,769	33,200	8.80
Total		6,030	68,700	11.39

Table 3.2: Last Touch Attribution for the Car Rental Campaign

We observed that in order to achieve high profits, the advertiser needs to be able to condition payment on estimates of the baseline as well as on the marginal increase of each publisher over the sets of other publishers. This result extends to the case of many publishers, where for a set of publishers M the advertiser will need to observe and estimate $2^{|M|}$ measurements.

Even small campaigns utilizing 7 publishers require more than 100 of these estimates to be used and reported. Current industry practices do not allow for such elaborate reporting resulting in advertisers using statistics of these values. The common practice is to report one value per publisher with the implicit assumption that if a publisher's attribution value is higher, so is its effectiveness.

Evidence of Baseline Exploitation and Detection of Free-Riding

Section 3.6 shows that publishers can target high baseline consumers to deceive the advertiser regarding their true effectiveness. To test the hypothesis that publishers target high baseline consumers, Table 3.3 shows the results of the logit estimates on the market share differences of each publisher combination in our data.¹⁴ The estimate shows that no publisher adds a statistically significant increase in utility for consumers compared to the baseline. More surprising is the result that a few publishers seem to decrease the response of consumers, thus supporting our hypothesis.

Section 3.5 predicts that using a last-touch method will lead publishers to strategically increase the number of ads shown, while attempting to free-ride on others. If publishers

¹³Publisher number 3 targets business travelers and yields more profit per attributed conversion.

¹⁴The estimation technique is described in Appendix B.3.

Publishers	logdiff coef	se
1	-0.657	(0.849)
2	-2.175***	(0.693)
3	-1.960***	(0.703)
4	-0.986	(0.751)
5	-1.559**	(0.691)
6	-1.689**	(0.744)
7	-0.588	(0.748)
8	-0.539	(0.813)
R ²	0.650	
Observations	88	
Standard errors in parentheses		
*** p<0.01, ** p<0.05, * p<0.1		

Table 3.3: Logit Estimates of Publisher Effectiveness

were not attempting to game the last-touch method, we would expect to see their marginal contribution estimates be close to their last-touch attribution in equilibrium. An issue that arises with using marginal estimates from the data, however, is that the timing of ads being displayed is endogenous and depends on a decision by the consumer to visit a publisher and by the publisher to display the ad. The advertiser does not observe and cannot control for this order, which might raise an issue with using ad view data as created by random experimentation.

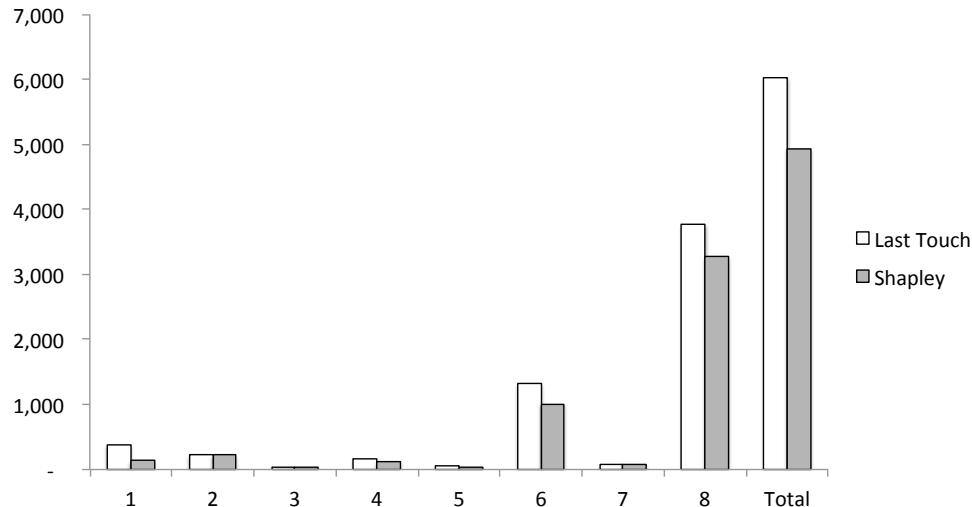
The use of the Shapley value, however, gives equal probability to the order of appearance of a publisher when a few publishers show ads to the same consumers. The effect is a randomization of order of arrival of ads when multiple ads are observed by the same consumer. Because of this fact, using the Shapley value as is to estimate marginal contributions will be flawed when not every order of arrival is possible. For example, the baseline effect needs to be treated separately while special publishers such as retargeting publishers and search publishers that can only show ads based on specific events need to be accounted for. An additional hurdle to using the Shapley value is the computation time required as it is exponential in the size of the input.

We developed a modified Shapley value estimation procedure to handle these issues. The computational issues are addressed by using specific structure of the advertising campaign data and will be described in Berman (2013).

Figure 3.6 compares the results from a last-touch attribution process to the Shapley value estimation.

More than 1,000 converters were reallocated to the baseline. In addition, a few publishers lost significant shares of their previously attributed contributions, showing evidence of baseline exploitation. Using these attribution measures the advertiser has reallocated its

Figure 3.6: Last Touch Attribution vs. Shapley Value



budgets and significantly lowered its cost per converter. We are currently collecting the data on the behavior of the publishers given this change in attribution method, to be analyzed in the future.

3.8 Conclusion

As multi-publisher campaigns become more common and many new publisher forms appear in the market, attribution becomes an important process for large advertisers. The more publishers are added to a campaign, however, the more complex and prone to errors the process becomes. Our two-publisher model has identified two issues that are detrimental to the process – free-riding among team members and baseline exploitation. This measurement issue arises because the data does not allow us to disentangle the effect of each publisher accurately and using statistics to estimate this effect gives rise to free riding. Thus, setting an attribution mechanism that does not take into account the equilibrium behavior of publishers will give rise to moral hazard even when the actions of the publishers are fully observable. On the other hand, if the performance of the campaign is not explicitly used in the compensation scheme through an attribution mechanism, adverse selection cannot be mitigated and ineffective publishers will be able to impersonate as effective ones.

The method of last-touch attribution, as we have showed, has the potential to make CPA campaigns more efficient than CPM campaigns under some conditions. In contrast, attribution based on the Shapley value yields well behaved pure strategy equilibria that increase profits over last-touch attribution when the noise is not too small. Adding experimentation as a requirement to the contract does not lower the profits of the advertiser too much, and allows for collection of the information required to calculate the Shapley value, as well as

estimating the magnitude of the baseline.

The analysis of the model and the data has assumed homogenous consumers. If the population has significant heterogeneity, which is observed by the publishers but not by the advertiser, the marginal estimates will be biased downwards, as the publishers will be able to truly target consumers they can influence. Another issue that arises from the analysis is that publishers may have access to exclusive customers who cannot be touched by other publishers.

Exclusivity can be handled well by our model as a direct extension. In those campaigns where a publisher has access to a large exclusive population, it may be beneficial to switch from CPM to CPA campaigns, or vice versa, depending on the overlap of other populations with other publishers.

To handle the heterogeneity of the baseline and consumers, we propose two solutions. To understand whether the baseline estimation affects the results significantly, we can compare the Shapley Value estimates with and without the baseline. In addition, the data include characteristics of consumers which can be used to estimate the baseline heterogeneity, and control for it when estimating the Shapley Value. Propensity Score Matching is a technique that will allow matching sets of consumers who have seen ads to similar consumers who have not seen ads and estimate the baseline for each set. One issue with this approach is that consumer data may include thousands of parameters per consumer including demographics, past behavior, purchase history and other information. Our tests have shown that using regularized regression as a dimensionality reduction technique performs well in this setting, and work is underway to implement it with a matching technique.

This study has strong managerial implications in that it identifies the source of the attribution issue that advertisers face. Advertisers today believe that if they improve their measurement mechanism campaigns will become more efficient. This conclusion is only correct if the incentive scheme based on this measurement is aligned with the advertiser's goal. If it is not, like last-touch methods, the resulting performance will be mediocre at best.

A key message of this chapter is that performance based incentive schemes require a good attribution method to alleviate moral hazard issues. The observations that proper estimates of marginal contributions as well as a proof based mechanism can solve these issues when employed together creates a path for solving this complex problem and providing advertisers with better performing campaigns.

Chapter 4

Reducing Sample Sizes in Large Scale Online Experiments

4.1 Overview

Online experiments have gained popularity as a leading method of performing market research, measuring ROI and producing new software for startups, advertisers and other firms. The two factors contributing to the increased popularity are the reduced cost of producing different versions to experiment with and cheap access to large consumer populations through the Internet. As a result of this trend, a recent survey¹ of 2,500 online marketers has determined “Conversion Rate Optimization” to be the top priority for the coming years, while another recent survey² has determined that the most popular method for determining marketing activity effectiveness is running an A/B Test.

A/B tests are randomized controlled trials in which two versions of a treatment to be tested (A and B) are assigned to consumers arriving to a website or using an app. An action by the consumer is designated as the target result of the experiment, and is called a conversion. Some examples of such treatments are exposing the consumer to an ad or displaying a different version of a webpage. Examples of conversions are the purchase of a product, filling out a form or providing an email address.

When running these experiments, marketers are required to invest time in designing the experiment through not only producing the different versions to test, but also by considering treatment allocation, experiment run-time, sample sizes and statistical tests. This approach fuses the traditionally separate positions of creative directors with that of planners and media buyers into one position requiring more rigorous and detailed analysis of the experiment a-priori. The applied business literature supports this approach by stressing the importance of properly applying the scientific method to business experiments, with examples from

¹SaleForce ExactTarget “2014 State of Marketing” report at:
<http://content.exacttarget.com/en/StateOfMarketing2014>

²Econsultancy “Conversion Rate Optimization Report 2013” at:
<http://econsultancy.com/reports/conversion-rate-optimization-report>

both marketing (Anderson and Simester, 2011) and entrepreneurship (Blank, 2013). Consequently, marketers who run online experiments typically utilize a software platform such as Optimize.ly, Google Analytics or Adobe Test for experimental design.

Given the treatments to test, the software automatically allocates them to consumers, tracks conversions, produces reports and performs statistical tests for the marketer. The determination of the experiment's sample size and the execution of hypotheses tests are then relegated to the software.

This chapter discusses the standard online experimental design proposed by the leading online platforms and focuses on the current methods used for sample size determination and hypothesis testing. Since in many cases even small effect size have large economic value in terms of profit for a website, experimenters set ambitious goals for detection of effects which may result in large sample requirements. The consequences of these large samples implies that experiments need to run for a long period of time until reaching the desired population and in many cases may result in an inability to properly measure the effect of a campaign due to large signal to noise ratios of the data as documented in Lewis and Rao (2012b) and Lewis et al. (2013).

The question arises, however, whether the standard test used by testing platforms, the statistical test of inequality of conversion rates with fixed sample size, is the most efficient test that can be used in an online setting. The intuition behind this question stems from three insights about online experiments. The first is that the goal of the experiment, or the decision to be taken given the result, impacts the required sample size to make the right decision. The second is that the data collected in an experiment is stochastic and collected sequentially, and the variation in it can be exploited. The third is that a-priori, the experimenter many times makes a best-effort guess regarding the underlying effect sizes to be determined, but it may turn out that the treatments are much worse or better than previously hypothesized. We therefore focus on several techniques an experimenter can use to lower the sample size needed in an experiment, either through matching the goal of the experiment with the statistical test used, or through using sequential analysis methods.

When calculating the required sample sizes in an experiment, typically two types of parameters are taken into account. The first type of parameters describe the desired effect size to be detected and many times a baseline conversion rate to detect the change from. For example, an experiment's goal may be to detect an increase of at least 10% in conversion rate above a 5% conversion rate. That is, we wish to detect treatments with conversion rates above 5.5%. The other set of parameters set the acceptable error rates of the decision process, namely the maximum levels for Type I and Type II error rates.

Typically there are two possible goals for an online experiment: to select the best treatment or to determine whether a new treatment is better than a control. We call the former a selection test and the latter a test of superiority, the difference depending on the cost of picking treatment A over B for eventual use. If both treatments A and B have the same cost of implementation and the goal is to pick the best one, we call the experiment a selection experiment. If switching from the control treatment to a new one will bear some additional cost, than the experimenter would like to make sure the new treatment outweighs

this cost by showing that the new treatment is superior to the control. We therefore label this experiment a test of superiority.

Section 4.2 formally introduces the distinction between the possible types of experiment goals and develops an understanding on the impact these have on reducing the possible sample sizes in experiments. We show that although the goal of most online experiments is to test superiority or select the best treatment, the current practice overestimates the sample size required for these tasks as it uses a test of equality. As an example, when selecting the best version out of two, it many times does not matter which version is selected if their performance measures are close enough. This indifference zone for experiment results effectively eliminates Type I errors and allows for achieving sample size requirements lower by more than 80% compared to the standard test.

In Sections 4.3 and 4.4 the concept of sequential statistical tests is introduced and applied to relevant online questions. The concept of sequential analysis was introduced by Wald et al. (1945) who developed the sequential ratio probability test (SPRT) during World War II to test the improved performance of anti-aircraft guns. The main idea behind the test is to use a sequence of likelihood ratios for the data to reject or accept the null hypothesis while an experiment is running when an extreme result occurs. Since its initial development, significant work has been done in the field of sequential analysis with the majority of applications carried out for medical trials. Ghosh and Sen (1991) contains a classical overview of the developments of the different tests, while Bartroff et al. (2012) contains a more modern treatment with focus on medical experimentation.

It is interesting to note that to the best of our knowledge sequential methods have seldom been used in the fields of Business, Economics and Psychology, and seem completely non-existent among marketing practitioners. Part of this puzzle can be attributed to the technical nature of the statistics required to perform the tests, as well as the abundant number of methods that may apply to a specific scenario. Another reason may be the lack of accessible software libraries for carrying out the experimental design and performing the statistical tests. This chapter therefore aims to synthesize the available literature and methods into a coherent overview that can serve as a guideline for applying sequential techniques for marketers. To this end, Sections 4.3, 4.4 and 4.5 review the most common scenarios online marketers may encounter in their experiments and carefully describe the applicable statistical approaches that can be used. A software library that has been developed while writing this chapter will be distributed on the author's site with the goal of making these techniques accessible and easily applicable.

Section 4.6 describes the application of a sequential test on observational data from an online experiment carried out by a software startup. The goal of the experiment was to show whether the firm's software has superior efficacy to the current best practice of a marketing website. As the results show the sequential test determines that the experiment has achieved its desired goal within 12 days, which would have allowed to reduce the length of the experiment by approximately 25% compared to the original a-priori determined sample size.

Lastly, Section 4.7 describes additional open questions and possible future avenues for

research in the application of sequential tests for marketing purposes.

4.2 Tests of Equality, Superiority and Selection

An A/B test is a simple online experiment that proceeds as following. Consumers arriving at a website are randomly assigned to one of two treatments, A and B . A conversion is counted as an action taken by a consumer exposed to one of those version, *e.g.* the purchase of a product, sign-up to a subscription or click on an ad, to name a few. At the end of the test, the conversion rate is calculated as the number of successful conversions divided by the size of the exposed population for each version. A statistical test is then performed to determine whether version A has outperformed B or vice versa. The parameters of the test, such as the acceptable error rate and target size of effect to identify determine the sample size of the experiment.

Formally, let $2n$ visitors arrive at a website sequentially, and assume n of these consumers are randomly assigned to treatment $i \in \{A, B\}$. Let $x_{ij} \in \{0, 1\}$ denote the response of consumer $1 \leq j \leq n$ to treatment i . A value of $x = 1$ denotes a conversion. We assume the consumers are independent and that the probability of conversion is p_i for all consumers exposed to treatment i . Denote $\hat{p}_i = \bar{x}_i = \frac{\sum_{j=1}^n x_{ij}}{n}$ the MLE estimate of p_i , which is the average observed conversion rate of each treatment.

We differentiate between three types of goals the experiment may have: determining inequality, superiority or selection.

A test of *equality* tests the hypothesis $H_0 : p_A = p_B$ against the alternative hypothesis $H_1 : p_A \neq p_B$. This is the typical test prescribed by online experimentation platforms, but as can be seen, it does not determine whether treatment A is better than B or vice-versa.

A test of *superiority* tests the hypothesis $H_0 : p_A \leq p_B$ against the alternative hypothesis $H_1 : p_A \geq p_B(1+d)$, with d being the minimal effect size the test wishes to detect. Compared to a test of inequality, it does help determine which treatment is better, and setting $d = 0$ is possible.

Finally, a *selection* experiment attempts to pick the treatment with highest conversion rate with high probability. The test simply selects the treatment with highest estimated conversion rate, \hat{p}_i , as the better treatment. The test has an indifference zone d , a value where if $\frac{1}{1+d} < \frac{p_A}{p_B} < 1+d$, the experimenter is indifferent among selecting one of the two options.

Sample Sizes for Fixed Size Tests

Statistical tests in which the sample sizes are determined in advance prior to the experiment and in which the test is performed following the data collection are called fixed sample size tests. The parameters of the test that determine the sample size are the maximum acceptable Type I and Type II error levels α and β , and the effect size at which type II error will be

controlled. For each of the tests, we let $d > 0$ denote the effect size the experimenter wish to detect if the alternative hypothesis is true.

Following are two examples that illustrate the difference between the three types of tests and the use of the parameters:

Example 1. *A startup has produced a website plug-in that attempts to increase the conversion rate of consumers subscribing to a magazine by offering more content to the consumers visiting the site. The company running the magazine would like to purchase the software assuming it increases conversion rate by at least 10% over the current baseline rate of the standard site. The company wishes to buy the software with 90% probability if the improvement is at least 10%, but not buy it with 95% probability if the software actually worsens the situation. This is a superiority test with $H_0 : p_A \leq p_B$ vs. $p_A \geq p_B(1 + d)$ with $d = 0.1$ and $\alpha = 0.05$, $\beta = 0.1$.*

Example 2. *In an online advertising campaign, the advertiser wishes to determine which of two different ad creatives, A and B, performs better. The advertiser would like to select the highest converting ad with 90% probability assuming the conversion rate is at least 10% higher than the other ad. If the difference in conversion rates is less than that, the advertiser is indifferent regarding which ad to use.*

This experiment is a selection experiment, to select the best of two versions, with $d = 0.1$ and $\beta = 0.1$. Each consumer is exposed to either ad A or B but not both. The creative that achieves the highest observed conversion rate is declared as best.

These examples show that in both of these common cases, a test of equality is not required yet the most common technique is to use one. The difference in required sample sizes for these tests is dramatic as will be shown below.

For each of exposition we use the Normal approximation to the binomial distribution for large sample sizes. We also focus on the difference of the conversion rates $\hat{p}_A - \hat{p}_B$ instead of its ratio $\frac{\hat{p}_A}{\hat{p}_B}$ as this is the typical measure prescribed by current testing platforms. When n is large, $\hat{p}_A - \hat{p}_B$ is approximately distributed $N(0, \frac{2p(1-p)}{n})$ under the null hypothesis of $p_A = p_B = p$.

Standard calculations show that the sample size for a test of inequality with level α and power $1 - \beta$ at $p_A - p_B = d$ is approximately:

$$n_{NEQ} = 2p(1-p) \left(\frac{\Phi^{-1}(1 - \alpha/2) + \Phi^{-1}(1 - \beta)}{d} \right)^2 \quad (4.1)$$

when Φ is the cdf of the Normal distribution.

Similarly, for a superiority test with the same parameters, the sample size equals approximately:

$$n_{SUP} = 2p(1-p) \left(\frac{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)}{d} \right)^2 \quad (4.2)$$

The case of the selection experiment is more interesting. As the goal defines, when the two treatments are close enough ($|p_A - p_B| < d$), the experimenter is indifferent among which treatment is selected as best. Effectively, this test has no Type I error, and we only need to make sure that the test selects the highest performing treatment with high probability. Assume, w.l.o.g that $p_A \geq p_B + d$, then the test that picks treatment A if $\hat{p}_A > \hat{p}_B$ has the following (lowest) power when $\theta = d$, termed the probability of correct selection:

$$Pr(\text{Correct Selection}) = Pr(\hat{p}_A - \hat{p}_B > d | p_A - p_B = d) = 1 - \Phi\left(\frac{-d}{\sigma}\right) = 1 - \beta \quad (4.3)$$

Solving for n yields

$$n_{SEL} = 2p(1 - p) \left(\frac{\Phi^{-1}(1 - \beta)}{d} \right)^2 \quad (4.4)$$

Calibrating with typical values of $\alpha = 0.05$ and $\beta = 0.1$, we can observe that the test of superiority requires $\sim 81.5\%$ of the sample size required by the test of non-equivalence, while the selection experiment only requires $\sim 15.6\%$ of the sample size. The improvement does not depend on the effect size d or the baseline rate p . This is a dramatic six-fold improvement in sample size that is achieved when running selection tests, which are very common for new products.

The caveat in this comparison is that the absolute sample size required to detect small effects compared to the baseline effect p may be very large compared to the timing constraints of an experiment. Figure 4.1 displays the required sample sizes to detect a 10% increase for different values of conversion rate p .

For a website with a thousand daily visitors willing to spend two weeks on an experiment, there are not enough visitors to detect even a 15% increase in conversion rate using a superiority test. This led Lewis and Rao (2012b) and Lewis et al. (2013) to provide convincing evidence that measuring small yet economically meaningful effects may be very hard for smaller firms and advertisers without access to extremely large populations.

Apart from using a test of inequality for superiority and selection experiments, the current approach has two noticeable deficiencies. The first deficiency is that the effect size, d , is selected as the difference of the two treatments regardless of the underlying baseline p of the experiment. The a-priori determined sample size is therefore highly sensitive to the specification of the baseline p . If p diverges significantly from the pre-specification, the test may end as being too powerful or not powerful enough to make a decision.

The second deficiency is the fact that the standard test is powered at a specific effect size, and does not adjust the sample size when the true effect is much larger or much smaller than hypothesized. The fact that consumers arrive sequentially over time to be included in the experiment means that early information about a substantially large or small effect size may be used to terminate the experiment early with the correct conclusion. The method behind this intuition is known as Sequential Analysis and will be introduced in the next section.

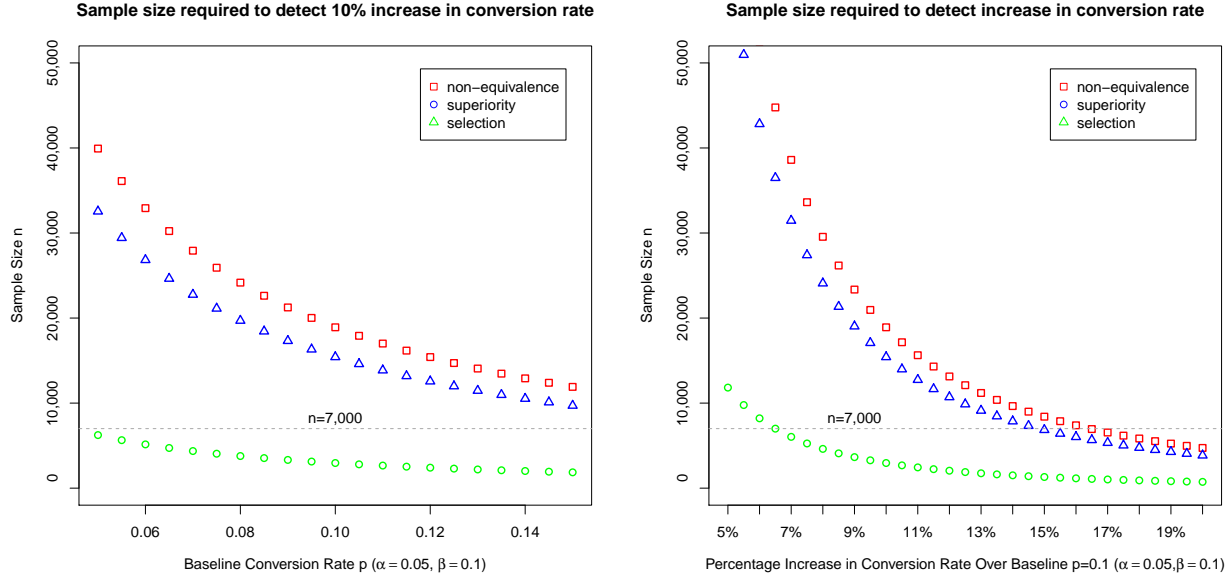


Figure 4.1: Minimum Required Sample Sizes

Left - detect 10% increase in conversion rate over different baselines. Right - detect various increases in conversion rates at a 10% baseline rate.

4.3 Sequential Analysis for Tests of Superiority

When the data about the experiment arrives sequentially, the experimenter can decide to stop the experiment early if the data provides enough evidence to accept or reject H_0 , at the expense of lowering the power and possibly the significance level of the test for a fixed sample size.

For example, it is well known that the naïve procedure of repeatedly testing for significance as n increases using the standard fixed test statistic has a much higher type I error than a single use of the same test. As a result, special care must be given to use tests that take into account the possibility of accumulating Type I and Type II error over time as the test progresses.

Wald et al. (1945) have introduced the Sequential Ratio Probability Test (SPRT) as a procedure to test the simple hypothesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ when X_i is i.i.d. with density $f(x_i|\theta)$. The test makes use of the log likelihood-ratio of the data for $x = (x_1, \dots, x_n)$ which is $LLR_n(x|\theta_0, \theta_1) = \sum_{i=1}^n \log \left(\frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} \right)$.

The test proceeds as follows:

1. For each new data point x_n , calculate $LLR_n(x|\theta_0, \theta_1)$.
2. If $LLR_n \leq \underline{c} = \log \left(\frac{\beta}{1-\alpha} \right)$ stop and accept H_0 .

3. If $LLR_n \geq \bar{c} = \log\left(\frac{1-\beta}{\alpha}\right)$, stop and accept H_1 .
4. Otherwise, when $\underline{c} < LLR_n < \bar{c}$, continue the test and take another sample.

To understand the intuition behind the test, Figure 4.2 graphs the log likelihood-ratio values of two random samples generated from Bernoulli distributions with parameters $p = 0.08$ and $p = 0.12$. The boundaries \underline{c} and \bar{c} are marked using the dashed lines. The values of the LLRs constitute a random walk. When each of the lines crosses one of the boundaries, the experiment is stopped and H_0 or H_1 is accepted. In the example's case, testing $H_0 : p = 0.1$ vs. $H_1 : p = 0.11$ would have required a sample size of 9,976 samples to achieve Type I error of $\alpha = 0.05$ and Type II error rate of $\beta = 0.1$. The simulation graphs show, however, that less than 2,000 draws were required to stop and accept H_1 for $p = 0.12$ and less than 1,100 draws were required before stopping to accept H_0 with $p = 0.08$.

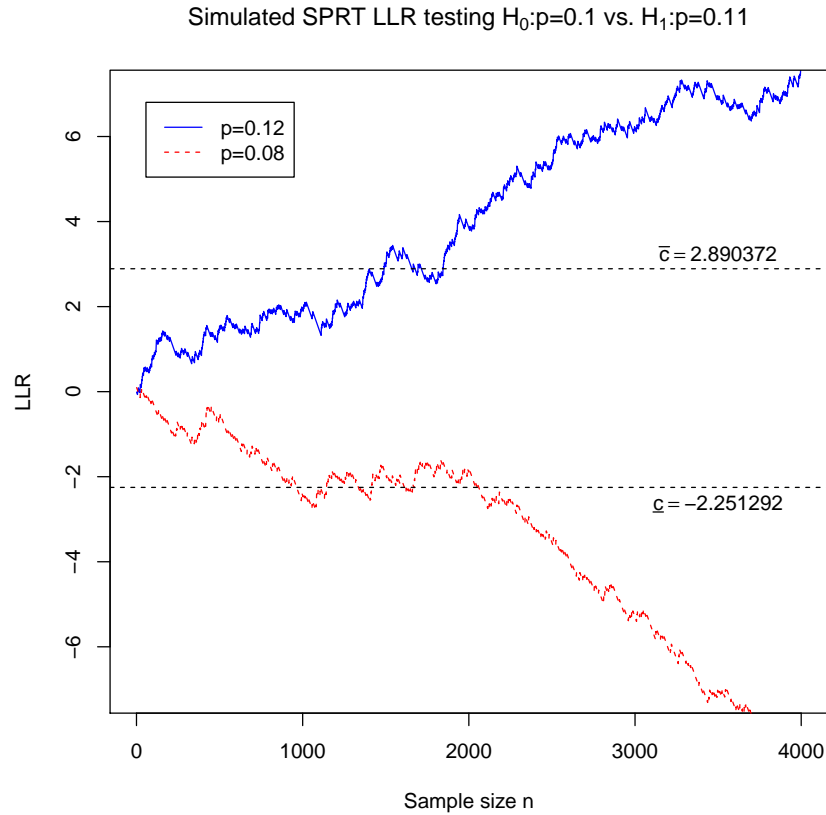


Figure 4.2: Simulated result of SPRT testing $H_0 : p = 0.1$ vs. $H_1 : p = 0.11$

The SPRT has been shown to approximately achieve a Type I error rate of α and a Type II error rate of β . As can be noticed from its description, the actual stopping time, denoted N , of the test is a random variable and depends on the arriving data. In theory, the experiment

may proceed indefinitely yielding very large sample sizes. Consequently substantial amount of research has been dedicated to minimize the expected sample size $E[N]$ of the test under various assumptions, while providing a definite upper bound for the maximum sample size of the procedure.

One common solution is to choose an upper bound N_{max} for the sample size n , and truncate the experiment if this bound is reached. At this point, H_0 is accepted if $LLR_{N_{max}} < 0$ and H_1 is accepted if $LLR_{N_{max}} > 0$, while a tie is broken arbitrarily. Another option is to design the stopping boundaries \underline{c} and \bar{c} to change with n , and converge to meeting at N_{max} .

An excellent overview of the research on the topic can be found in Ghosh and Sen (1991); Jennison and Turnbull (1999); Bartroff et al. (2012). The majority of the applications of sequential analysis to date have been the medical trials field, where the cost of experimentation is high and ethical issues about adverse effects of drugs raise the need to stop experiments early.

Except for the possibly unlimited sample size, a major issue with applying the standard SPRT to comparing to Bernoulli trials is that the test applies only to simple hypotheses of single parameter distributions, and most extensions apply to the exponential family of distributions, to which the two-armed bernoulli trial does not belong. In our case of conversion rate comparison, however, we are interested in testing composite hypotheses $H_0 : p_A \leq p_B$ vs. $H_1 : p_A \geq p_B(1 + d)$. These hypotheses are composite since the underlying values of p_A and p_B are unknown in advance. As a result, we are interested in making use of the observed data to estimate p_A and p_B which may bring the ability of stopping earlier for very high or very low values of p_A .

We present two modified versions of sequential tests based on Hoel et al. (1976) and Kulldorff et al. (2011) which take a maximum likelihood approach to the estimation of the test statistic. The first procedure, the equality constrained maximum likelihood ratio SPRT (eqMaxSPRT), uses a two sided boundary and allows for early stopping to both accept H_0 or H_1 . This approach is useful to minimize the expected sample sizes of the experiment when H_0 and H_1 are true with equality, and in addition to stop early with high power when $p_A \gg p_B$ or when $p_A \ll p_B$.

The second approach, which we term ineqMaxSPRT uses an inequality constrained maximum likelihood estimator for the test statistic, and only an upper bound \bar{c} for stopping. In other words, $\underline{c} = -\infty$, and the test will never stop early to accept H_0 . The test is truncated at a maximum sample N_{max} , at which point H_0 is accepted. This test is appropriate when there is high probability that H_1 is correct, or high cost to not stopping when H_1 is correct and the experimenter would like to stop as early as possible when this is true. Otherwise, if both treatments are equivalent, the cost of continuing to experiment should be low. This test, for example, can be used to detect a new treatment that is significantly worse than a control by properly reversing the hypotheses or counting non-converters as converters.

Equality Constrained MaxSPRT

We first notice that the set of hypotheses $H_0 : p_A = p_B$ vs. $H_1 : p_A = p_B(1+d)$ is composite, as there are sets of values for p_A and p_B that can satisfy these. The log-likelihood of the data with $s_A = \sum_j x_{Aj}$ and $s_B = \sum_j x_{Bj}$ conversions for arms A and B respectively, given values p_A and p_B of the true conversion rates and n samples from each arm is:

$$LLR_n(s_A, s_B | p_A, p_B) = s_A \log(p_A) + (n - s_A) \log(1 - p_A) + s_B \log(p_B) + (n - s_B) \log(1 - p_B) \quad (4.5)$$

To calculate the eqMaxSPRT statistic, we maximize the log-likelihood under each hypothesis, to receive the log likelihood-ratio:

$$eqMaxLLR_n(s_A, s_B) = \max_{p_A=p_B(1+d)} LLR_n(s_A, s_B | p_A, p_B) - \max_{p_A=p_B} LLR_n(s_A, s_B | p_A, p_B) \quad (4.6)$$

The left-hand additive estimates the log likelihood of the data under H_1 by solving the quadratic equation resulting from the first order condition to maximize the constrained likelihood. The right-hand additive solution results in an estimate of $\hat{p}_A = \hat{p}_B = \frac{s_A + s_B}{2n}$.

To show the applicability of this test, we simulated 2,000 experiments for different values of p_A when $p_B = 0.1$ and $d = 0.1$, with up to 50,000 draws in each experiment. For each experiment we calculated the eqMaxSPRT statistic for each state, and determined the stopping time of the experiment using the boundaries \underline{c} and \bar{c} described above, with $\alpha = 0.05$ and $\beta = 0.1$. Table 4.1 shows the probability of rejecting H_0 and the probability of not having stopped by using n samples with $n = 10,000$, $n = 30,000$ and $n = 50,000$ draws per arm.

p_A	n=10,000		n=30,000		n=50,000	
	Prob. Reject	Prob. No	Prob. Reject	Prob. No	Prob. Reject	Prob. No
	H_0	Stop	H_0	Stop	H_0	Stop
0.08	0	0	0	0	0	0
0.09	0.0005	0.0045	0.0005	0	0.0005	0
0.1	0.0305	0.256	0.0435	0.0105	0.0435	0.0005
0.105	0.22	0.482	0.4345	0.0695	0.4655	0.012
0.11	0.5935	0.331	0.8915	0.01	0.9015	0
0.12	0.988	0.0105	0.9985	0	0.9985	0

Table 4.1: Simulation results of probability of rejecting H_0 and of not stopping eqMaxSPRT

$$p_B = 0.1, d = 0.1, \alpha = 0.05 \text{ and } \beta = 0.1.$$

As expected, the power of the test increases as p_A gets farther below 0.1 or above 0.11. In addition the longer the maximum sample sizes n , the higher the power of the test. Finally, for the range of values between $p_A = 0.1$ and $p_A = 0.1 \times (1 + 0.1) = 1.1$, we see that the probability of not stopping early is highest, as the test statistic is not powerful enough to discriminate at these rates. This is the same phenomenon that would happen with a fixed size test for values of $\theta_0 < \theta < \theta_1$.

Analyzing the expected sample size until stopping $E[N]$, however, sheds light on the advantages of using the sequential method. Table 4.2 summarizes the expected sample sizes

until stopping to accept H_0 or H_1 for the series of simulated experiments described above. The values in parenthesis show the ratio between the expected sample size $E[N]$ to the one-arm sample size of 16,094 required to test $H_0 : p_A < p_B$ vs. $H_1 : p_A \geq p_B$ for $p_B = 0.1$, with $\alpha = 0.05$ and $\beta = 0.1$ when $d = \frac{p_A}{p_B} = 0.1$.

p_A	$E[N]$	$E[N]$	$E[N]$
	n= 10,000	n= 30,000	n= 50,000
0.08	1,778 (0.11)	1,778 (0.11)	1,778 (0.11)
0.09	2,917 (0.18)	2,952 (0.18)	2,952 (0.18)
0.1	4,942 (0.31)	7,678 (0.48)	7,964 (0.49)
0.105	5,571 (0.35)	10,595 (0.66)	12,151 (0.75)
0.11	5,599 (0.35)	8,804 (0.55)	9,086 (0.56)
0.12	3,569 (0.22)	3,656 (0.23)	3,656 (0.23)

Table 4.2: Simulation results of expected sample sizes $E[N]$ of eqMaxSPRT

$$p_B = 0.1, d = 0.1, \alpha = 0.05 \text{ and } \beta = 0.1.$$

As can be noticed, when the data is drawn from distributions with $p_A \geq p_B(1+d)$ or with $p_A \leq p_B$, the expected sample size is substantially smaller than the fixed sample test size, leading to improvement of 40% and more. This feature of the test, being more powerful and requiring smaller samples with more extreme data farther outside the indifference zone is a result of the monotonicity of the estimate of the constrained \hat{p}_A in the number of successes s_A , which is itself monotone in the true p_A in expectation. When the test is less powerful, however, for values of $p_B < p_A < p_B(1+d)$, the expected sample size increases and may reach values close to the fixed sample test that will not warrant the expense of having the possibility of not stopping the test.

The problem of minimizing the maximum expected sample size $E[N]$ is known as the Kiefer-Weiss problem (Kiefer et al., 1957). Several applicable solutions to a slight variation of this problem for exponential families have been developed, a good example of which appears in Huffman (1983). These solutions unfortunately do not apply to the case of comparing two Bernoulli populations with composite hypotheses.

Inequality Constrained MaxSPRT

In several cases it may be desired to stop the test early only if H_1 is to be accepted, while waiting until a fixed value of n is reached to reject H_0 . This, for example, can be the case when a new version of software is being installed and monitored for failures compared to the control. As long as the failure rate is not worse than the control, there is no reason to revert to the old version. However, if the failure rate (non-conversion in this case) is high, it may be desired to revert. Another case is when testing a new version of product for which each test is expensive. If it is desired to stop as early as the new version proves superior to the old version, it is possible to increase the power of the test for positive results with lower samples at the expense of later stopping to reject H_0 .

The inequality constrained MaxSPRT calculates the test statistic:

$$ineqMaxLLR_n(s_A, s_B) = \max_{p_A \geq p_B(1+d)} LLR_n(s_A, s_B | p_A, p_B) - \max_{p_A = p_B} LLR_n(s_A, s_B | p_A, p_B) \quad (4.7)$$

The other major difference is that the test only has an upper bound for early stopping \bar{c} and a maximum sample size N_{max} chosen such that:

$$\sum_{j=1}^{N_{max}} Pr(ineqMaxLLR_n(s_A, s_B) > \bar{c} | H_0) \leq \alpha \quad (4.8)$$

$$\sum_{j=1}^{N_{max}} Pr(ineqMaxLLR_n(s_A, s_B) > \bar{c} | H_1) \geq 1 - \beta \quad (4.9)$$

In their novel development of the inequality constrained MaxSPRT, Kulldorff et al. (2011) compare two-armed problems whose distributions can be reduced to single parameter distributions. As a result they could use numerical integration or exact calculations to find N_{max} and \bar{c} . For our purpose, we can achieve similar results using simulation methods that estimate $Pr(ineqMaxLLR_n(s_A, s_B) > \bar{c})$ under H_1 and H_0 to find B and N_{max} .

Table 4.3 shows the probability of rejecting $H_0 : p_A = p_B$ under different conditions. Comparing to Table 4.1 we can see that higher power is achieved for smaller sample sizes with the inequality constrained MaxSPRT. Since the test is built using simulation techniques, however, the error rate guarantees are only approximate, as can be seen for the values of $p_A = 0.1$, which reaches a Type I error of 0.0565 for a large sample. In our experiments these values fluctuated between 0.02 and 0.07.

p_A	n= 10,000	n= 30,000	n= 50,000
0.08	0.005	0.0045	0.0045
0.09	0.014	0.0105	0.0105
0.1	0.0685	0.0575	0.0565
0.105	0.1685	0.3395	0.456
0.11	0.489	0.9165	0.9825
0.12	0.976	1	1

Table 4.3: Inequality MaxSPRT probability of rejecting H_0

$p_B = 0.1$, $d = 0.1$, $\alpha = 0.05$ and $\beta = 0.1$.

Another characteristic of the inequality constrained MaxSPRT is its expected sample size behavior. A key feature is that rejecting H_0 will always require the maximum sample size, while rejecting H_1 might come at much earlier stages. Table 4.4 documents simulation results for $E[N]$ for various conditions and should be compared with Table 4.2. As can be seen, for values of p_A much lower than p_B or higher than $p_B(1 + d)$ the expected sample sizes can be much smaller for the inequality constrained version. Closer to these values, however, the expected sample size might inflate above the original fixed sample test size. This undesirable property should be taken into account when choosing which test to use. Comparing to the

equality constrained version, however, we notice that the inequality constrained MaxSPRT allows testing $d = 0$ vs. $d > 0$, while the equality constrained version cannot test this assumption. Therefore, by setting $d = 0$ for the inequality test, one can avoid the inflated sample sizes, and achieve a powerful test for detecting smaller effects of the treatment.

p_A	$E[N]$	$E[N]$	$E[N]$
	n= 10,000	n= 30,000	n= 50,000
0.08	78 (0.00)	87 (0.01)	88 (0.01)
0.09	519 (0.03)	693 (0.04)	696 (0.04)
0.1	1,588 (0.10)	3,804 (0.24)	4,016 (0.25)
0.105	3,573 (0.22)	13,114 (0.81)	19,938 (1.24)
0.11	4,555 (0.28)	11,058 (0.69)	12,845 (0.80)
0.12	3,098 (0.19)	3,498 (0.22)	3,520 (0.22)

Table 4.4: Inequality MaxSPRT expected sample size

$p_B = 0.1$, $d = 0.1$, $\alpha = 0.05$ and $\beta = 0.1$. Values in parenthesis are the ratio of the expected sample size to the fixed sample test size.

4.4 Sequential Analysis for Tests of Selection

As section 4.2 noted, when selection is the major goal of a procedure, there are no Type I errors and there is an indifference zone in which both treatments are considered equivalent. An intuitive sequential procedure to terminate sampling is when one arm is successful enough compared to the other, and the number of future remaining samples is such that the disadvantaged arm will not be able to “catch up” with it. A summary of this procedure and its properties appears in Bechhofer (1985), which we call the curtailed difference procedure.

Formally, if the n draws were made from each arm, yielding s_i successes, and if N_{max} is the maximum allowed sample size, the following procedure makes the same decision as the fixed sample test from Section 4.2

- If $s_i > s_{-i} + N_{max} - n$, declare i as the best arm.
- Otherwise, continue sampling until $n = N_{max}$ and declare the arm with the highest value of s_i as the best.

It is easy to show that this procedure will reach the same decision as the fixed sample procedure, hence reaching the same probability of correct selection (Power). On the other hand, this procedure has the advantage of being able to stop earlier if one of the arms turns out to be much better than the other. Another advantage of this procedure is that the possible states for $s_i - s_{-i}$ after n samples are discrete and finite ranging from $-n$ to n . We can therefore exactly calculate the probability of stopping and probability of error given this procedure.

Although the curtailed difference procedure has the advantage of a deterministic stopping time, calculations show that the decrease in expected sample sizes is moderate at best. The

reason is that the procedure itself does not make use of the indifference zone to relax its requirement for stopping.

We therefore propose to use the equality constrained MaxSPRT procedure to test the hypothesis: $H_0 : p_A = \frac{p_B}{1+d}$ vs. $H_1 : p_A = p_B(1 + d)$. If H_0 is accepted, we choose arm 2 as the best, while if H_1 is accepted, we pick arm 1 as the best. The procedure is similar to the one described in Section 4.3 with a different null hypothesis and with setting $\alpha = \beta$. This ensures that the probability of correct selection is $1 - \beta$ as desired.

p_A	Prob. Correct Selection	Prob. No Stop	E[N]
0.08	0.9965	0	1,067 (0.36)
0.09	0.918	0.016	1,700 (0.58)
0.1	1	0.056	2,050 (0.69)
0.105	0.723	0.0415	1,928 (0.65)
0.11	0.9005	0.0095	1,639 (0.55)
0.12	0.9825	0	1,003 (0.34)

Table 4.5: Simulation results for eqMaxSPRT test of Selection.

$$n = 6,000, \beta = 0.1, d = 0.1, p_B = 0.1.$$

Table 4.5 summarizes the results of using this procedure to pick the highest among p_A and p_B , with p_B fixed at 0.1, for various values of p_A , and an indifference zone of 10% ($d = 0.1$). The fixed sample size test requires $n = 2,957$ from each arm to reach a 90% probability of correct selection. As can be seen, setting $n = 6,000$ for the sequential test yields a high probability of stopping early with the required probabilities of correct selection. The major advantage of this approach is the lowered expected sample size compared to the curtailed difference approach. The results in the table show that the expected sample sizes go as low as 34% of the fixed test size, and do not surpass 70%.

4.5 Extensions

Unknown Sample Sizes

Various online scenarios do not allow for determining the total sample size of exposed consumers, but rather provide information only on the number of converters resulting from the different exposures. As an example, when running an online advertising campaign to test two ad creatives, a budget is allocated to a network that will be used to display both versions of the ad. The network can guarantee a certain ratio of ad displays between the versions (e.g. 1:1 or 1:2), but cannot guarantee, and many times cannot determine the number of consumers exposed to the ads, as exposures are done on an impression by impression basis. Another example is using a search advertising campaign that gives the number of impressions per ad, but not the number of consumers exposed to each ad.

In such cases it is impossible to tell how many of the treatments resulted in non-conversions (failures). It is possible, however, to test an approximate hypothesis about

the ratios of p_A and p_B assuming their values are small and the samples are large. Suppose that a sample of n consumers was exposed to treatment 2, and that the ratio of sample sizes between treatment 2 and 1 is z . That is, n/z consumers were exposed to treatment 1. The expected number of successes from treatment 1 is $(np_A)/z$, and its probability is $Pr(\text{Converter from 1}) = \frac{np_A}{np_A + nzp_B}$. Dividing by np_B and letting $r = \frac{p_A}{p_B}$ we receive $Pr(\text{Converter from 1}) = \frac{r}{r+z}$.

Thus, we can consider the allocation of converters among the treatments as a Bernoulli variable with parameter $\frac{r}{r+z}$. Testing for $H_0 : r = 1$ vs $H_1 : r = 1 + d$ using the previously described tests will allow to test both superiority as well as performing selection without knowing n .

The disadvantage in this approach is that the Bernoulli approximation is not exact and loses the information from non-converters that could be used to better distinguish the likelihood-ratios (Cook et al., 2012). As a result, reaching the same powerful test will require a larger number of converters and a total sample size. Other approaches which may be beneficial include matching, where only instances of paired results of $(x_A, x_B) = (\text{success}, \text{failure})$ and $(x_A, x_B) = (\text{failure}, \text{success})$ are being collected from the data. The interested reader is directed to Wald et al. (1945) and to Tamhane (1985) for details.

Data Grouping

Performing a continuous sequential test with every additional observation is many times undesirable and sometimes impossible. For example, many online tracking systems provide only aggregated daily values for the number of exposures and converters in an experiment.

This restriction is a blessing in disguise as the limited number of tests to perform increases the power of the test and lowers the potential error rate. Suppose the samples arrive in K groups (*e.g* days), each of size n_k , with s_{Ak} and s_{Bk} the number of success in group k . Each s_{ik} is then distributed binomially with parameter p_i , and its probability mass function can be calculated exactly. The standard eqMaxSPRT statistic can be used, but the limits \bar{c} and \underline{c} can be adjusted to be less conservative and allow for earlier stopping. The theory for group sequential tests is well developed in Jennison and Turnbull (1999), and the R software library gsDesign by Merck Corp. implements these techniques.

4.6 Application

To illustrate the use of the equMaxSPRT test, we apply the technique to data collected by Reactful.com in an experiment to compare the efficacy of their software on a branding website. Reactful.com is a startup producing add-on software for websites that allow the website to “react” dynamically to customer visits to enhance user experience. As an example, the add-on may detect confusion by the consumer evident in its mouse hovering between too many options and react with a pop-up screen to suggest an explanation. Another example

is detecting when a consumer is about to enter a purchasing process and suggesting more information to the consumer to help them make a decision.

The experiment was run for 14 days in January 2014 on a branding website (known as a “mini-site”) whose goal was to educate consumers about a new product on the market of a well known CPG brand. The conversion was defined as the event of a consumer ordering a free trial of the product.

The experiment was set-up so that consumers were randomly allocated with equal probability to a static version of the mini-site (Control) and the dynamic version using Reactful (Treatment). The goal was to show superiority of the Reactful product over the control with at least 10% improvement. Every day data was collected about the total number of visitors for each version of the site (n_A and n_B), version A being the treatment and B the control. In addition, the number of converters under each version was counted.

The data is displayed in Table 4.6. Although the software assigned each consumer to a treatment randomly with equal probability, the samples are not balanced. In addition, the control version had an eventual overall 11.4% conversion rate.

Day	n_A (Reactful)	s_A (Reactful)	n_B (Control)	s_B (Control)
1	661	102	681	85
2	1,044	136	1,030	124
3	651	58	691	50
4	1,108	84	1,102	68
5	1,111	117	1,007	102
6	737	126	719	105
7	973	145	923	134
8	1,527	194	1,531	195
9	804	108	741	93
10	471	67	533	77
11	778	109	739	85
12	671	87	689	83
13	547	63	474	52
14	558	69	573	49
Total	11,641	1,465	11,433	1,302
Total Conversion Rate		0.126		0.114

Table 4.6: Results of Reactful.com experiment on CPG trial website

The fixed sample size for a test of superiority able to detect a 10% increase with $\alpha = 0.05$ and $\beta = 0.1$ is 13,886 samples per arm. This is less than the sample sizes actually achieved in the experiment, and would have required approximately 2,500 additional samples for each arm. The question is whether a sequential approach could determine if the treatment is superior to the control.

As the data is daily aggregate data and not continuous, the standard SPRT techniques should be modified to handle grouped data as discussed in Section 4.5. We should note, however, that if at any point the group data crosses one of the boundaries for stopping, it would have crossed that boundary in a possibly earlier time for a continuous test. Thus, we can tell whether the test should have been stopped earlier or not. In addition, if we assume that the test statistic can be approximated by straight lines between the group points, we can use the standard eqMaxSPRT to decide if to accept H_0 or H_1 .

Figure 4.3 compares the test statistic values to the stopping boundaries of the test. As can be seen, the test could have been stopped after day 12 with a conclusion that Reactful’s software increases conversions by at least 10%.

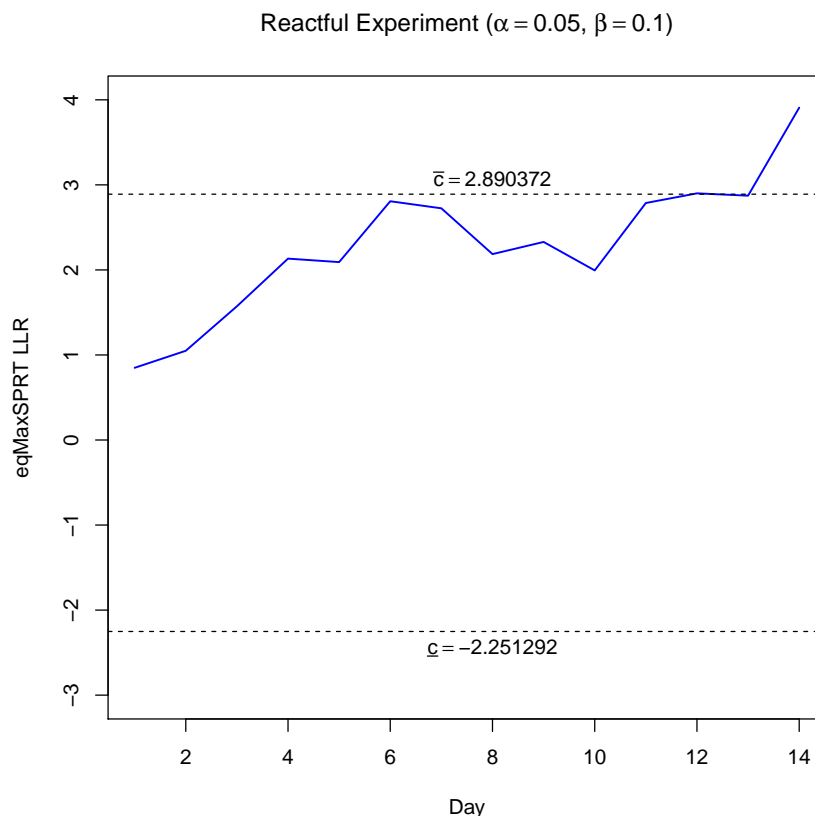


Figure 4.3: eqMaxSPRT Log likelihood-ratio for the Reactful experiment

4.7 Conclusion

The standard technique of testing for equality of conversion rates in A/B tests can be inefficient when not matched with the goal of the test and when not exploiting the sequential nature of arriving data. In this chapter we have shown two approaches that when combined can lead to a substantial decrease in expected sample sizes of online experiments.

The first and simplest approach is to match the statistical test with the goal of the experiment. By realizing that many experiments are aimed at selection, marketers can frequently design powerful experiments which end later than previously hypothesized. The second approach uses sequential analysis techniques to stop the experiment early when the results show enough evidence for the efficacy or lack of efficacy of the treatment.

The field of sequential analysis is wide and contains many varieties of tests and designs that can be used for a diverse set of online scenarios. Applying these techniques, however, requires developing software to make the techniques accessible to researchers and practitioners. Combining sequential techniques with Bayesian methods is a natural avenue for further exploration of the topic. It should be noted, however, that there is a clear advantage of using the frequentist approach when applying these dynamic techniques as they make interpretation and application easy.

Future directions for research include the dynamic allocation of treatments to consumers based on the historical result of the experiment to date. These adaptive methods balance exploration and exploitation of the treatment arms to maximize the value generated by the experiment, and are related to the classic multi-armed bandit problem. Performing these experiments while combining them with an ongoing sequential procedure will prove invaluable to the current development of online experimental techniques.

Bibliography

- Eric T. Anderson and Duncan Simester. A step-by-step guide to smart business experiments. *Harvard Business Review*, 89(3):98, 2011.
- S. Athey and D. Nekipelov. A Structural Model of Sponsored Search Advertising Auctions. *working paper*, 2010.
- Susan Athey and Glenn Ellison. Position auctions with consumer search. *The Quarterly Journal of Economics*, page forthcoming, 2012.
- Jay Bartroff, Tze Leung Lai, and Mei-Chiung Shih. *Sequential Experimentation in Clinical Trials: Design and Analysis*, volume 298. Springer, 2012.
- Michael R. Baye and Heidrun C. Hoppe. The strategic equivalence of rent-seeking, innovation, and patent-race games. *Games and Economic Behavior*, 44(2):217–226, 2003.
- Robert E Bechhofer. An optimal sequential procedure for selecting the best bernoulli process—a review. *Naval research logistics quarterly*, 32(4):665–674, 1985.
- Ron Berman. An application of the shapley value for online advertising campaigns. *Work in Progress*, 2013.
- Thomas Blake, Chris Nosko, and Steven Tadelis. Consumer heterogeneity and paid search effectiveness: A large scale field experiment. *NBER Working Paper*, pages 1–26, 2013.
- Steve Blank. Why the lean start-up changes everything. *Harvard Business Review*, 91(5):63–72, 2013.
- Y. Chen and C. He. Paid placement: Advertising and search on the Internet. *The Economic Journal*, 121(556):309–328, 2011.
- Andrea J Cook, Ram C Tiwari, Robert D Wellman, Susan R Heckbert, Lingling Li, Patrick Heagerty, Tracey Marsh, and Jennifer C Nelson. Statistical approaches to group sequential monitoring of postmarket safety surveillance data: current state of the art for use in the mini-sentinel pilot. *Pharmacoepidemiology and drug safety*, 21(S1):72–81, 2012.
- Morris H DeGroot. Optimal statistical decisions. 1970.

- Xavier Dreze and François-Xavier Husherr. Internet advertising: Is anybody watching? *Journal of interactive marketing*, 17(4):8–23, 2003.
- Alex Gershkov, Jianpei Li, and Paul Schweinzer. Efficient tournaments within teams. *The RAND Journal of Economics*, 40(1):103–119, 2009.
- Anindya Ghose and Sha Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.
- Bhaskar Kumar Ghosh and Pranab Kumar Sen. *Handbook of sequential analysis*. CRC Press, 1991.
- DG Hoel, GH Weiss, and R Simon. Sequential tests for composite hypotheses with two binomial populations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 302–308, 1976.
- Bengt Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340, 1982.
- Michael D Huffman. An efficient approximate solution to the kiefer-weiss problem. *The Annals of Statistics*, pages 306–316, 1983.
- Christopher Jennison and Bruce W Turnbull. *Group sequential methods with applications to clinical trials*. CRC Press, 1999.
- Przemyslaw Jeziorski and Ilya Segal. What Makes Them Click: Empirical Analysis of Consumer Demand for Search Advertising. *SSRN eLibrary*, 2009.
- Zsolt Katona and Miklos Sarvary. The race for sponsored links: Bidding patterns for search advertising. *Marketing Science*, 29(2):199–215, 2010.
- Jack Kiefer, Lionel Weiss, et al. Some properties of generalized sequential probability ratio tests. *The Annals of Mathematical Statistics*, 28(1):57–74, 1957.
- Pavel Kireyev, Koen Pauwels, and Sunil Gupta. Do display ads influence search? attribution and dynamics in online advertising. *Working Paper*, 2013.
- René Kirkegaard. Favoritism in asymmetric contests: Head starts and handicaps. *Games and Economic Behavior*, page forthcoming, 2012.
- Kai A Konrad. Strategy in contests: An introduction. Technical report, Discussion papers//WZB, Wissenschaftszentrum Berlin für Sozialforschung, Schwerpunkt Märkte und Politik, Abteilung Marktprozesse und Steuerung, 2007.
- Vijay John Morgan Krishna. The winner-take-all principle in small tournaments. *Advances in Applied Microeconomics*, 28:849–862, 2007.

- Martin Kulldorff, Robert L Davis, Margarette Koleczak, Edwin Lewis, Tracy Lieu, and Richard Platt. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis*, 30(1):58–78, 2011.
- Anja Lambrecht and Catherine Tucker. When does retargeting work? timing information specificity. *Timing Information Specificity (Dec 02, 2011)*, 2011.
- Randall Lewis, Justin M Rao, and David H Reiley. Measuring the effects of advertising: The digital frontier. Technical report, National Bureau of Economic Research, 2013.
- Randall A Lewis and Justin M Rao. On the near impossibility of measuring advertising effectiveness. Technical report, Working paper, 2012a.
- Randall A Lewis and Justin M Rao. On the near impossibility of measuring advertising effectiveness. Technical report, Working paper, 2012b.
- Alice Li and P.K. Kannan. Modeling the conversion path of online customers. *Working Paper*, 2013.
- Puneet Manchanda, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, pages 98–108, 2006.
- R. Preston McAfee and John McMillan. Optimal contracts for teams. *International Economic Review*, pages 561–577, 1991.
- Oliver J. Rutz and Randolph E Bucklin. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research*, 48(1):87–102, 2011.
- Ravi Sen. Optimal search engine marketing strategy. *Int. J. Electron. Commerce*, 10(1):9–25, 2005.
- Lloyd S Shapley. A value for n-person games. 1952.
- Lee Sherman and John Deighton. Banner advertising: Measuring effectiveness and optimizing placement. *Journal of Interactive Marketing*, 15(2):60–64, 2001.
- Ron Siegel. All-pay contests. *Econometrica*, 77(1):71–92, 2009.
- Dana Sisak. Multiple-prize contests—the optimal allocation of prizes. *Journal of Economic Surveys*, 23(1):82–114, 2009.
- Ajit C Tamhane. Some sequential procedures for selecting the better bernoulli treatment by using a matched samples design. *Journal of the American Statistical Association*, 80(390):455–460, 1985.

- Greg Taylor. Search quality and revenue cannibalisation by competing search engines. *Journal of Economics & Management Strategy*, page forthcoming, 2012.
- Catherine Tucker. The implications of improved attribution and measurability for online advertising markets. 2012.
- Abraham Wald et al. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16(2):117–186, 1945.
- Alexander White. Search engines: Left side quality versus right side profits. *working paper, Toulouse School of Economics*, 2009.
- Bo Xing and Zhangxi Lin. The impact of search engine optimization on online advertising market. In *ICEC '06: Proceedings of the 8th international conference on Electronic commerce*, pages 519–529, New York, NY, USA, 2006. ACM. ISBN 1-59593-392-1. doi: <http://doi.acm.org/10.1145/1151454.1151531>.
- Lizhen Xu, Jianqing Chen, and Andrew B. Whinston. Price competition and endogenous valuation in search advertising. *Journal of Marketing Research*, 48(3):566–586, 2011.
- Lizhen Xu, Jianqing Chen, and Andrew B. Whinston. Effects of the presence of organic listing in search advertising. *Information Systems Research*, page forthcoming, 2012.
- S. Yang and A. Ghose. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, 29:602–623, 2010.
- Song Yao and Carl F. Mela. A dynamic model of sponsored search advertising. *Marketing Science*, 30(3):447–468, 2011.
- Yi Zhu and Kenneth C. Wilbur. Hybrid advertising auctions. *Marketing Science*, 30(2): 249–273, 2011.

Appendix A

Appendix for Chapter 2

A.1 Proofs

Proof of Proposition 1:

Let $F_{\varepsilon_i - \varepsilon_j}$ be the c.d.f. of a triangle distribution $\varepsilon_i - \varepsilon_j \sim T[-\sigma, \sigma]$ with mean zero and $f_{\varepsilon_i - \varepsilon_j}$ be its p.d.f. Each website faces the following first order condition w.r.t. to their scores resulting from the profit function:

$$v_i \cdot f_{\varepsilon_i - \varepsilon_j}(\bar{s}_i - \bar{s}_j) = \frac{\bar{s}_i - q_i}{\alpha^2}, \quad (\text{A.1})$$

where $\bar{s}_i = \mathbb{E}_{\varepsilon_i}[s_i]$. Let $x = \bar{s}_i - \bar{s}_j$ and $\mu = q_i - q_j$. By subtracting both F.O.Cs and using the fact that $f_{\varepsilon_i - \varepsilon_j}$ is symmetric around zero we can rewrite the condition as:

$$f_{\varepsilon_i - \varepsilon_j}(x) = \frac{x - \mu}{\alpha^2(v_i - v_j)} \quad (\text{A.2})$$

An interior solution x^* would require both F.O.Cs and the S.O.Cs to hold as well as $-\sigma \leq x^* \leq \sigma$. When $v_i > v_j$ and $\alpha^2 \geq -\sigma \frac{\mu}{v_i - v_j}$, or when $v_i < v_j$ and $\alpha^2 < \sigma \frac{\mu}{v_j - v_i}$, the equilibrium solution is $s_i^* - s_j^* = x_R^* = \frac{\sigma^2 \mu + \sigma \alpha^2 (v_i - v_j)}{\sigma^2 + \alpha^2 (v_i - v_j)}$. When $v_i < v_j$ and $\alpha^2 \geq \sigma \frac{\mu}{v_j - v_i}$, or when $v_i > v_j$ and $\alpha^2 < -\sigma \frac{\mu}{v_i - v_j}$, the equilibrium solution is $s_i^* - s_j^* = x_L^* = \frac{\sigma^2 \mu + \sigma \alpha^2 (v_i - v_j)}{\sigma^2 - \alpha^2 (v_i - v_j)}$.

We can immediately verify that the condition $\sigma > \mu$ ensures that $-\sigma \leq x \leq \sigma$, while $\alpha^2 < \frac{\sigma^2}{v_H}$ ensures that both the F.O.Cs and the S.O.Cs hold. Under the condition on α , the equilibrium point is a unique extremum, and thus a global maximum.

To examine the effects of the equilibrium SEO investment on the ranking efficiency and consumer satisfaction, we let $P(\alpha)$ denote the probability that the player with the highest quality wins the organic link. Assume $q_H = q_1 > q_2 = q_L$. In the perfectly correlated case $x^* = x_R^*$. We then have $P(\alpha) = F_{\varepsilon_1 - \varepsilon_2}(x_R^*)$ and $P'(\alpha) = f_{\varepsilon_1 - \varepsilon_2}(x_R^*) \frac{\partial x}{\partial \alpha} \Big|_{x=x_R^*} > 0$. In the perfectly negatively correlated case, when $\rho = -1$, we have $x^* = x_L^*$, thus $P'(\alpha) =$

$f_{\varepsilon_1-\varepsilon_2}(x_L^*) \frac{\partial x}{\partial \alpha} \Big|_{x=x_L^*} < 0$. Building on the two extreme cases, one can show for intermediate correlation values that $P(\alpha) > P(0)$ for certain $\alpha > 0$ and $0 < \rho < 1$.

To prove the second part, when $\rho = -1$ taking the derivative with respect to α of the profit functions of both players shows that at the limit of $\alpha \rightarrow 0$, the profit never increases for any of the equilibrium conditions. It should be noted that at some conditions the profit might increase for higher values of α . When $\rho = 1$ solving directly for player 1:

$$\pi_i(\alpha) - \pi_i(\alpha = 0) = \pi_i(\alpha)|_{x^*} - v_1 F_{\varepsilon_1-\varepsilon_2}(q_H - q_L) > 0 \quad (\text{A.3})$$

yields the conditions $v_H > 2v_L$ or $\alpha^2 > \frac{\sigma^2(2v_L-v_H)}{(v_H-v_L)^2}$, where the latter condition is ruled out if α is small enough. For player 2, the same exercise shows there is no solution for $\alpha > 0$ that increases player 2's profit. \square

Proof of Proposition 2: We use backward induction and first determine the sponsored bids given the allocation of the organic link, then the SEO investments in three different cases with respect to the site qualities. Initially, we assume that consumers start with the organic link. Later we will show that this is an equilibrium strategy and that starting with the sponsored link cannot be (*Part 1* of the proposition). We will also determine the threshold \bar{c} . We start with the $r < v_L$ case and then show how the analysis changes for $v_L \leq r < v_H$. Let w_O denote the organic and w_S the sponsored winner. The main technique we use is to compare the profits in equilibrium when the player occupies and does not occupy the organic link. The difference between these profits is the value of the organic link for that player.

Case I: When $q_i = q_j = q_H$, consumers stop searching at the organic link and do not search further. This renders the sponsored link useless for both players leading to no valid bids above the reserve price, r . The SEO game is therefore equivalent to the case with no sponsored links.

Case II: When $q_i = q_j = q_L$, consumers will not be satisfied with the organic link and continue to the sponsored link as long as it does not lead to the same site. If site i is the organic winner, then $ctr_i = 0$ for the sponsored link, leaving the sponsored link for site $j \neq i$ to win at a price per click equal to r . Since $q_i = q_j$, consumers do not go back to the organic link, leaving 0 profits for site i . The organic link is worthless, therefore no site will invest in SEO.

Case III: When $q_i = q_H$ and $q_j = q_L$, consumers will stop at the organic link if $w_O = i$. Just as in Case I, no site will submit a valid bid higher than r . If $w_O = j$, consumers will not be satisfied with a low quality organic link and will continue searching, as long as the sponsored link is different from the organic. As in Case II, $ctr_i = 1$ and $ctr_j = 0$, leading to $w_S = i$ at a price per click of r . Hence site i , with the high quality, will capture all the demand regardless of which position it is in. When $w_O = i$, this will lead to $\pi_i^O = v_i$, but when $w_O = j$ and $w_S = i$, site i has to pay for the sponsored link and $\pi_i^S = v_i - r$. The value of winning the organic link will therefore be $\pi_i^O - \pi_i^S = r$ for site i and $\pi_j^O - \pi_j^S = 0$ for site j . Applying the results of Proposition 1 with $v'_i = r, v'_j = 0, q'_i = q_H, q'_j = q_L$, we get

the optimal SEO efforts and the probability of a high quality organic link:

$$e_i^* = \frac{\alpha r(\sigma - q_H + q_L)}{\alpha^2 r + \sigma^2}, \quad e_j^* = 0, \quad P^* = P(\alpha | q_i = q_H, q_j = q_L) = 1 - \frac{1}{2} \left(\frac{\sigma(\sigma - q_H + q_L)}{\alpha^2 r + \sigma^2} \right)^2. \quad (\text{A.4})$$

P^* is increasing in α , that is, $w_O = i$ becomes more likely as α increases regardless of ρ , proving *Part 2* of the Proposition.

In *Part 3*, when $v_L \leq r < v_H$ the analysis is identical to the above except in *Case III*, when $w_O = j$ and $v_i = v_L < r$. In this case site i with $q_i = q_H$ cannot afford the sponsored link and will profit $\pi_i^O - \pi_i^S = v_L - 0 = v_L$ from getting the organic link, whereas site j will profit $\pi_j^O - \pi_j^S = v_H - 0 = v_H$. According to Proposition 1 a higher α decreases $\Pr(w_O = i)$, but the probability of this case is $\Pr(q_i = q_H, q_j = q_L, v_i = v_L, v = v_H) = \left(\frac{1-\rho}{4}\right)^2$, which decreases with ρ and reaches 0 when $\rho = 1$. Thus, SEO will only increase the probability of the high quality site acquiring the organic link if ρ is high enough, proving *Part 3*.

Returning to *Part 1*, combining the three cases, it is clear that the organic link is more likely to be of high quality than the sponsored link. It is therefore rational for consumers to start their search with the organic link. On the other hand, assuming that consumers start with the sponsored link, redoing the same analysis shows that even then the organic link is more likely to be high quality. Starting with the sponsored link is therefore never an equilibrium strategy. Furthermore, in order to determine \bar{c} , we need to calculate the expected benefit of continuing the search when finding q_L . This is simply $(q_H - q_L) \Pr(q_{w_S} = q_H | q_{w_O} = q_L) = (q_H - q_L) \frac{(1/2)(1-P^*)}{(1/4)+(1/2)(1-P^*)}$, where P^* is defined in (A.4). For a consumer to even start searching it is sufficient to assume $c < q_L$. Therefore,

$$\bar{c} = \min \left(q_L, (q_H - q_L) \frac{1 - P^*}{3/2 - P^*} \right) \quad (\text{A.5})$$

To prove *Part 4*, we only need to examine *Case III*, since neither consumer welfare nor search engine revenue is affected by SEO in *Case I* and *Case II*. In *Case III*, consumers always find q_H eventually, but they are better off finding it right away, when $w_O = i$. Therefore, consumer welfare increases iff $P(\alpha)$ increases. On the other hand, search engine revenues are higher when the low quality site acquires the organic link, that is, the revenue increases iff $P(\alpha)$ decreases, proving *Part 4*. \square

Proof of Corollary 1: Consumers only click the sponsored link if the organic link is of low quality. Thus, the search engine's revenue is $R^{SE} = (1 - P(\alpha)) \cdot r$, since the search engine makes exactly r when the low quality site gets the organic link. From the proof of Proposition 1, we can derive $P(\alpha) = P(\alpha, r) = \frac{3}{4} - \frac{\sigma^2(\sigma - q_H + q_L)^2}{4(\sigma^2 + r\alpha^2)^2}$ which is clearly increasing in r . Differentiating the revenue with respect to r yields $\frac{\partial R^{SE}}{\partial r} = 1 - P(\alpha, r) - r \cdot \frac{\partial P(\alpha, r)}{\partial r} = \frac{1}{4} + \frac{\sigma^2(\sigma - q_H + q_L)^2(\sigma^2 - r\alpha^2)}{4(\sigma^2 + r\alpha^2)^3}$. The above derivative is positive if r is below a suitable $\hat{r}(\alpha)$, leading to an inverse U-shaped revenue function below v_L . The implicit function theorem yields that $\hat{r}(\alpha)$ is decreasing. \square

Proof of Corollary 2: When $r < v_L$ the higher quality site has an effective valuation of r for the organic link, whereas the low quality site has an effective valuation of 0. From the proof of Proposition 1, it is clear that the high quality site has an increasing chance of acquiring the organic link and its profit increases as α increases. \square

A.2 SEO Contest with a General Error Distribution

In this section we check whether our main result in Proposition 1 are robust to different F_ε error distributions when σ is high enough. Let us assume that $F_\varepsilon(x) = F\left(\frac{x}{\sigma}\right)$, where $f() = F'()$ and $f_\varepsilon(x) = \frac{f\left(\frac{x}{\sigma}\right)}{\sigma}$ is the p.d.f of the distribution. Let us assume that the distribution f has an infinite support. The p.d.f. of the difference of two independent errors will be $f_{\varepsilon_i - \varepsilon_j}(x) = \frac{f_\Delta\left(\frac{x}{\sigma}\right)}{\sigma}$ where $f_\Delta = f * -f$ is the convolution of $f()$ and $-f()$. Therefore f_Δ is symmetric around 0 and also has an infinite support.

Recall that in the proof of Proposition 1, we derive our results in general until equation (A.1). Plugging in $f_{\varepsilon_i - \varepsilon_j}(x) = \frac{f_\Delta\left(\frac{x}{\sigma}\right)}{\sigma}$, we get

$$f_\Delta\left(\frac{s_i - s_j}{\sigma}\right) = \frac{\sigma(s_i - q_i)}{v_i \alpha^2} = \frac{\sigma e_i}{v_i \alpha} \quad (\text{A.6})$$

Since f_Δ is symmetric around 0, the left hand side is the same for both players, revealing that players will exert efforts proportional to their valuations. A solution to the first order condition always exists and it will correspond to a unique maximum as long as σ is high enough. The above equation immediately yields the solution for $v_i = v_j$. When $v_i > v_j$, we plug in $s_j = \frac{v_j}{v_i} s_i$ to obtain

$$f_\Delta\left(\frac{s_i}{\sigma} \left(1 - \frac{v_j}{v_i}\right)\right) = \frac{\sigma(s_i - q_i)}{v_i \alpha^2}. \quad (\text{A.7})$$

Again, if σ is high enough this yields a unique s_i^* solution providing $s_j^* = \frac{v_j}{v_i} s_i^*$ and the $e_i^* = \frac{s_i^* - q_i}{\alpha}$, $e_j^* = \frac{s_j^* - q_j}{\alpha}$ equilibrium efforts. We can then show that $P(\alpha)$ is increasing (decreasing) depending on the relationship between (v_i, v_j) and (q_i, q_j) in the exact same fashion as in the proof of Proposition 1.

A.3 SEO with Errors Observed by Players - Relation to All-Pay Auctions

Our model of SEO on the organic side is closely related to research on all-pay auctions and contests. Many applications exists from innovation to patent-race games that are strategically equivalent (Baye and Hoppe, 2003). Our analysis specifically takes into account asymmetries among websites as well as ranking error of the search engine. Kirkegaard (2012)

describes the equilibria in contests with asymmetric players, while Siegel (2009) analyzed such games under more general conditions. Our application is unique in that it considers the cases where the initial asymmetry is biased by noise inherent in the quality measurement process. Krishna (2007) and Athey and Nekipelov (2010) are two of the few examples taking noise into consideration in an auction setting.

In this section we assume that two Web sites compete for a single organic link, but unlike in our main model, we assume that sites observe the error made by the search engine in assigning scores to them. Before deciding on their SEO investments, sites therefore observe $s_i^S = q_i + \epsilon_i$. Let the distribution of the error be simple: it takes the values of σ or $-\sigma$ with equal probabilities. We assume $\sigma > |q_1 - q_2|/2$ to ensure that the error can affect the ordering of sites, otherwise the error never changes the order of results and the setup is equivalent to one with no error. We assume that valuations are exogenously given v_1, v_2 and that qualities are $q_1 > q_2$ since in the case of equal qualities, SEO does not matter.

When search engine optimization is not possible, i.e., when $\alpha = 0$, sites cannot influence their position among the search results. Since $q_1 \geq q_2$, the probability that the higher quality site gets the organic link is $P(0) = \frac{3}{4}$. When search engine optimization becomes effective, i.e., when $\alpha > 0$, websites have a tool to influence the order of results knowing the score that has been assigned to them s_i^S , which includes the error. The game thus becomes an all-pay auction with headstarts.

Lemma 3. *The game that sites play after observing their starting scores is equivalent to an all-pay auction with headstarts.*

Proof. All pay-auctions with headstarts are generalizations of basic all-pay auctions. In traditional all-pay auctions players submit bids for an object that they have different valuations for. The player with the highest bid wins the object, but all players have to pay their bid to the auctioneer (hence the term “all-pay auction”). When the auctioneer does not collect the revenues from the bids which are sunk, the game is called a contest. If players have headstarts then the winner is the player with the highest score - the sum of bid and headstart.

The level of headstart in our model depends on the starting scores and hence on the error. For example, if $q_1 > q_2$ and $\varepsilon_1 = \varepsilon_2 = 1$, the error does not affect the order (which is $q_1 \geq q_2$) nor the difference between the starting scores ($q_1 - q_2$). Since SEO effectiveness is α , an investment of b only changes the scores by αb , therefore the headstart of site 1 is $\frac{q_1 - q_2}{\alpha}$. As the size of the headstart decreases with α , the more effective SEO is, the less the initial difference in scores matters. Even if site 1 is more relevant than site 2, it is not always the case that it has a headstart. If $\varepsilon_1 = -1$ and $\varepsilon_2 = 1$ then $s_1^S = q_1 - \sigma < s_2^S = q_2 + \sigma$ given our assumption on the lower bound on σ . Thus, player 2 has a headstart of $\frac{q_2 + 2\sigma - q_1}{\alpha}$. By analyzing the outcome of the all-pay auction given the starting scores, we can determine the expected utility of the SE and the websites.

We decompose the final scores of both sites into a headstart h and a bid as follows: $\tilde{s}_1^F = h + b_1$ and $\tilde{s}_2^F = b_2$ where $h = \frac{s_1^S - s_2^S}{\alpha}$. The decomposed scores have the property that

$\tilde{s}_1^F \geq \tilde{s}_2^F \iff s_1^F \geq s_2^F$ for every b_1, b_2 and thus preserve the outcome of the SEO game. Since the investments are sunk and only the winner receives the benefits (with the exception of a draw) the SEO game is equivalent to an all-pay auction with a headstart of $h = \frac{s_1^S - s_2^S}{\alpha}$. In the following, we present the solution of such a game to facilitate the presentation of the remaining proofs. \square

All-pay auctions with complete information typically do not have pure-strategy Nash-equilibria. In a simple auction with two players with valuations $v_1 > v_2$, both players mix between bidding 0 and v_2 with different distributions. The generic two player all-pay auction with headstarts has a unique mixed strategy equilibrium. When players valuations are $v_1 \geq v_2$ and player 1 has a headstart of h then s/he wins the auction with the following probabilities:

$$W_1(h) = Pr(1 \text{ wins} | h \geq 0) = \begin{cases} 1 & h > v_2 \\ 1 - \frac{v_2}{2v_1} + \frac{h^2}{2v_1v_2} & h \leq v_2 \end{cases}$$

$$W_1(h) = Pr(1 \text{ wins} | h < 0) = \begin{cases} 1 - \frac{v_2}{2v_1} & h \geq v_2 - v_1 \\ \frac{v_1^2 - h^2}{2v_1v_2} & -v_1 \leq h < v_2 - v_1 \\ 0 & \text{otherwise} \end{cases}$$

For completeness, we specify the players' cumulative bidding distributions. When h is positive,

$$F_1(b) = \begin{cases} 0 & b \leq 0 \\ \frac{h+b}{v_2} & b \in (0, v_2 - h] \\ 1 & b > v_2 - h \end{cases} \quad F_2(b) = \begin{cases} 0 & b \leq 0 \\ 1 - \frac{v_2 - h}{v_1} & b \in (0, h] \\ 1 - \frac{v_2 - b}{v_1} & b \in (h, v_2] \\ 1 & b > v_2 \end{cases} \quad (\text{A.8})$$

When h is negative,

$$F_1(b) = \begin{cases} 0 & b \leq h \\ \frac{b-h}{v_2} & b \in (h, v_2 + h] \\ 1 & b > v_2 + h \end{cases} \quad F_2(b) = \begin{cases} 0 & b \leq 0 \\ 1 - \frac{v_2 - b}{v_1} & b \in (0, v_2] \\ 1 & b > v_2 \end{cases} \quad (\text{A.9})$$

In our model, the value of the headstart is determined by the different realizations of the errors $\varepsilon_1, \varepsilon_2$. There are four possible realizations with equal probability: $h_1 = h_2 = \frac{q_1 - q_2}{\alpha}$, $h_3 = \frac{q_1 - q_2 + 2\sigma}{\alpha}$ and $h_4 = \frac{q_1 - q_2 - 2\sigma}{\alpha}$. Player 1, having the higher valuation, wins with the higher probability of $v_1/2v_2$ and player 2's surplus is 0. Thus, only the player with the highest valuation makes a positive profit in expectation, but the chance of winning gives an incentive to the other player to submit positive bids. In the case of an all-pay auction with headstarts the equilibrium is very similar and the player with the highest potential score (valuation plus headstart) wins with higher probability and the other player's expected surplus is 0. The winner's expected surplus is equal to the sum of differences in valuations and headstarts. Figure A.1 illustrates the probabilities that the two sites win and their payoffs as a function of the headstart.

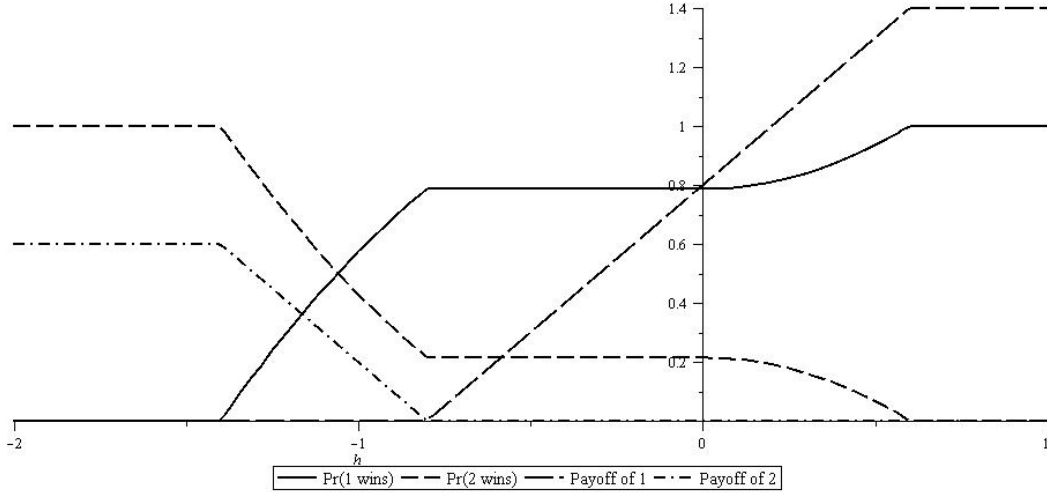


Figure A.1: Mixed strategy equilibrium of an all-pay auction as a function of the headstart of player 1.

Sites' valuations are $v_1 = 1.4$ and $v_2 = 0.6$. The probability that player 1 (player 2) wins is weakly increasing(decreasing) in the headstart, similarly to the payoffs.

As we have seen when SEO is not possible and $\alpha = 0$, we have $P(0) = 3/4$. Our goal is therefore to determine whether the probability exceeds this value for any positive α SEO effectiveness levels. It is useful, however, to begin with analyzing how the probability depends on valuations and qualities for given α and σ values. The following Lemma summarizes our initial results.

Lemma 4. *For any fixed α and σ , $P(\alpha; \sigma, v_1, v_2, q_1, q_2)$ is increasing in v_1 and q_1 and is decreasing in v_2 and q_2 .*

Proof. Since $P(\alpha) = \frac{1}{2}W_1(h_1) + \frac{1}{4}W_1(h_3) + \frac{1}{4}W_1(h_4)$ and the headstart does not depend on v_1 and v_2 , it is enough to show that $W_1(\cdot)$ is increasing in v_1 and decreasing in v_2 . These easily follow from the definition of $W_1(\cdot)$. The results on q_1 and q_2 follow from the fact that h_1, h_3, h_4 are all increasing in q_1 and decreasing in q_2 , and $W_1(\cdot)$ depend on them only through h in which it is increasing. \square

To show that our main results hold in this case, we derive the following.

Proposition 8.

1. *For any $\sigma > |q_1 - q_2|/2$, there exists a positive $\hat{\alpha} = \hat{\alpha}(\sigma, v_1, v_2, q_1, q_2)$ SEO effectiveness level such that $P(\hat{\alpha}) \geq P(0)$.*
2. *If $v_1/v_2 > 3/2$ then for any $\sigma > |q_1 - q_2|/2$, there exists a positive $\hat{\alpha} = \hat{\alpha}(\sigma, v_1, v_2, q_1, q_2)$ such that $P(\hat{\alpha}) > P(0)$.*

3. If $v_1 < v_2$ and $\sigma \geq \frac{v_2+v_1}{v_2-v_1} \frac{q_1-q_2}{2}$ then for any $\alpha > 0$ we have $P(\alpha) \leq P(0)$.

Proof. We use the notation $P_i = Pr(1 \text{ wins} | h_i)$. Given the above described equilibrium of the two-player all-pay auctions we have $P_i = W_1(h_i)$. We further define $\alpha_1 = \frac{q_1-q_2}{v_2}$, $\alpha_3 = \frac{q_1-q_2+2\sigma}{v_2}$, $\alpha_4 = \frac{q_2-q_1+2\sigma}{v_1}$, $\alpha'_4 = \frac{q_2-q_1+2\sigma}{v_1-v_2}$. Note that $P_1 = P_2$, since the headstarts in the first two case are equal. Thus $P(\alpha) = \frac{1}{2}P_1 + \frac{1}{4}P_3 + \frac{1}{4}P_4$, and $P_1 = 1$ iff $\alpha \leq \alpha_1$, $P_3 = 1$ iff $\alpha \leq \alpha_3$, $P_4 = 1 - \frac{v_2}{2v_1}$ iff $\alpha \geq \alpha'_4$. Furthermore, it is easy to check that $\alpha_1 \leq \alpha_3$, $\alpha_4 \leq \alpha_3$, and $\alpha_4 \leq \alpha'_4$.

We proceed by separating the three parts of the proposition:

- Part 1: By setting $\alpha = \alpha_1$, we have $P_1 = P_3 = 1$, and thus $P(\alpha) \geq 3/4$ for any σ .
- Part 2: In order to prove this part, we determine the α value that yields the highest efficiency level for a given σ if $v_1/v_2 > 3/2$. As noted above, $P(\alpha)$ is a linear combination of $W_1(h_1)$, $W_1(h_3)$, $W_1(h_4)$. Since $W_1(\cdot)$ is continuous and h_1, h_3, h_4 are all continuous in α , it follows that $P(\alpha)$ is continuous in α . However, $P(\alpha)$ is not differentiable everywhere, but there are only a finite number of points where it is not. Therefore it suffices to examine the sign of $P'(\alpha)$ to determine whether it is increasing or not. This requires tedious analysis, since depending on the value of σ the formula describing $P(\alpha)$ is different in up to five intervals. We identify five different formulas that $P(\alpha)$ can take in different intervals and take their derivatives:

$$\begin{aligned}
 P'(\alpha) &= P'_I(\alpha) = \frac{(q_1 - q_2 - 2\sigma)^2}{4\alpha^3 v_1 v_2} \text{ if } \alpha_4 \leq \alpha \leq \alpha_1 \& \alpha'_4, \\
 P'(\alpha) &= P'_{II}(\alpha) = -\frac{(q_1 - q_2)^2}{2\alpha^3 v_1 v_2} \text{ if } \alpha_1 \leq \alpha \leq \alpha_4, \\
 P'(\alpha) &= P'_{III}(\alpha) = -\frac{2(q_1 - q_2)^2 + (q_1 - q_2 + 2\sigma)^2}{4\alpha^3 v_1 v_2} \text{ if } \alpha_3 \& \alpha'_4 \leq \alpha, \\
 P'(\alpha) &= P'_{IV}(\alpha) = \frac{4\sigma^2 - (q_1 - q_2)(4\sigma + q_1 - q_2)}{4\alpha^3 v_1 v_2} \text{ if } \alpha_1 \& \alpha_4 \leq \alpha \leq \alpha'_4, \\
 P'(\alpha) &= P'_V(\alpha) = -\frac{(q_1 - q_2)(4\sigma + q_1 - q_2)}{2\alpha^3 v_1 v_2} \text{ if } \alpha_3 \& \alpha'_4 \leq \alpha.
 \end{aligned}$$

In any other range the derivative of $P(\alpha)$ is 0. It is clear from the above formulas that $P'_I(\alpha)$ is always positive and that $P'_{II}(\alpha)$, $P'_{III}(\alpha)$, and $P'_V(\alpha)$ are always negative. Furthermore, one can show that

$$P'_{IV}(\alpha) > 0 \text{ iff } \sigma > \frac{1 + \sqrt{2}}{2}(q_1 - q_2).$$

This allows us to determine the maximal $P(\alpha)$ for different values of σ in four different cases.

1. If $\frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1}{v_2} \frac{q_1 - q_2}{2}$ then $\alpha_4 \leq \alpha'_4 \leq \alpha_1 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: $0, P'_I(\alpha), 0, P'_{II}(\alpha), P'_{III}(\alpha)$. Therefore $P(\alpha)$ is first constant, then increasing, then constant again and then strictly decreasing. Thus, any value between α'_4 and α_1 maximizes $P(\alpha)$. Using the notation of Corollary 5, $\hat{A}(\sigma) = [\alpha'_4, \alpha_1]$.
2. If $\frac{v_1}{v_2} \frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1 + v_2}{v_2} \frac{q_1 - q_2}{2}$ then $\alpha_4 \leq \alpha_1 \leq \alpha'_4 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: $0, P'_I(\alpha), P'_{IV}(\alpha), P'_{II}(\alpha), P'_{III}(\alpha)$. Therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing, then depending on the sign of $P'_{IV}(\alpha)$ increasing or decreasing, and finally strictly decreasing. Therefore if $\sigma < \frac{1 + \sqrt{2}}{2}(q_1 - q_2)$ then α_1 maximizes $P(\alpha)$, that is $\hat{A}(\sigma) = \{\alpha_1\}$. If $\sigma = \frac{1 + \sqrt{2}}{2}(q_1 - q_2)$ then $P(\alpha)$ is constant between α_1 and α'_4 , that is $\hat{A}(\sigma) = [\alpha_1, \alpha'_4]$. Finally, if $\sigma = \frac{1 + \sqrt{2}}{2}(q_1 - q_2)$ then $\hat{A}(\sigma) = \{\alpha'_4\}$.
3. If $\frac{v_1 + v_2}{v_2} \frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1}{2v_2 - v_1} \frac{q_1 - q_2}{2}$ then $\alpha_1 \leq \alpha_4 \leq \alpha'_4 \leq \alpha_3$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: $0, P'_{II}(\alpha), P'_{IV}(\alpha), P'_{II}(\alpha), P'_{III}(\alpha)$. In this case $P'_{IV}(\alpha) > 0$ since $\sigma \geq \frac{v_1 + v_2}{v_2} \frac{q_1 - q_2}{2} \geq (1 + \frac{3}{2}) \frac{q_1 - q_2}{2} > (1 + \sqrt{2}) \frac{q_1 - q_2}{2}$. Therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing again and finally strictly decreasing. Thus, there are two candidates for the argmax: α_1 and α'_4 . One can show that $P_{IV}(\alpha'_4) > P_{II}(\alpha_1)$ iff $v_1 > \sqrt{2}v_2$, therefore α'_4 maximizes $P(\alpha)$ in this case.
4. If $\frac{v_1}{2v_2 - v_1} \frac{q_1 - q_2}{2} \leq \sigma$ then $\alpha_1 \leq \alpha_4 \leq \alpha_3 \leq \alpha'_4$ and the derivative of $P(\alpha)$ takes the following values in the five intervals respectively: $0, P'_{II}(\alpha), P'_{IV}(\alpha), P'_V(\alpha), P'_{III}(\alpha)$. Similarly to the previous case $P'_{IV}(\alpha) > 0$, therefore $P(\alpha)$ is first constant, then decreasing, then strictly increasing again and finally strictly decreasing. Comparing the two candidates for the argmax yields that $P_{IV}(\alpha_3) > P_{II}(\alpha_1)$ iff $v_1 > (3/2)v_2$, that is α_3 maximizes $P(\alpha)$ in this case.

In each of the cases above, it is clear that the maximum is higher than $P(0) = 3/4$. In cases 1 and 2, $P(\alpha)$ is strictly increasing after a constant value of $3/4$ and in cases 3 and 4 we directly compared to $P_{II}(\alpha_1) = 3/4$. This completes the proof of Part 2.

- Part 3: One can derive the efficiency function for different cases as in Part 2. It follows that if $\sigma \geq \frac{v_2 + v_1}{v_2 - v_1} \frac{q_1 - q_2}{2}$ then $P'(\alpha)$ is first 0 then negative and finally positive. Therefore $P(\alpha)$ either has a maximum in $\alpha = 0$ or as it approaches infinity. However,

$$P(\alpha) \xrightarrow{\alpha \rightarrow \infty} \frac{v_1}{2v_2} \leq \frac{1}{2} < \frac{3}{4} = P(0).$$

□

The results are consistent with our main model, where the errors are not observed by the firms prior engaging in SEO. We also examine how the benefits of SEO change with the magnitude of the error. Let $\hat{A}(\sigma)$ denote the set of α SEO effectiveness levels that maximize

the search engine's traffic. For two sets $A_1 \subseteq \mathbb{R}$ and $A_2 \subseteq \mathbb{R}$, we say that $A_1 \succeq A_2$ if and only if for any $\alpha_1 \in A_1$ there is an $\alpha_2 \in A_2$ such that $\alpha_2 \leq \alpha_1$ and for any $\alpha'_2 \in A_2$ there is an $\alpha'_1 \in A_1$ such that $\alpha'_1 \geq \alpha'_2$.

Corollary 5. *If $v_1/v_2 > 3/2$, then the optimal SEO effectiveness is increasing as the variance of the measurement error increases. In particular, for any $\sigma_1 > \sigma_2 > 0$, we have $\hat{A}(\sigma_1) \succeq \hat{A}(\sigma_2)$.*

Proof. In the proof of Proposition 8, we determined the values of α that maximize $P(\alpha)$ for different σ 's. In summary:

$$\hat{A}(\sigma) = \begin{cases} [\alpha'_4, \alpha_1] & \text{if } \frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1}{v_2} \frac{q_1 - q_2}{2} \\ \alpha_1 & \text{if } \frac{v_1}{v_2} \frac{q_1 - q_2}{2} < \sigma \leq (1 + \sqrt{2}) \frac{q_1 - q_2}{2} \\ [\alpha_1, \alpha'_4] & \text{if } \sigma = (1 + \sqrt{2}) \frac{q_1 - q_2}{2} \\ \alpha'_4 & \text{if } (1 + \sqrt{2}) \frac{q_1 - q_2}{2} < \sigma \leq \frac{v_1 + v_2}{v_2} \frac{q_1 - q_2}{2} \\ \alpha'_4 & \text{if } \frac{v_1 + v_2}{v_2} \frac{q_1 - q_2}{2} \leq \sigma \leq \frac{v_1}{2v_2 - v_1} \frac{q_1 - q_2}{2} \\ \alpha_3 & \text{if } \frac{v_1}{2v_2 - v_1} \frac{q_1 - q_2}{2} \leq \sigma \end{cases}$$

It is straightforward to check that all of α_1 , α_3 , and α'_4 are increasing in σ and that the $\hat{A}(\sigma)$ is increasing over the entire range. \square

A.4 SEO with Multiple Organic Links

Here we show that our main results on SEO are robust under more general assumptions. We focus on showing that SEO is beneficial to improving the ranking of organic links, and defer analysis on the impact of sponsored links and search engine profits to future work. First, we extend our model to allow multiple websites to compete for multiple links in one ranked list. Second, we relax the assumption on the distribution of the search engine's measurement error. Finally, we consider the case of an incomplete information structure, where websites do not know the values of the measurement errors induced by the search engine's algorithm, and analyze the resulting Bayesian Nash equilibrium.

The analysis is highly simplified by the use of a multiplicative scoring function instead of an additive one. Thus, the ranking score of site i with quality q_i is $\tilde{s}_i = \tilde{q}_i \cdot \tilde{b}_i^\alpha \cdot \tilde{\varepsilon}_i$. This scoring function is equivalent to taking an exponent of our original additive function and maintains its ordinal properties. Here, we assume that a website's effort of \tilde{b}_i costs \tilde{b}_i , which results in a convex cost function.

The game still consists of n websites that are considered by the search engine for inclusion in the organic list consisting of k links with qualities $\tilde{q}_1 > \tilde{q}_2 > \dots > \tilde{q}_n$. Let $\tilde{q} = (\tilde{q}_1, \dots, \tilde{q}_n)$, and let $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_n)$ be the SEO expenditures of the n sites. Regarding the error $\tilde{\varepsilon}_i$, we allow its distribution to be arbitrary with c.d.f $F_{\tilde{\varepsilon}}$ having finite support, a mean of zero and a finite variance normalized to 1. Let $\underline{\varepsilon}$ and $\bar{\varepsilon}$ be the lower and upper boundaries of the support respectively, and assign $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n)$. Similarly to Section 2.3, we assume that

the error is large enough that it makes a difference, that is, we assume that $\tilde{q}_i \underline{\varepsilon} < \tilde{q}_{i+1} \bar{\varepsilon}$ for each $1 \leq i \leq n$. Furthermore, let Φ_i^j be an indicator for site i appearing in location j among the top k sites.

We treat consumer search as an exogenous process and assume that when site i is displayed in location j of the organic list, it receives β_j clicks from a mass one of consumers. We call this quantity the click-through rate. Given sites' click-through rates, we define t_i as the total amount of visitor traffic a site receives in a list of k sites:

$$t_i(\tilde{b}, \tilde{q}) = \mathbb{E}_{\tilde{\varepsilon}} \left[\sum_{j=1}^k \beta_j \Pr(\Phi_i^j = 1) \right]. \quad (\text{A.10})$$

The profit of site i is thus $\pi_i(\tilde{b}, \tilde{q}) = R_i(t_i(\tilde{b}, \tilde{q})) - \tilde{b}_i$. We let $\pi = (\pi_1, \dots, \pi_n)$. The first order conditions necessary for equilibrium are given by

$$\frac{\partial t_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q})}{\partial \tilde{b}_i} = \frac{1}{r_i(t_i(\tilde{b}, \tilde{q}))}. \quad (\text{A.11})$$

Our construction fulfills the conditions of Theorem 1 in Athey and Nekipelov (2010). To see this, we first prove that a proportional increase in the bids of all other players decreases site i 's profit by \tilde{b}_i , which is a variation on Lemma 1 in Athey and Nekipelov (2010).

Lemma 5. *Assume that $\frac{\partial}{\partial \tilde{b}'} \pi(\tilde{b}, \tilde{q})$ is continuous in \tilde{b} . Suppose that $t_i(\tilde{b}, \tilde{q}) > 0$ for all i . Then \tilde{b} is a vector of equilibrium bids satisfying the first order conditions in (A.11) iff $\frac{d}{d\tau} \pi_i(\tilde{b}, \tau \tilde{b}_{-i}, \tilde{q})|_{\tau=1} = -\tilde{b}_i$ for all $i \leq k$.*

Proof. We denote by $P_i^j(\tilde{b}, \tilde{q})$ the probability that site i appears in location j among the top k sites. This probability equals $P_i^j(\tilde{b}, \tilde{q}) = \int \Phi_i^j(\tilde{b}, \tilde{q}, \tilde{\varepsilon}) dF_{\tilde{\varepsilon}}(\tilde{\varepsilon})$. The total number of clicks site i gets, t_i , is therefore $t_i(\tilde{b}, \tilde{q}) = \sum_{j=1}^k \beta_j P_i^j(\tilde{b}, \tilde{q})$.

A proportional increase of all bids in \tilde{b} does not change the expected rankings of the sites, and keeps the expected number of clicks constant for all sites: $t_i(\tilde{b}, \tilde{q}) = t_i(\eta \tilde{b}, \tilde{q})$ for any $\eta \neq 0$. Since t_i is homogeneous of degree zero, by Euler's homogeneous function theorem, $\sum_{l=1}^k \tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} R_i(t_i(\tilde{b}, \tilde{q})) = r_i \sum_{l=1}^k \tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} t_i(\tilde{b}, \tilde{q}) = 0$.

As a result, the following holds:

$$\sum_{l=1}^k \frac{\partial}{\partial \tilde{b}_l} \pi_i(\tilde{b}, \tilde{q}) \cdot \tilde{b}_l = \sum_{l=1}^k \left(\tilde{b}_l \frac{\partial}{\partial \tilde{b}_l} R_i(t_i(\tilde{b}, \tilde{q})) \right) - \tilde{b}_i = -\tilde{b}_i \quad (\text{A.12})$$

This identity can be rewritten as:

$$\frac{\partial}{\partial \tau} \pi_i(\tilde{b}_i, \tau \tilde{b}_{-i}, \tilde{q})|_{\tau=1} + \tilde{b}_i \frac{\partial}{\partial \tilde{b}_i} \pi_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q}) = -\tilde{b}_i \quad (\text{A.13})$$

Thus, the FOC $\frac{\partial}{\partial \tilde{b}_i} \pi_i(\tilde{b}_i, \tilde{b}_{-i}, \tilde{q}) = 0$ holds for $\tilde{b}_i > 0$ iff $\frac{\partial}{\partial \tau} \pi_i(\tilde{b}_i, \tau \tilde{b}_{-i}, \tilde{q})|_{\tau=1} = -\tilde{b}_i$ \square

Using Lemma 5, we can rewrite the first order conditions by defining a mapping $\tilde{b} = \lambda(\tau)$ that exists in some neighborhood of $\tau = 1$:

$$\tau \frac{d}{d\tau} \pi_i(\lambda_i(\tau), \tau \lambda_{-i}(\tau), q) = -\tilde{b}_i \quad (\text{A.14})$$

We let $V = [0, v_1] \times \dots \times [0, v_k]$ be the support of potential bids of players 1 to k , and define $D_0(\tilde{b}, \tilde{q}) = \frac{\partial}{\partial \tilde{b}} \pi(\tilde{b}, \tilde{q})$ with the diagonal elements replaced with zeros. The following theorem from Athey and Nekipelov (2010) establishes the conditions under which the mapping $\lambda(\tau)$ exists locally around $\tau = 1$ and globally for $\tau \in [0, 1]$, which yields the equilibrium bids of the players.

Theorem 1 (Athey and Nekipelov (2010)). *Assume that D_0 is continuous in \tilde{b} . Suppose that for each $i = 1, \dots, k$, $t_i(\tilde{b}, \tilde{q}) > 0$, and that each π_i is quasi-concave in \tilde{b}_i on V and for each \tilde{b} its gradient contains at least one non-zero element. Then*

1. *An equilibrium exists if and only if for some $\delta > 0$ the system of equations (A.14) has a solution on $\tau \in [1 - \delta, 1]$.*
2. *The conditions from part 1 are satisfied for all $\delta \in [0, 1]$ and so an equilibrium exists, if $D_0(\tilde{b}, \tilde{q})$ is locally Lipschitz and non-singular for $\tilde{b} \in V$ except for a finite number of points.*
3. *There is a unique equilibrium if and only if for some $\delta > 0$ the system of equations (A.14) has a unique solution on $\tau \in [1 - \delta, 1]$.*
4. *The conditions from part 3 are satisfied for all $\delta \in [0, 1]$, so that there is a unique equilibrium, if each element of $\frac{\partial}{\partial \tilde{b}} \pi(\tilde{b}, \tilde{q})$ is Lipschitz in \tilde{b} and non-singular for $\tilde{b} \in V^1$.*

The theorem shows that under very general conditions, websites would spend non-zero efforts on SEO in equilibrium. We now proceed to analyze how positive levels of SEO effectiveness α affect the satisfaction of consumers from the ranking of the organic list. To analyze the incentives of the different websites, it is easier to transform the multiple links contest into a game where websites choose the amount of traffic they would like to acquire from organic clicks, which implicitly determines their bids. We define the vector of traffic for each site i given the SEO effectiveness α and the vector of bids \tilde{b} as $t^\alpha(\tilde{b}) = (t_1^\alpha(\tilde{b}), \dots, t_n^\alpha(\tilde{b}))$. For each player i , fixing the bids of other players as b_{-i} , we can rewrite the first order condition of each player as $\frac{\partial \pi_i}{\partial t_i} = 0$. The expected utility of consumers when searching through links with traffic vector t^α is $EU(t^\alpha) = \sum_i q_i t_i^\alpha$.

Analyzing the result of the SEO game with multiple links is hard. In addition, under certain conditions, such as when the errors are small or α is very large, multiple equilibria might exist as shown in Siegel (2009). We therefore proceed to analyze the special cases

¹Athey and Nekipelov (2010) give example conditions for the non-singularity of the matrix D_0 in their Lemma 2.

defined by Theorem 1 where an internal equilibrium exists for all players and the first order conditions hold for players in equilibrium. For every α we define $\mathbb{T}^\alpha = \{t^\alpha | EU(t^\alpha) \geq EU(t^0)\}$ as the group of all traffic distributions over sites where the expected consumer utility is higher than under the benchmark traffic distribution t^0 .

The following proposition shows that under certain conditions, a positive level of SEO can improve consumer satisfaction. These conditions are sufficient, but by no means necessary. We conjecture that much weaker conditions can be found under which SEO improves consumer satisfaction.

Proposition 9. *For each α such that there exists a vector of non-negative functions $M(t) = (M_1(t), \dots, M_k(t))$ with*

$$\frac{M_i(t)}{M_{i+1}(t)} > \frac{t_i \left| \frac{\partial \pi_{i+1}}{\partial t_{i+1}}(t) \right|}{t_{i+1} \left| \frac{\partial \pi_i}{\partial t_i}(t) \right|} \quad (\text{A.15})$$

for every $t \in \mathbb{T}^\alpha$ and $\frac{\partial \pi_{i+1}}{\partial t_{i+1}}(t) \neq 0$, the equilibrium distribution of traffic t^{α} satisfies $EU(t^{\alpha*}) > EU(t^0)$*

Proof. Recall that \mathbb{T}^α contains all traffic distributions $t = (t_1, \dots, t_n)$ for which the expected utility of consumers is weakly greater with an SEO effectiveness level of α than with $\alpha = 0$, implying $EU(t) = \sum_i q_i \cdot t_i \geq \sum_i q_i \cdot t_i^0$.

Let $\beta = \sum_j \beta_j$ be the sum of the exogenous click-through rates. If we normalize the sum of clicks $\sum_i t_i$ to 1 we have $\beta = \sum_i t_i$. We then define, for each α , the mapping

$$F_\alpha : \frac{(t_1, \dots, t_n)}{\beta} \rightarrow \frac{(M_1(t) \cdot \left| \frac{\partial \pi_1}{\partial t_1}(t) \right| + t_1, \dots, M_n(t) \cdot \left| \frac{\partial \pi_n}{\partial t_n}(t) \right| + t_n)}{\beta + \sum_i M_i(t) \left| \frac{\partial \pi_i}{\partial t_i}(t) \right|}.$$

Above, for convenience of notation, α was dropped and the first orders $\frac{\partial \pi_i}{\partial t_i}$ as well as the traffic distributions t_i are given under the specific α for each F_α . To simplify exposition we assign $\tilde{t} = \frac{t}{\beta}$ as the normalized traffic vector. This mapping has several special properties:

- The mapping maps a given traffic distribution to another, implicitly setting the required bids to reach this traffic distribution. The input and output distributions are normalized to one, so the mapping is closed on traffic distributions. In addition, the mapping is continuous.
- The fixed points of each mapping F_α are the equilibrium distributions of the SEO game. To see this, note that when the first order conditions hold and are equal zero, the mapping has a fixed point, and vice-versa.
- The set of traffic distributions superior to $U(t^0)$ (which is \mathbb{T}^α) is convex.

As a result, showing that the fixed points of F_α are superior to t^0 would prove that SEO increases consumer utility in equilibrium. To see this, let $t \in \mathbb{T}^\alpha$. Then

$$U(F(t)) - U(\tilde{t}) = \sum_i q_i \left(\frac{t_i + M_i \left| \frac{\partial \pi_i}{\partial t_i} \right|}{\beta + \sum_j M_j \left| \frac{\partial \pi_j}{\partial t_j} \right|} - \frac{t_i}{\beta} \right) = \sum_i q_i \frac{\beta M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| - t_i \sum_j M_j \left| \frac{\partial \pi_j}{\partial t_j} \right|}{\beta(\beta + \sum_j M_j \left| \frac{\partial \pi_j}{\partial t_j} \right|)} \quad (\text{A.16})$$

As $M_i(t)$ are non-negative and $\beta = \sum_i t_i$, the difference in utilities is positive when:

$$\sum_i q_i \left(\beta M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| - t_i \sum_j M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| \right) = \sum_j t_j \left(\sum_i M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| (q_i - q_j) \right) > 0 \quad (\text{A.17})$$

Fix i, j and assume $i < j$, then $q_i \geq q_j$. Looking at the couples of additions in the sum for i, j we get

$$t_j M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| (q_i - q_j) + t_i M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| (q_j - q_i) = \left(t_j M_i \left| \frac{\partial \pi_i}{\partial t_i} \right| - t_i M_j \left| \frac{\partial \pi_j}{\partial t_j} \right| \right) (q_i - q_j) \quad (\text{A.18})$$

which is larger than zero when condition (A.15) holds.

This shows that the set \mathbb{T}^α is convex and closed under the continuous mapping F_α . As a result, Brouwer's fixed point theorem tells us that a fixed point of F_α exists in \mathbb{T}^α , which concludes the proof. \square

The conditions in (A.15) imply that the sequence of bounding function limits the changes in profits of the different players from increased organic traffic. As a result, the existence of such a sequence means that extra traffic does not yield “too steep” changes in players profits and thus their incentives to decrease their expected amount of clicks in equilibrium. In such cases, allowing $\alpha > 0$ improves consumer satisfaction from the resulting quality of ranking and increases total traffic to the search engine.

A.5 Simultaneous SEO and Sponsored Auction

In this section we present a robustness check by examining a model where decisions on the SEO investments and the sponsored auction bids are made simultaneously. The setup is otherwise identical to what we present in Section 2.2 of the paper. We focus on the case when r is sufficiently low and rederive the results of Section 2.3 and show that the results do not change.

Proposition 10. *When (i) the decisions about the sponsored auction and SEO are made simultaneously, (ii) consumers have a small, but positive search cost c , and (iii) $r < v_L$:*

1. *The game has a unique equilibrium in which all consumers start their search with the organic link.*

2. *The likelihood of a high quality organic link is increasing in α for any $-1 \leq \rho \leq 1$.*
3. *The search engine's revenue increases in α iff the likelihood of a high quality organic link decreases.*

Proof. The proof is very similar to that of Proposition 2. We focus on the case when sites have different qualities ($q_1 = q_H > q_2 = q_L$), otherwise consumers are indifferent about which site is displayed. We begin by assuming that consumers start with the organic link. Since sites make simultaneous decisions affecting their position both on the organic and sponsored side, we need to determine how much profit they make in each of the possible cases. First, we note that since consumers are actively searching for a high quality link, the low quality site will only be able to attract any customers if it possesses both links. However, its sponsored CTR will be zero, since consumers recognize that it is the same link as in the organic position. In every other case, the high quality site will get at least one link driving all the demand there. If it gets the organic link, its revenue (net of SEO investments) will be v_1 , whereas if it gets only the sponsored links, it will have to pay the price for the sponsored link and its revenue (net of SEO investments, but including sponsored payments) will be $v_1 - r$.

Now consider potential equilibrium strategies. Each site has to specify an SEO investment and a bid for the sponsored link. Clearly, the low quality site's only chance to make positive profits is to obtain the sponsored link. However, its ctr would be 0 if it did, forcing it out of the sponsored side. That is, the low quality site (site 2) will not be able to make positive profits, hence its SEO investment will be 0. Due to the uncertainty in the SEO process, site 2 may still get the organic link, thus site 1 has an incentive to invest in SEO. By submitting a high enough bid to surpass the minimum bid (which we assume is low), its payoff will be

$$v_1 - r + r \Pr(s_1 > s_2) - \frac{e_1^2}{2}.$$

Aside from the fixed $v_1 - r$ that the site make regardless of its SEO investment, this a special case of what we saw in equation (2.2) in the paper. Site 1 will thus behave as if it had a valuation r for the organic link, while its opponent had 0. The likelihood of site 1 acquiring the organic link will be $P(\alpha; \sigma, r, 0, q_H, q_L)$ which is increasing in α regardless of ρ . This proves Part 2. For Part 1, it is easy to see that consumers are better off starting on the organic side in this equilibrium. Similarly to the proof of Proposition 2, we can prove that an equilibrium where consumers start on the sponsored side does not exist by redoing the above steps assuming that they do start on the sponsored side. Finally, to prove Part 3, it is trivial to see the search engine makes less money if the high quality site acquires the organic link and as SEO becomes more efficient, this is more likely. \square

A.6 Heterogeneous Search Costs

Although search costs play an important role in consumers' searching behavior, we did not fully explore their role in the paper. In this section we analyze a more realistic structure of

heterogeneous search costs across consumers. An important advantage of this setup is that it allows us to examine consumers' decision to visit the search engine and to understand how SEO affects the search engine's traffic.

Instead of fixing each consumer's search cost at $c \geq 0$, we now assume that consumers have potentially different non-negative search costs distributed according to a distribution with a support of $[0, \infty)$ and a differentiable c.d.f., G . An important implication of having consumers with different search costs is that some of them might have relatively high costs so that they would only want to visit a single link. This leads to the emergence of an equilibrium where consumers start their search with the sponsored link. We distinguish the different types of equilibria depending on which side consumers start the search. We call the equilibrium where consumers start with the organic link an O-type equilibrium, and we call an equilibrium S-type if consumers start on the sponsored side.

Proposition 11. *There is always one O-type equilibrium in which consumers start with the organic link. When ρ is high enough and a large enough proportion of consumers have high search costs, there is a second, S-type equilibrium in which consumers start with the sponsored link.*

Proof. We begin by showing that there is an O-type equilibrium, similarly to the proof of Proposition 2. When consumers start with the organic link only the advertiser who does not acquire the organic link will have a chance to get the sponsored link. When a high quality player is in the organic position, the low quality competitor will not benefit from the sponsored link. When a low quality player obtains the organic link, consumers with a low search cost will click the sponsored link to find out if it is higher quality. Let $\hat{c}(p)$ denote the expected benefit of continuing to the sponsored link where p is the probability that a high quality advertiser obtains the organic link when advertisers have different qualities and valuations. Thus, consumers whose search costs is lower than the above benefit will search. The proportion who continues is $\vartheta(p) = G(\hat{c}(p))$ which is continuous in p . Performing the same calculations as in the proof of Proposition 2, we get that the value for the site with the high quality (denoted as site 1) to get the organic link is $(1 - \vartheta(p))v_1 + \vartheta(p)r$, whereas site 2's value is $(1 - \vartheta(p))v_2$. Using the function $P(\alpha, v_i, v_j, q_H, q_L)$ from the proof of Proposition 8, we obtain an equilibrium by solving $p = P(\alpha, (1 - \vartheta(p))v_1 + \vartheta(p)r, (1 - \vartheta(p))v_2, q_H, q_L)$. Since the derivative of the continuous $P()$ function is less than 1 and the function takes a positive value at $p = 0$ and less than 1 at $p = 1$, we obtain a unique solution. As long as α is not too high, the organic link will be more likely to be high quality than low quality. Therefore, consumers do not have an incentive to deviate and start with the sponsored link.

To show that existence of an S-type equilibrium assume that consumers start with the sponsored link. When the organic link is obtained by the high quality site the sponsored competition will be for the $(1 - \vartheta)$ proportion of consumer who only click on the first (sponsored) link they encounter. When the low quality site obtains the organic link, the sponsored competition is for all consumers (as the high quality either gets them all or none). The sponsored link will thus always go to the advertiser with the higher per-click valuation

(as CTR's will be the same, 1 for both players) as long as the minimum bids are exceeded, otherwise there will be no sponsored link. In order for the minimum bid to be exceeded, we need $v_H(1 - \vartheta) > r$, that is $1 - \vartheta > r/v_H$. This condition is satisfied if enough consumers have a high enough search cost so that they would never search, for example, $1 - G(q_H - q_L) > r/v_H$. If ρ is high enough then the player with the higher valuation is likely to be the high quality advertiser. This makes it rational for consumers to start with the sponsored link, as there is always a positive probability that the organic link will be acquired by the low quality player. \square

The first type of equilibrium is a direct generalization of the one described in Proposition 2. If consumers start with the organic link, the sponsored link only serves as a backup and the high quality advertiser has a higher chance of getting the organic link. Therefore, sites take SEO seriously and the organic link will offer a higher expected quality to consumers who rationally start their search there. However, when there are enough people with high search costs who will never click more than one link there is an equilibrium in which consumers start with the sponsored link. If sites expect a significant proportion of consumers to only click the sponsored link they will take the sponsored auction seriously. If the site with the higher quality is more likely to have a higher valuation (high ρ), it will win the sponsored link no matter who acquires the organic link. Therefore, it is rational for consumers to start with the organic link. SEO is not as important in the S-type equilibrium since most of the competition will happen on the sponsored side and the organic link serves as a backup.

Although the S-type equilibrium only exists for a limited parameter range, it is at least as important as the O-type equilibrium. When some consumers do very limited searches and advertisers' qualities are correlated with their valuations for consumers, it is plausible for consumers to start with the sponsored link and advertisers to fight hard for them. The multiplicity of equilibria may indeed explain the substantial differences observed between sponsored click-through rates for different keywords (Jeziorski and Segal, 2009).

Taking the above analysis a step further, we directly examine how search costs impact the outcome of SEO. In order to compare different search cost distributions, let $G_1 \succ G_2$ denote the relation generated by first-order stochastic dominance.

Corollary 6. *Suppose $G_1 \succ G_2$.*

1. *The likelihood of a high quality organic link in the O-type equilibrium is higher (lower) under G_1 than under G_2 for high (low) values of ρ .*
2. *In the S-type equilibrium the likelihood of a high quality organic link is lower under G_1 than under G_2 .*

Proof.

Part 1: Since $G_1 \succ G_2$, we have $\vartheta_1(p) = G(\hat{c}(p)) \leq \vartheta_2(p) = G_2(\hat{c}(p))$. That is, when search costs are higher, fewer consumer continue to the sponsored link. To determine the probability of a high organic link, we obtain the solution of $p = P(\alpha, (1 - \vartheta(p))v_1 + \vartheta(p)r, (1 -$

$\vartheta(p)v_2, q_H, q_L)$ as in the proof of Proposition 11. Note from the proof of Proposition 8 that $P(\alpha, v_i, v_j, q_H, q_L)$ is increasing in $v_i - v_j$. In our case $v_i - v_j = (1 - \vartheta)(v_1 - v_2) + \vartheta r$ which is decreasing in ϑ when $v_1 > v_2$, that is when $\rho = 1$ and increasing when $v_1 < v_2$, that is when $\rho = -1$. Since $\vartheta_1(p)$ is lower than $\vartheta_2(p)$, the solution for $G_1()$ is higher (lower) than for G_2 when ρ is high (low).

Part 2: The case of an S-type equilibrium is very similar, but this type of equilibrium only exists when ρ is high. When the high quality advertiser has a high valuation, its benefit of getting the organic link is ϑv_L , the extra sponsored payment it would have to incur when not getting the organic link. The low quality player has 0 valuation for the organic link, therefore $v_i - v_j = \vartheta v_L$ which is increasing in ϑ , which completes the proof the same was as in Part 1. \square

The results illustrate the different roles of SEO in the two types of equilibria. When consumers start searching with the organic link, higher search costs generally lead to tougher competition in SEO. The intuition is that higher search costs decrease consumers' search incentives and advertisers' only chance to attract an increasing proportion of consumers is through the organic link. Therefore, when valuations and qualities are correlated higher search costs lead to a higher SEO investment on the high quality advertiser's part. Consequently, a lower percentage of consumers will move on to the sponsored side. This hurts the high quality advertiser when its low quality competitor possesses the organic link and incentivizes it to invest more in SEO. The opposite is true when valuations and qualities are negatively correlated: as search costs go up the low quality advertiser (now with a high valuation) has an increased incentive to fight for the organic link that becomes the only link that an increasing proportion of consumers clicks on.

On the other hand, when consumers start with the sponsored link, the organic link only serves as a backup. As search costs go up, fewer consumers continue to the organic link, therefore its importance declines. Since this equilibrium only exists for correlated qualities and valuation the smaller number of click on the organic link will reduce the high quality advertiser's incentive and chance of obtaining the organic link.

Finally, in addition to examining the differences in consumers searching behavior once they arrive to the search engine, we also study their initial decision to visit. Comparing their expected net payoff from visiting the search engine with an outside option of 0 allows us to determine the amount of traffic and revenue the search engine receives.

Corollary 7. *The search engine's revenue is always decreasing as a high quality organic link becomes more likely, even when the traffic to the search engine is increasing.*

Proof.

In case of an O-type equilibrium: The expected benefit of moving on to the sponsored link when encountering a low quality organic link is

$$\hat{c} = (q_H - q_L) \frac{(1 - \rho)^2(1 - p_1) + (1 - \rho)^2(1 - p_{-1}) + (1 + \rho)(1 - \rho)(2 - p_L - p_H)}{(1 - \rho)^2(1 - p_1) + (1 - \rho)^2(1 - p_{-1}) + (1 + \rho)(1 - \rho)(2 - p_L - p_H) + 2},$$

where p_1 is the probability of a high quality organic link when the advertisers have different qualities and perfectly correlated valuations, p_{-1} when they perfectly negatively correlated valuations, p_H when both of them have high valuations and p_L when both of them have low valuations. The person with a search cost of \hat{c} will be therefore indifferent between stopping and continuing. The same person will have an expected benefit of $\hat{c}' = \frac{q_H}{8}(2 + (1 + \rho)^2 p_1 + (1 - \rho)^2 p_{-1} + (1 + \rho)(1 - \rho)(p_H + p_L))$ from visiting the search engine. Since $\hat{c}' > \hat{c}$, some consumers will stop searching after visiting the organic link. As the p values increase, the benefit from visiting the search engine increases, but the \hat{c} threshold for clicking the sponsored link decreases in each p . Therefore, even though the traffic to the search engine strictly increases as any or all of the p values increases, the search engine's revenue will strictly decrease (as each click generates a revenue of r).

In case of an S-type equilibrium: The expected benefit of moving on to the organic link when encountering a low quality sponsored link is $\hat{c} = (q_H - q_L) \frac{2p_{-1}(1-\rho)^2 + (1+\rho)(1-\rho)}{2(1+\rho)^2 + 4(1-\rho)^2 + 2(1+\rho)(1-\rho)}$. The same person will have an expected benefit of $\hat{c}' = \frac{q_H}{16}(4 + 3(1 + \rho)^2 + (1 - \rho)^2)$ from visiting the search engine. Since $\hat{c}' > \hat{c}$ when the S-type of equilibrium exists ($\rho > 0$ is necessary), the highest search costs visitors will only click on the sponsored link. Therefore, high search cost consumers will not benefit from an increased expected organic quality and traffic, thus revenue will not increase. \square

This result sheds more light on the fundamental tension between the search engine and its visitors in their preference for a high quality organic link. We have already identified the basic misalignment of incentives in the last part of Proposition 2, but in that case the traffic to the search engine was exogenously fixed. Here, we show that even though a high quality organic link makes consumers better off and attracts more traffic, it does not increase revenues. The intuition is based on how consumers search. In the O-type equilibrium those with a search cost below the expected benefit from visiting the search engine make the first (organic) click. However, not all of them make the second (sponsored) click that would generate revenue as the expected benefit of the second click is always lower than the first one. Therefore, even though the search engine can attract more visitors by having a higher expected quality organic link, the extra visitors will not generate revenue. In the S-type equilibrium all visitors generate revenue, but the promise of a higher organic link will not attract more visitors, as it only benefits low search cost consumers who visit anyway.

In both cases the misalignment between consumers' and the search engine's incentive is clear. Although a higher quality organic link increases consumer welfare, it reduces the search engine's revenue. This phenomenon may explain why large search engines, such as Google, take a stance against SEO that might potentially improve the quality of search results.

A.7 Heterogeneous Preferences

Throughout the paper we have assumed that advertisers are vertically differentiated and that consumers have homogeneous preferences. However, an important challenge that search engines face is that different consumers would rank sites in a different order. Here, we examine the implications of SEO under heterogeneous consumer preferences. We modify our basic model, by distinguishing between mainstream sites niche sites instead of high and low quality. We assume that a $1 > \beta > 1/2$ proportion of consumers are mainstream who derive a q_H utility from visiting a mainstream site and q_L from visiting a niche site. The remaining $1 - \beta$ consumers have the opposite preferences.

Since the majority of consumers prefers the mainstream sites, we assume that the search engine intends to put a mainstream site into the organic position if one exists. Identically to our basic model, we assume that there are two sites which can be either mainstream or niche with equal probability and that sites can have a valuation of v_H or v_L for a consumer. We assume $v_H > v_L$ are sufficiently different and that the minimum bid, r is small.² The correlation between whether a site is mainstream and its valuation for consumers is given by ρ as before. As in Section 2.3 of the paper, we assume that all consumers have the same small search cost.

Since consumers know their preference-type (mainstream vs. niche), they will have different search strategies depending on which type they prefer.

Proposition 12.

1. *When ρ is a sufficiently low negative correlation, mainstream consumers will start with the organic link whereas niche consumers will start with the sponsored. A mainstream organic link is less likely as α increases.*
2. *When ρ is sufficiently high positive correlation, mainstream consumers will start with the sponsored link whereas niche consumers will start with the organic. A mainstream organic link is more likely as α increases.*

Proof. We only need to examine the case when the two sites are of different type. If $v_H > \max\left(\frac{1}{\beta}, \frac{\beta}{1-\beta}\right) v_L$ and at least one consumer groups starts with the sponsored link, it is straightforward to show that the advertiser with the highest valuation acquires the sponsored link. Therefore, for low values of ρ the sponsored link is likely to be niche, whereas for high values of ρ it is likely to be mainstream. Due to the error in the SEO process both sites have a positive chance to get the organic link, therefore niche (mainstream) consumers will start with the sponsored link in the former (latter) case and mainstream (niche) will start with the organic link. The total sponsored payment will be $v_L(1 - \beta)$ when the mainstream site gets the organic link and $v_L\beta$ otherwise. In the case of $\rho = 1$ the valuations for getting the organic

²We assume $v_H > \max\left(\frac{1}{\beta}, \frac{\beta}{1-\beta}\right) v_L$ and $r < (1 - \beta)^2 v_L$. Note that for fixed v_L, v_H, r these conditions limit the value of β , therefore the results of this section only hold under a sufficient level of heterogeneity.

link will be $v_H(1 - \beta) + (2\beta - 1)v_L$ for the mainstream site and $v_L(1 - \beta)$. Since $\beta \geq 1 - \beta$, the former is higher and SEO increases the chance of a mainstream organic link when ρ is close to 1. When $\rho = -1$, the value of the organic link will be βv_L and $\beta v_H + (1 - 2\beta)v_L$ for the mainstream and niche, respectively. In the assumed parameter range the latter is higher resulting in a more likely niche organic link when ρ is close to -1 . \square

The results contribute to our understanding of the different types of equilibria in the previous section. We again identify two types of equilibria; one in which the majority of consumers (mainstream) start with the organic link and another one in which the majority of consumers start with the sponsored link. However, the equilibria in this case are unique in each of the above parameter regions. The first part examines a typical mainstream-niche scenario, where the mainstream advertiser cannot command as high of a margin as its niche competitor. The high valuation of the niche firm will ensure its position in the top sponsored position incentivizing niche customers to start searching on the sponsored side. The mainstream product will have an advantage on the organic side as the search engine wants to cater to the majority. However, the niche firm will invest more heavily in SEO which decreases the likelihood of a mainstream organic link. SEO in this case decreases consumer welfare,³ but increases the search engine's revenue.

The second part identifies a more surprising scenario. When the product preferred by the majority is able to command a higher margin then the majority of consumers start with the sponsored link. We think of this type of market as one with a minority of consumers who are well informed about a product category and a majority who are less informed. The less informed customers can be charged a higher price which the more informed customers are not willing to pay. Even though the majority of the customers start with sponsored link, SEO will still increase consumer welfare by increasing the likelihood of a majority preferred organic link. However, this reduces sites' incentives to pay for the sponsored link, hurting the search engine's revenues. It is interesting to contrast these results with Proposition 2 where consumers always start with the organic link. What makes this case different is a sufficient level of consumer heterogeneity that leads some customers to focus on the sponsored link, inducing more advertiser competition on the sponsored site, which in turn leads to more consumer attention to the sponsored side. The transition is not continuous and the required level of heterogeneity depends on the minimum bid r . Since the search engine is clearly better off under the heterogeneous outcome, it should pay particular attention to setting the minimum bid.

³Mainstream consumers are worse off, whereas niche consumers are indifferent.

Appendix B

Appendix for Chapter 3

B.1 Proofs

Proof of Proposition 3. To find p^A , notice that the profit of the advertiser is $(2q^A)^\rho(1-2p)$. Since $q^A \sim p^{\frac{1}{2-\rho}}$, we can drop the constants and solve for $p^A = \arg \max_p p^{\frac{\rho}{2-\rho}}(1-2p)$, yielding $p^A = \frac{\rho}{4}$, and $q^A = \frac{\left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}}{2}$. The second order condition of each agent is:

$$\rho(\rho-1)(q_i + q^A)^{\rho-2} p^A - 1 < 0 \quad (\text{B.1})$$

For $\rho \leq 1$ it always holds, while for $1 < \rho < 2$ it holds if $q_i > \left(\frac{\rho^2(\rho-1)}{4}\right)^{\frac{1}{2-\rho}} - q^A$ after plugging p^A and collecting terms. The right hand side is negative if $2^{-\rho-1}(\rho-1) < 1$, which holds for every $1 < \rho < 2$.

To show that $q^* > q^M$, we notice that $\frac{q^*}{q^M} = 2^{\frac{1}{2-\rho}} > 1$ for $0 < \rho < 2$. Similarly, $\frac{q^M}{q^A} = \left(\frac{2}{\rho}\right)^{\frac{1}{2-\rho}} > 1$ for $0 < \rho < 2$, which proves part 1 of the proposition.

To prove part 2, since $q^M > q^A$, the total revenue generated by the advertiser $x(q_1, q_2)$ is always larger under CPM. The share of profit given to the publisher under CPM is $\frac{2(q^M)^2}{(2q^M)^\rho} = \frac{\rho}{2}$. This is the same share p^A given under a CPA contract. As a result, since revenues are strictly larger and the same share is given, profits under CPM are larger.

To prove part 3, the difference in profit $u^A - u^M$ of the publisher has a numeric root on $[0, 2]$ at $\rho^c = 0.618185$. The function has a unique extremum in this range at $\rho = 0.246608$, which is a local maximum, and the difference is zero at $\rho = 0$. Thus, it is positive below ρ^c and negative above ρ^c proving part 3. \square

Proof of Corollary 3. In the single publisher case, $q^M = p^M$ similarly to before, and solving the advertiser optimization problem yields $q^M = \left(\frac{\rho}{2}\right)^{\frac{1}{2-\rho}}$. Under CPA, $q^A = \left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}$. We immediately see that $q^A > q^M \iff \rho > 1$.

The share of revenue given as payment to the publishers equals $\frac{\rho}{2}$ in both cases. As a result, when $\rho > 1$, $\pi^A > \pi^M$, and vice versa when $\rho < 1$. \square

Proof of Proposition 4 and Corollary 4. For completeness, we specify the resulting distribution function, $f_1(\frac{q_1}{q_2})$:

$$f_1\left(\frac{q_1}{q_2}\right) = \begin{cases} 1 & q_1 \geq dq_2 \\ -\frac{d^2 q_2^2 - 2((d-1)d+1)q_1 q_2 + q_1^2}{2(d-1)^2 q_1 q_2} & dq_2 < q_1 < dq_2 \\ \frac{(q_2 - dq_1)^2}{2(d-1)^2 q_1 q_2} & \frac{q_2}{d} < q_1 < q_2 \\ \frac{1}{2} & q_1 = q_2 \\ 0 & q_1 \leq \frac{q_2}{d} \end{cases} \quad (\text{B.2})$$

In a symmetric equilibrium, $f'(1) = \frac{1}{2} \frac{d+1}{d-1}$. Solving for the symmetric equilibrium allocation of the publishers under CPA reaches $q^{A-LT} = (2^{\rho-1} p [\frac{d+1}{d-1} + \frac{\rho}{2}])^{\frac{1}{2-\rho}}$. Plugging into the advertiser optimization problem, we find that $p^A = \frac{\rho}{2}$, yielding the values of q^A specified in the proposition. The SOC at the symmetric point is only negative for $\rho < \frac{3}{2} + \frac{1}{2} \sqrt{\frac{7+25d}{d-1}}$. The profit is positive for $\rho < 2 - \frac{4}{d-1} < \frac{3}{2} + \frac{1}{2} \sqrt{\frac{7+25d}{d-1}}$, proving this is an equilibrium.

A simple comparison to q^* and among q^{A-LT} and q^M yields the conditions in the proposition and finalizes the proof.

Since $q^{A-LT} > q^M > q^A$ and the share of revenues given to the publishers under each scheme is equal, the profit under last touch attribution is higher. \square

Proof of Proposition 5. The first order condition the publisher faces in a symmetric equilibrium is: $\frac{\rho(2q)^{\rho-1} + \rho q^{\rho-1}}{2} p^A = q$. The solution, after calculating the equilibrium share offered by the principal is: $q^{A-S} = \left(\frac{\rho^2}{4}(2^{\rho-1} + 1)\right)^{\frac{1}{2-\rho}}$.

The second orders are negative at $q_1 = 0$ and at $q_1 \rightarrow \infty$, while the third order is always negative between these two, implying the second order condition holds. To prove part 2, we recall that $q^{A-LT} = (2^{\rho-1} \frac{\rho}{2} [\frac{d+1}{d-1} + \frac{\rho}{2}])^{\frac{1}{2-\rho}}$ and $q^A = \left(\frac{\rho^2}{2}\right)^{\frac{1}{2-\rho}}$. In this case, $q^{A-S} > q^A$ iff $\frac{1}{2}(2^\rho + 1) > 1$ which holds for every $0 < \rho < 2$. $q^{A-S} > q^{A-LT}$ always when $\rho < 2 - \frac{4}{d-1}$. Comparing to the CPM quantity, $q^{A-S} > q^M$ iff $\rho > 1$.

Finally, since the share of revenue given by the advertiser to the publishers is $\frac{\rho}{2}$, which is equal to the share given under regular CPA campaigns and under last-touch attribution, we find that profit is higher for Shapley value attribution when q^{A-S} is highest. \square

Proof of Lemma 1. Maximizing the expectation w.r.t to s yield $q^E = 1 - \mu$.

Plugging in q^E yields the profit $\pi_{min} = \frac{N}{2} ((\mu - 1)^2 + 2\mu)$, which is smaller than π^{max} by $\frac{N}{2} (\mu - \mu^2)$. \square

Proof of Lemma 2. The proof was built for $\theta \in [0, 1]$ assuming the effectiveness of advertising is θq . In the text $\theta = 1$.

The total profit from experimenting is:

$$\frac{\beta\theta^2(2\alpha + \beta^2 + \beta(\alpha + N + 1) + N) - 2\beta\theta^2\sqrt{\alpha(\beta + 1)(\alpha + \beta + N)} + 2\alpha N(\alpha + \beta + 1)}{2(\alpha + \beta)(\alpha + \beta + 1)} \quad (\text{B.3})$$

The second order with respect to n is $-\frac{\alpha\beta\theta^2(\alpha+\beta+N)}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+n)^3}$ and is negative for all $\alpha > 0, \beta > 0$ and $N > 0$. At $n = 0$, the first order is positive when $N > \frac{\beta(\alpha+\beta)(1+\alpha+\beta)}{\alpha}$ implying the optimal sample size is positive.

The solution to the first order condition is $n^* = \sqrt{\frac{\alpha(N+\alpha+\beta)}{1+\beta}} - (\alpha + \beta)$ which is independent of θ . Finally, calculation of the change with respect to α, β and N yield the conditions stated in the lemma. \square

Proof of Proposition 6. The difference in profit from π^{min} is

$$\frac{\beta\theta^2 \left((\alpha + \beta) \left(\alpha(\beta + 2) + \beta^2 + \beta - 2\sqrt{\alpha(\beta + 1)(\alpha + \beta + N)} \right) + \alpha N \right)}{2(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (\text{B.4})$$

Whenever $N > \beta \frac{1-\mu}{\sigma^2}$, the firm can achieve this difference by Lemma 2. It can be verified that this difference is positive for all $\alpha > 0$ and $\beta > 0$, and is increasing in θ . Specifically, it is positive for $\theta = 1$ when N is large enough. \square

Proof of Proposition 7. In a CPM campaign, the publisher can choose to show q_b ads to baseline consumers and q ads to non-baseline consumers. The profit of the publisher is:

$$u = N \left[(q_b s + q(1 - s))p^M - s \frac{q_b^2}{2} - (1 - s) \frac{q^2}{2} \right] \quad (\text{B.5})$$

Maximizing the profit of the publisher yields $q_b = q = p^M$. Plugging into the advertiser's profit and maximizing over the expectation of s yields $p = \frac{1-\mu}{2}$, resulting in an advertiser profit of $N \left[\mu + \frac{(1-\mu)^2}{4} \right]$.

Performing a similar exercise for a CPA campaign, the publisher will opt not to show ads to baseline consumers, as it receives commission for their conversions regardless of showing them ads. Maximizing the publisher's profit yields $q_b = 0$ and $q = p^A$, which yields $p = \frac{1-2\mu}{2(1-\mu)}$ when plugged into the advertiser's profit and maximized. This value is higher than 0 only for $\mu < \frac{1}{2}$, and for $\mu > 1/2$ the advertiser will prefer not to use a CPA campaign. Comparing q^M to q^A and q^* yields the second part of the proposition. The profit of the advertiser is then $N \frac{1}{1-\mu}$ which is lower than the CPM profit for any $\mu < \frac{1}{2}$, concluding the proof. \square

Proof of Proposition 14. The following lemma from McAfee and McMillan (1991) establishes conditions for payments offered by the advertiser to maximize its profit, and applies to our model:

Lemma 6 (McAfee and McMillan (1991) Lemma 1). *Suppose the payment functions b_i satisfy*

$$\mathbb{E}_{x, \theta_{-i}}[b_i(\theta)]|_{\theta_i=0} = 0 \quad (\text{B.6})$$

and evoke in equilibrium outputs $y^(\theta)$. Then the payments maximize the advertiser profits subject to publisher individual rationality and incentive compatibility.*

Theorem 2 of McAfee and McMillan (1991) shows that the payments in (B.17) yield an optimal mechanism under the conditions that agents are complements in production and when $\frac{\partial y_i^*(\theta)}{\partial z_j} \geq 0$ for $j \neq i$. These conditions do not apply in our case when publishers are substitutes in production.

We therefore proceed to show that the linear mechanism is optimal also under substitute production. To prove these payments yield an optimal truthful mechanism, we first prove the following:

Lemma 7. *The optimal allocation of efficiency units $y^*(\theta)$ is unique, positive for positive effectiveness and increases with self-reported type and decreases with reported types of other publishers:*

- $\frac{\partial y_i^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \geq 0$
- $\frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \leq 0$ for $j \neq i$

Proof. Let $\pi(y, \theta) = x(y) - \gamma_1(y_1, \theta_1) - \gamma_2(y_2, \theta_2)$.

The first order conditions are:

$$1 - \frac{2 - \theta_i y_i}{\theta_i^3} - y_{-i} = 0 \quad (\text{B.7})$$

With solutions:

$$y_i^* = \frac{\theta_i^3 (\theta_{-i}^3 + \theta_{-i} - 2)}{\theta_i^3 \theta_{-i}^3 - \theta_i \theta_{-i} + 2\theta_i + 2\theta_{-i} - 4} \quad (\text{B.8})$$

These are positive for $1 \geq \theta_i > 0$.

The first principal minors of π are zero and the second is positive. As a result π is negative semidefinite, and y^* is the unique maximum.

Using the implicit function theorem:

$$\frac{\partial y_i^*}{\partial \theta_i} = \frac{2(\theta_i - 3)(2 - \theta_{-i})y_1}{\theta_i (\theta_i^3 \theta_{-i}^3 - \theta_i \theta_{-i} + 2\theta_i + 2\theta_{-i} - 4)} \geq 0 \quad (\text{B.9})$$

$$\frac{\partial y_i^*}{\partial \theta_{-i}} = \frac{2\theta_1^3(\theta_2 - 3)y_2}{\theta_2 (\theta_1^3 \theta_2^3 - \theta_1 \theta_2 + 2\theta_1 + 2\theta_2 - 4)} \leq 0 \quad (\text{B.10})$$

□

The expected profit of a publisher with type θ_i reporting $\hat{\theta}_i$ is

$$u_i(\hat{\theta}_i, \theta_i, y_i) = \alpha_i(\hat{\theta}_i, \theta_{-i}) \left[x(y_i, y_{-i}^*(\hat{\theta}_i, \theta_{-i})) - x(y^*(\hat{\theta}_i, \theta_{-i})) \right] \quad (\text{B.11})$$

$$+ c(y^*(\hat{\theta}_i, \theta_{-i}), \hat{\theta}_i) - \int_0^{\hat{\theta}_i} \frac{\partial c}{\partial \theta}(y_i^*(s, \theta_{-i}), s) ds - c(y_i, \theta_i) \quad (\text{B.12})$$

The optimal allocation of efficiency units is $y_i^* = \frac{\theta_i^3(\theta_{-i}^3 + \theta_{-i} - 2)}{\theta_i^3\theta_{-i}^3 - \theta_i\theta_{-i} + 2\theta_i + 2\theta_{-i} - 4}$.

We note that $u_i|_{\hat{\theta}_i=0} = 0$ since $\alpha_i = 0$ in this case, which is the first sufficient condition of Lemma 6.

The publisher will then choose to show y_i ads that solve the first and second order conditions:

$$\frac{\partial u_i}{\partial y_i} = \alpha_i(\hat{\theta}_i, \theta_{-i}) \frac{\partial x}{\partial y_i} - \frac{\partial c}{\partial y} = 0 \quad (\text{B.13})$$

$$\frac{\partial^2 u_i}{\partial y_i^2} = \alpha_i(\hat{\theta}_i, \theta_{-i}) \frac{\partial^2 x}{\partial y_i^2} - \frac{\partial^2 c}{\partial y^2} = -\frac{\partial^2 c}{\partial y^2} < 0 \quad (\text{B.14})$$

The publisher will therefore choose \hat{y}_i s.t.

$$\alpha_i(\hat{\theta}_i, \theta_{-i}) = \frac{-\frac{\partial c}{\partial y}(\hat{y}_i, \hat{\theta}_i)}{\frac{\partial x}{\partial y_i}(\hat{y}_i, y_{-i}^*(\hat{\theta}_i, \theta_{-i}))} \quad (\text{B.15})$$

We note that $\hat{y}_i|_{\hat{\theta}_i=\theta_i} = y_i^*$, which is the second sufficient condition of Lemma 6.

We therefore need to prove that the payments b_i are incentive compatible for the publishers.

Denote $x_i = \frac{\partial x}{\partial y_i}$ and $c_y = \frac{\partial c}{\partial y}$.

Differentiating (B.15) with respect to $\hat{\theta}_i$ yields:

$$\frac{\partial \hat{y}_i}{\partial \hat{\theta}_i} = \frac{\frac{\partial \alpha_i(\hat{\theta}_i, \theta_{-i})}{\partial \hat{\theta}_i} + \frac{c_y}{x_i^2} x_{ij} \frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i})}{\frac{c_{yy}}{x_i}} \geq 0 \quad (\text{B.16})$$

This inequality holds since $x_{ij} \frac{\partial y_j^*}{\partial \theta_i}(\hat{\theta}_i, \theta_{-i}) \geq 0$ by Lemma 7 and $\frac{\partial \alpha_i(\hat{\theta}_i, \theta_{-i})}{\partial \hat{\theta}_i} \geq 0$ by Theorem 3 of McAfee and McMillan (1991).

The remainder of the proof follows the proof of Theorem 2 in McAfee and McMillan (1991), p. 574. Using the fact that $\frac{\partial \hat{y}_i}{\partial \hat{\theta}_i} \geq 0$ is then sufficient to prove incentive compatibility. \square

Proof of Lemma 8. Let $q_i^* = \frac{y^*}{\theta_i}$. Then $\frac{\partial q_i^*}{\partial \theta_i} > 0$. As showing any number of ads not in q_i^* will yield zero revenue with positive costs, the publisher will prefer to show ads in q_i^* only.

Since q_i^* and y_i^* are both monotonically increasing in θ_i , choosing to show the optimal number of ads such that $\hat{\theta}_i = \theta_i$ is an equilibrium strategy for the publisher.

Finally, $b_i(x, q)$ is well defined. Suppose there are $\theta_1^1 \neq \theta_1^2$ s.t. $q_i^*(\theta_1^1) = q_i^*(\theta_1^2)$ yet $b_i(x, \theta_1^1) > b_i(x, \theta_1^2)$. Then because the utility of the publisher increases with the payment, the publisher would prefer to claim its type is θ_1^1 when its true type is θ_1^2 . This contradicts the truthfulness of the direct mechanism. Hence $b_i(x, \theta_1^1) = b_i(x, \theta_1^2)$. \square

B.2 Asymmetric Publishers

We briefly overview the modeling of asymmetric effectiveness of publishers and results about the impact on campaign effectiveness.

When publishers are asymmetric the advertiser may want to compensate them differently depending on their contribution to the conversion process. If we assume the advertiser has full knowledge of the effectiveness level of each publisher, we can treat publisher one's effectiveness as fixed, and use the relative performance of publisher two as influencing its costs. Specifically, we let the cost of publisher two be $\frac{q_2^2}{2\theta}$.

When $\theta = 1$, we are back at the symmetric case. When $\theta < 1$, for example, publisher one is more effective as its costs of generating a unit of contribution to conversion are lower.¹

Solving for the decision of the publishers and the advertiser under CPM and CPA contracts yields the following results:

Proposition 13. *When publishers are asymmetric:*

- Under a CPM contract the same price $p^M = \left(\frac{\rho(\theta+1)^{\rho-1}}{2}\right)^{\frac{1}{2-\rho}}$ per impression will be offered to both publishers.
- Under a CPA contract, if $\theta < 1$, the advertiser will contract only with publisher one. If $\theta > 1$, the advertiser will only contract with publisher two.

We observe that asymmetry of the publishers creates starkly different incentives for the advertiser and the publishers. Under CPM campaigns having more effective publishers in the campaign increases the price offered by the advertiser to all publishers. As a result publisher one will benefit when a better publisher joins the campaign yet will suffer when a worse one joins.

Performance based campaigns using CPA, in contrast, make the advertiser exclude the worst performing publisher from showing ads. The intuition is that because conversions are generated by symmetric “production” input units of both publishers, the advertiser may just as well buy all of the input from the publisher who has the lowest cost of providing them. The only case when it is optimal for the advertiser to make use of both publishers is when $\theta = 1$ and they are symmetric.

¹It should be noted that this specification is equivalent to specifying the costs as being equal while the conversion function being $x(q_1, \zeta q_2)$ for some value ζ .

Using a single publisher is significantly less efficient when two are available to the publisher. Adding an attribution process creates an opportunity for this shut-out publisher to compensate for its lower effectiveness with effort. The resulting asymmetric equilibrium is currently under investigation to understand the ramifications of the attribution process on such a campaign.

Asymmetric Information with Asymmetric Publishers

When the publishers may be asymmetric yet their relative asymmetry is unknown to the advertiser, the problem exhibits adverse selection. The mechanism design literature has dealt with similar scenarios when either moral hazard is present, *i.e.*, the effort of publishers is unobserved, or with a scenario when both moral hazard and adverse selection are present. A novel result by McAfee and McMillan (1991) has developed second-best mechanisms for the case of team production when agents are complements in production.

We extend this result to the case of substitute production and note that publishers can be seen as contributing a measure of output we call *efficiency units* to the performance of the campaign x (McAfee and McMillan, 1991). This measure of input to the advertising process is not observed by the advertiser, but the mechanism will elicit optimal choice of efficiency units in equilibrium.

Let $y_i = \theta_i q_i$ be the output of publisher i measured in efficiency units, and let $y = (y_1, y_2)$, $\theta = (\theta_1, \theta_2)$ be the vectors of efficiency units and effectiveness of publishers. We assume $\theta_i \sim U[0, 1]$. Then the expected observed performance will be $x(y) = (y_1 + y_2)^\rho$, and the cost of each publisher will be $c(y_i, \theta_i) = \frac{y_i^2}{2\theta_i}$.

We denote by $\gamma(y_i, \theta_i) = c(y_i, \theta_i) - (1 - \theta_i) \frac{\partial c}{\partial \theta}(y_i, \theta_i)$. This is the virtual cost as perceived by the advertiser resulting from the actual cost of the publisher and the cost of inducing the publisher to reveal its true effectiveness.

We focus on a direct mechanism where publishers report their effectiveness which we denote as $\hat{\theta}_i$. By the revelation principle any incentive compatible mechanism can be mimicked by such a direct mechanism which is truthful in equilibrium. As the assumption that publishers can send a message about their effectiveness θ_i departs from reality, we later show how this assumption translates to a world where publishers can only make a choice about the number of ads to show.

The scheme the advertiser will offer to the publishers is $\{\hat{\theta}, x, b_1(x, \hat{\theta}), b_2(x, \hat{\theta})\}$, where x is the observed performance and b_i are the payments offered to publishers based on the observed output and reported effectiveness. During a campaign, publishers will report their types $\hat{\theta}_i$, and after the output x is determined, they will receive the payment $b_i(x, \hat{\theta})$.

Define the payments b_i as following when publishers report types $\hat{\theta}_i$ and performance x is observed:

$$b_i(x, \theta) = \alpha_i(\theta) [x - x(y^*(\theta))] + c(y^*(\theta), \theta_i) - \int_0^{\theta_i} \frac{\partial c}{\partial \theta}(y_i^*(s, \theta_{-i}), s) ds \quad (\text{B.17})$$

In the above specification:

$$\alpha_i(\theta) = \frac{\frac{\partial c}{\partial y}(y^*(\theta), \theta_i)}{\frac{\partial x}{\partial y_i}(y^*(\theta))} \quad \text{and} \quad y^*(\theta) = \arg \max_y x(y) - \gamma(y_1, \theta_1) - \gamma(y_2, \theta_2) \quad (\text{B.18})$$

To understand the intuition behind the definition of these payment functions, we first note that y^* finds the optimal allocation of efficiency units the advertiser would like to employ if the cost of advertising amounted to the virtual cost. The advertiser needs to consider the virtual cost since the mechanism is required to incentivize high effectiveness publishers to truthfully report their type and not try to impersonate lower effectiveness publishers. The advertiser then calculates a desired performance level for the campaign given publishers' reports, and pays publishers only if the output matches or exceeds this level.

The payment gives each publisher a share α_i of $x - x(y^*(\theta))$, which is the difference in performance from the expected optimal output given the publishers reports. In addition, the publisher is paid the expected optimal costs of showing ads $c(y^*(\theta), \theta_i)$ corrected for the expected information rent.

We can now prove the following result that shows the payments in (B.17) yield the optimal result for the advertiser:

Proposition 14. *When $\frac{\partial^2 x}{\partial y_i \partial y_j}(y^*(\theta)) \cdot \frac{\partial y_j^*(\theta)}{\partial \theta_i} \geq 0$ the payments in (B.17) yield optimal profit to the advertiser. They are incentive compatible and individually rational for publishers. In equilibrium, publishers will choose to generate $y^*(\theta)$ efficiency units.*

Proposition 14 shows that when publishers are substitutes in production and when the equilibrium allocations are substitutes, then the linear contract is optimal, extending McAfee and McMillan (1991) for the case of substitution among the publishers.

The intuition behind this result is subtle. When the other publishers $-i$ are of higher effectiveness they will produce more output in equilibrium. The resulting externality on publisher i 's profit will then be stronger and as a result it will decide to increase its own output to compensate and redeem its share of the profits. In equilibrium these effects cause the publisher to increase its output with its type, which is a result similar to the standard monotonicity result in single agent mechanism design.

The optimal mechanism allows the advertiser to efficiently screen among publishers at the cost of giving positive rent to very effective publishers. The payment scheme is built as a sum of two separate payments: payment for performance and payment for effort of displaying ads. Using the ratio of the marginal cost to the marginal productivity of the publisher as the share of performance given to the publisher, the advertiser is able to align the incentives of the publisher at the margin. In equilibrium the most effective publishers will show the full information (first-best) number of ads, but will also receive the highest share of profit from the advertiser. The publishers with the lowest effectiveness will be excluded from the campaign, and will receive zero profits. An interesting aspect of this payment scheme is that publishers receive less expected rent when compared to the two standard CPA and CPM

schemes, which is a result of using a combination of reported types and observed performance of the campaign.

The optimal mechanism yields improved results compared to standard compensation schemes at the cost of requiring the assumption that publishers can report their effectiveness. In an advertising campaign, however, publishers can only choose the effort they spend in terms of number of ads they show. To remedy this technical issue we employ the taxation principle to transform the direct mechanism into an indirect mechanism where publishers choose the number of ads to show and the output they will achieve. Based on the observed performance and effort, the advertiser will pay b_i in the following way:

Lemma 8. *Let $b_i(x, q) = b_i(x, \hat{\theta})$ when $q = (y_i^*(\hat{\theta})/\hat{\theta}_i)$ and $b_i = 0$ otherwise. Then $b_i(x, q)$ yields the same equilibrium as the mechanism in (B.17).*

Although this result is standard in mechanism design, it typically requires the assumption that publishers can be punished with arbitrary severity when they do not choose output that the advertiser would prefer. We are able to show that in our case, not paying anything sufficient for the mechanism to still be a truthful equilibrium.

The caveat, however, is that the resulting mechanism is highly non-linear in the effort of publishers. The monotonicity of publisher effort also does not hold for many specifications, which gives rise to multiple equilibria of the indirect mechanism. Another issue that arises is the effect of the baseline conversion rate of consumers. This was not considered previously and will prove to be detrimental to these mechanisms.

As it is highly unlikely that advertisers can implement such a mechanism in the reality, we choose to develop a simpler mechanism that holds potential for achieving profits that are closer to the full information (first-best) profits.

B.3 Estimation of Publisher Effectiveness

We let $x_i = 1$ denote exposure by consumers to ads from publisher $i \in N$, and specify the following discrete choice model: The utility of a converting consumer j exposed to a subset of ads $I \subseteq 2^N$ is specified as $u_{jI} = s + \sum_{i \in I} b_i x_i + \epsilon_{jI}$, with s the basic utility of consumers in the baseline. A consumer converts if $u_{jI} > s + \epsilon_{j\emptyset}$.

If we assume the ϵ_{jI} are distributed i.i.d. extreme value, we expect to see the population conversion rate $y_I = \frac{e^{\sum_{i \in I} b_i x_i}}{1 + e^{\sum_{i \in I} b_i x_i}}$. For each subset I observed in the data we have exact values for this conversion rate, as well as the total population value who were exposed to ads. We therefore do not need to make assumptions about the total population size or estimate it from the data.

We then have

$$\ln y_I - \ln(1 - y_I) = \sum_{i \in I} b_i x_i \quad (\text{B.19})$$

which is estimated using OLS.